AD 651 784

PSYCHOLOGICAL RESEARCH ON PILOT TRAINING

Neal E. Miller

Army Air Forces
Washington, D.C.

1947

Best Available Copy

AD-651784

Psychological Research on Pilot Training

REPORT NO. 8

Army Air Forces
Aviation Psychology Program
Research Reports

Psychological Research on Pilot Training

REPORT NO. 8

Edited by
NEAL E. MILLER
Research Associate in Psychology
(Associate Professor Rank)
Yale University

1947

Preface

The purpose of this report is to provide a record of the research on pilot training conducted by aviation psychologists under the Office of the Surgeon in the AAF Training Command. It has been prepared in compliance with directives from Headquarters, Army Air Forces, and Headquarters, Training Command. No attempt is made to give comprehensive references to psychological research conducted outside of the AAF Training Command because most of that work has not yet been officially published. The final chapter (No. 15) is a summary presented for the benefit of those readers who are more interested in an integrated picture of the main results than in the technical details of the separate studies.

Most of the research summarized in this report was conducted by Psychological Research Project (Pilot), an organization which was established and made responsible for psychological research on pilot training on 1 February 1944. A smaller but important amount of research was performed by other psychological units.

The research which is reported could not have been conducted without the strong support of higher headquarters, and especially from the aviation psychologists in Headquarters, Army Air Forces, and Headquarters, Training Command. Valuable assistance has been received from the Surgeon, the Pilot Section of the A-3 Division, and the Statistical Control Unit at Central Flying Training Command; the Director of Training and Training Advisory Boards of the Central Instructors School at Randolph Field; and the Director of Training and Standardization Boards at the Instrument Pilot School at Bryan. During the crucial initial stages of the development of Psychological Research Project (Pilot), particularly helpful, insightful support was received from the following officers in the Central Instructors School at Randolph Field: Maj. Charles H. Roadman, Director of the Ground School; Lt. Col. Charles M. Wharton, Commanding Officer of the Bomber Training Group; and Col. Merill J. Rech, Surgeon. Pilots from a number of different organizations have supplied expert advice and criticism on the technical aspects of flying. The organizations which were the chief sources of such assistance are listed immediately following this preface.

The research reported represents the cooperative, creative work of a number of psychologists. A roster of personnel who have worked in the Pilot Project precedes the table of Contents of this report. Particular note should be made of the fact that the lower army status of the enlisted men did not necessarily mean that their work was of lower professional calibre.

The authors, whose names appear at the head of each chapter, were responsible for the supervision of most of the work reported in their chapters, with the exception of Captain Ben-Avi who was given the difficult task of reporting work with which he had no first-hand contact. Although it is difficult to assign credit equitably for cooperative work of this kind, an attempt has been made to cite the chief contributors to each of the research studies reported.

Six men deserve special mention here because their contributions were of a general nature which is not credited in the individual studies. Sgt. Allen J. Sprow performed an excellent job of writing the semimonthly reports summarizing the research of the Pilot Project and of assisting with the annual reports. This work, which required insight into all of the research involved, left no time free for conducting studies of his own. T/Sgt. Robert R. Blake was so useful as a general administrative assistant in roles ranging from liaison with other organizations on the field and establishing filing systems, through that of serving as a general critic and conscience, that he was given little time for anything else. Similar capable service in a number of administrative roles was rendered by T/Sgt. Robert E. Dixon. During his three months with the Pilot Project, Lt. Maurice Deigh displayed initiative and resourcefulness in dealing with problems of supply and assisting in many of the administrative details involved in the preparation of this report. In the final stages of the preparation of the manuscript intelligent help was received from Dr. John T. Cowles and Capt. John T. Dailey, the new Director and Assistant Director of the Pilot Project.

NEAL E. MILLER, *Major, Air Corps.*

Randolph Field, Tex., 11 February 1946.

Organizations Supplying Technical Consultants

AAF Board, AAF Tactical Center.
AAF Instrument Flying Standardization Board, Bryan Army Air Field.
AAF School of Applied Tactics, AAF Tactical Center.
Advanced Single-Engine Training Advisory Board, Central Instructors School (Pilot), Randolph Field.
Advanced Two-Engine Training Advisory Board, Central Instructors School (Pilot), Randolph Field.
Basic Training Advisory Board, Central Instructors School (Pilot), Randolph Field.
Flight Test Section, Proving Ground Command, Eglin Field.
Ground School, Lockbourne Army Air Base (Pilot Instructors School, B–17).
Ground School, Smyrna Army Air Field (Pilot Instructors School, B–24).
Ground Training Technical Advisory Department, Central Instructors School (Pilot), Randolph Field.
Instrument Board, Central Instructors School (Pilot), Randolph Field.
Instrument Training Squadron, Pampa Army Air Field.
Office of Flying Safety.
Physiological Section, Proving Ground Command, Eglin Field.
Physiology Department, School of Aviation Medicine, Randolph Field.
Pilot Section, A–3 Division, Headquarters, AAF Central Flying Training Command, Randolph Field.
Pilot Section, A–3 Division, Headquarters, AAF Training Command, Fort Worth.
Primary Training Advisory Board, Central Instructors School (Pilot), Randolph Field.
Station Standardization Board, Instructors School (Instrument Pilot), Bryan Army Air Field.

Roster of Personnel

No.	Name	*Number of months duty at PRP*
1.	Aborn, Murray	1
2.	Abner, Charles W	2
3.	Ben-Avi, Avrum H	3
4.	Benson, Richard W	1
5.	Bickhart, Benjamin H., Jr	2
6.	Blake, Robert R	17
7.	Boss, James F	1
8.	Bowles, J. W., Jr	2
9.	Brown, Walter T	1
10.	Cobb, Bart B., Jr	2
11.	Connery, Maurice F	22
12.	Davy, Elvin B., Jr	2
13.	Deigh, Maurice	3
14.	Delman, Louis	1
15.	Derry, James	1
16.	Derthick, Charles	3
17.	Dixon, Robert E	10
18.	Ericksen, Stanford C	21
19.	Ewing, Thomas N	9
20.	Fassnacht, George A	10
21.	Ficere, Raymond S	2
22.	Foulke, Stuart	10
23.	Freedman, Abraham	3
24.	Freedman, Morris	11
25.	French, Benjamin L., Jr	1
26.	Friedman, Sidney	2
27.	Galt, William E	21
28.	Gershenson, Charles P	19
29.	Glaser, Ezra	5
30.	Glass, Robert E	10
31.	Gleason, John G	13
32.	Grant, Henry	1
33.	Grier, Daniel J	19
34.	Gutsche, Lyle D	3
35.	Hagin, William V	11
36.	Hagy, Harold H	5
37.	Hansen, Vern J	2
38.	Harris, William E	14
39.	Heathers, Glen L	3
40.	Heidenreich, Frank W	2
41.	Hellmer, Leo A	6
42.	Hemphill, John K	6
43.	Hodgson, Thomas J	2
44.	Holton, Elizabeth	1
45.	Ismael, Walter W	15
46.	Johnson, J. R	1
47.	Johnston, Roland E	1
48.	Kamman, James F	10
49.	Knicely, Robert S	3
50.	Krasnow, Robert H	1
51.	Latham, Albert J	4
52.	Lauber, Robert W	2
53.	Levine, Robert	11
54.	Lewis, Quentin A	10
55.	Lindberg, John E	1
56.	Lyster, Harold W	3
57.	Matheny, William G	16
58.	McCarthy, Warren J	1
59.	Miller, Clarence K	8
60.	Miller, Florence E	1
61.	Miller, Neal E.[1]	23
62.	Nieder, Leonard	10
63.	Nygard, John W	5
64.	Pope, Henry T	10
65.	Race, Lillie M	1
66.	Resnick, Anne S	1
67.	Robbins, Irving	20
68.	Rohrs, John R	20
69.	Roitblat, Alvin	10
70.	Rust, Ralph M	4
71.	Showalter, Ralph E	12
72.	Smith, Robert W	10
73.	Spivak, Daniel C	6
74.	Sprow, Allen J	19
75.	Stratton, James W	14
76.	Stubblefield, Eben M	3
77.	Stuckman, William D	9
78.	Teachout, Hillard D	1
79.	Tibbetts, Palmer G	10
80.	Tucker, Anthony C	1
81.	Van Nostrand, Maurice A	4
82.	Waeltermann, John J	12
83.	Walker, Stuart B	10
84.	Weller, Mary S	1
85.	Wiesenberg, William H	10
86.	Wilson, Richard W	1
87.	Wurst, Ralph, Jr	7
88.	Youtz, Richard P.[2]	23

[1] Director.
[2] Assistant Director.

Contents

APPENDIXES

List of Tables

703322—47——2

List of Figures

CHAPTER ONE

Areas and Conditions of Research

Maj. Neal E. Miller

INTRODUCTION

Purpose of Report

This report is written as a record of the research on pilot training conducted during the second World War by aviation psychologists in the AAF Training Command. It is one of a series dealing with the various types of wartime psychological research under the direction of the Office of the Air Surgeon.

The two chapters following this one present a sketch of the pilot-training program and a description of the task of learning to fly. They will serve as a background for understanding the research reported. The main body of the report, chapters 4 through 14, is intended to be a source of technical information concerning the procedures and results of psychological research on pilot training during the war. The final chapter summarizes the main highlights for those readers who are more interested in the chief conclusions than in technical details, or who want to get a perspective before immersing themselves in the separate studies. A glossary of military, aeronautical, and statistical terms is included at the end of the report.

Gradual Development of Opportunity for Psychological Research on Training

Aviation psychologists were originally called into the Air Forces to construct tests for the selection and classification of pilots, bombardiers, and navigators. During the first years of the war their efforts were almost entirely devoted to developing and administering these tests. Research on training developed gradually out of this original work on selection and classification. The three lines of work leading from classification into training were: (1) Job analysis studies conducted to provide a basis for test development; (2) investigations of measures of flying proficiency conducted to learn more about the criteria available for validating the aptitude tests; and (3) attempts to apply the techniques used successfully in developing a battery of pilot-selection tests to the problem of instructor selection.

From the very first, a number of aviation psychologists tried to visit pilot training schools and obtain some flying training in order to be able to make first-hand, job-analysis observations as a basis for developing the pilot aptitude tests. Because of the pressure of developing an extensive battery of aptitude tests quickly and administering them to ever-increasing numbers of students in the emergency and because of certain administrative difficulties, it was not until early in 1943 that an opportunity was finally secured for two officers to spend 6 weeks at a pilot training school and to receive a certain amount of flying training.[1]

At the same time, studies were being made of the criteria used to validate the pilot aptitude tests. The three original psychological research units made statistical studies of the pass-fail criterion, investigating the variability of elimination rates at different schools. The Field Studies Unit, Psychological Section, Headquarters, Training Command, visited schools in the spring and summer of 1942 to introduce a rating scale which was designed in an attempt to get more specific measures of various aspects of pilot performance to use as differential criteria in validating aptitude tests.[2]

As validation data accumulated, demonstrating that the aptitude tests were successful in selecting those students who were least likely to be eliminated for flying deficiency, the individuals responsible for pilot training tended to seek the aid of aviation psychologists on other problems. Officers at Psychological Research Unit No. 2 [3] were asked to try to develop a battery of tests for the selection of civilian flying instructors who were to be employed in Primary schools.

Meanwhile a research detachment which originally had been sent to a Gunnery Instructors School to work on the selection of flexible gunners had grown into a full-fledged Unit and attracted considerable attention by the technical assistance which it was giving to the gunnery training program.

Furthermore, a staff study in Headquarters, Army Air Forces recommended that aviation psychologists be used for research on training as well as on selection.

These developments finally led to the establishment on 1 February 1944 of an organization, the Psychological Research Project (Pilot), which was eventually given primary responsibility for psychological research on pilot training in the Training Command. This Project was on Randolph Field and was attached first to the Central Instructors School and later to the Headquarters of Central Flying Training

[1] The officers were Maj. Neal E. Miller and Capt. Donald E. Super from Psychological Research Unit No. 1. An earlier brief contact with training had occurred when Lt. Col. Lawrence F. Shaffer and Maj. Neal E. Miller gave lectures on the psychology of teaching to pilots in the Central Instructors School at Maxwell Field. This contact was terminated when Psychological Research Unit No. 1 was moved from Maxwell Field, Ala., to Nashville, Tenn.

[2] Lt. Col. Richard T. Sollenberger was the officer who at first was most responsible at Headquarters, Training Command, for fostering the development of psychological research on problems related to pilot training. After he left for the Continental Air Forces, this role was ably assumed by Maj. S. Rains Wallace.

[3] Capts. Richard P. Youtz, John T. Dailey and Glen Finch were chiefly responsible for initiating this work.

Command.[4] Similar organizations were also established at the Navigator and the Bombardier Central Instructors Schools and were given the responsibility for psychological research related to training in those specialties. Meanwhile the three much larger, original psychological research units continued to conduct studies related to training as a sideline to their main mission of developing aptitude tests.

AREAS OF RESEARCH

The general policy of the psychology program was that aviation psychologists who were working in an area such as pilot training, should conduct fundamental research on problems of practical importance rather than perform service functions. The specific problems on which the Psychological Research Project (Pilot) worked were determined by directives from higher headquarters. The two main problems specified were to develop techniques for selecting and evaluating flying instructors and to develop objective measures of flying skill. The Project was also assigned responsibility for developing printed tests of flying information, and for making a job analysis of the student pilot's task. It was able to find a little time in which to perform training experiments.

Instructor Selection

When the Pilot Project was first established, Headquarters, AAF Training Command, directed it to work on problems of instructor selection and to confine its research completely to such problems. Work on instructor selection consisted of analyzing the job of teaching flying, constructing a battery of tests and information blanks to get at factors believed to be related to success as an instructor, and examining possible criteria for measuring the success of instructors. Since no ready-made criteria were found, it was necessary to develop rating scales to measure teaching proficiency. Finally, these scales were used to determine the validity of a number of instructor-selection tests and to make an extensive study of factors related to success as a flying instructor. This work is described in chapter 14.

Objective Measures of Flying Skill

In May 1944, before the work on instructor selection was completed, the Pilot Project was directed to concentrate on investigating the feasibility of constructing an objective scale of flying skill. In this work, Colonel Flanagan, Chief of the Psychological Branch, AFTAS, believed that a large stock of measures should be accumulated before work was devoted to other aspects of the problem, such as combining the measures into a total score, analyzing sources of variance, computing intercorrelations, or comparing the relative merits of check rides

[4] A roster of personnel who have worked in the Pilot Project is given preceding the Table of Contents of this report.

composed of objective, subjective, or a combination of these two types of measures. He therefore instructed the Pilot Project to devote its time to the construction of a large number of objective measures and to the preliminary evaluation of the suitability of each of these measures.

Because of the magnitude of the task of constructing and evaluating items to measure the entire range of flying skill at all levels of training, the Pilot Project concentrated, in the first development of objective measures, on two crucial areas: the early part of Primary training, and instrument flying. The early part of Primary training was selected because that was the time when most eliminations occurred. Therefore, any improvement in the measures used as a basis for elimination at that level would have special practical importance at that time. It was also somewhat easier to work at this level because the range of talent had not yet been so greatly restricted by elimination of the poor flyers. Instrument flying was selected because of its extreme importance in combat and transport and because the fact that it is conducted solely with reference to instruments made it more adaptable to objective measurement.

In addition to these two main areas, some work was done on fixed gunnery. This area was selected because the number of hits on a target appeared to be a readily available, objective measure of an important aspect of the fighter pilot's task.

In general, the injunction to concentrate the first work on item construction and evaluation was adhered to closely with the exception of one large-scale objective study of the effects of additional training on flying skill. This study was conducted at the personal request of the Chief of Staff of the Training Command.

It was possible to design some of the item-evaluation studies so that, as a byproduct, they yielded information on sources of variability and on total scores. These results are given more space in the main body of this report than would be proportional to the amount of research time involved because they appear to have a wider range of applicability than information on the reliability and validity of each one of a long list of items.

Before the stage of item construction and preliminary evaluation was quite completed, so that the Project could move on to the next problems involved in this work, the end of the war disrupted training to such an extent that further research had to be temporarily suspended.

The research on objective measures of flying skill is described in chapters 6 through 11. Some of the theoretical problems involved are discussed in chapter 4.

Printed Tests of Flying Information

As a supplement to its work on developing objective measures of flying skill, the Pilot Project initiated the development of printed tests of flying information and was subsequently assigned chief responsibility for such tests in the pilot area. There was a need for these tests because, in addition to flying skill, the good combat pilot had to have a large fund of specialized knowledge. In order to be able to get maximum performance and cope with emergencies, he needed a thorough, practical understanding of the aerodynamics and flying characteristics of his airplane, and the design, functions and limitations of its equipment. He also needed to know crucial facts about weather and navigation. Some of these types of information could be measured more efficiently on the ground than in the air. The development of printed tests to measure these types of information is described in chapter 12.

Subjective Measures of Flying Proficiency

As has already been indicated, the three original psychological research units investigated various subjective measures of flying proficiency as possible criteria to be used in validating aptitude tests. These studies are described in chapter 5.

Job Analysis

As a basis for developing aptitude tests and objective measures of flying skill, various organizations in the aviation psychological program analyzed different aspects of the task of learning to fly. Chief responsibility for this work was eventually assigned to the Pilot Project. This work is summarized in chapter 3.

Training Experiments

One of the more useful functions of aviation psychologists should be in serving as consultants on the design of training experiments. The pressure of work directed by higher headquarters on projects more closely related to the aptitude testing program, however, was so great during the war that almost no opportunity was left for aviation psychologists to work on pilot training experiments. Two minor studies, which were conducted under conditions preventing the best experimental design, are described in chapter 13 in order to illustrate the type of questions which should be answered by the experimental method.

CONDITIONS OF RESEARCH

Applied Science

The research of aviation psychologists on pilot training during the war was applied, not pure, science. The direction of work was guided by rigid practical needs instead of more flexible scientific curiosity. The need to achieve specific goals within a reasonably short time excluded the possibility of exploring interesting incidental leads (like the one which led Fleming to the discovery of penicillin) to obtain knowledge which in the long run might be more important than the original problem.

Place of Pilot Project in the Structure of the Army

The Pilot Project was a part of a program which had representatives in higher headquarters. Direct channels on technical matters between the Project and these representatives were most useful in affording an overall perspective on those problems which were likely to become most significant, in giving the strong administrative support necessary for conducting research, and in obtaining authority to translate research findings into administrative action.

Of necessity, the great mass of army procedures was established to continue to deal with large numbers in the same standardized way for as long as possible. Research, by its very nature, involves dealing with relatively small numbers in different ways and continually moving on to new problems. Any research organization, therefore, was fated to confront the typical army station with a much greater number of exceptions to routine procedure than was proportionate to its size. The fact that these problems involved small numbers did not reduce their nuisance value, but only increased the possibility of the matter getting sidetracked by more pressing issues. It was necessary, therefore, for army research to be a part of a strong program having special access to High Headquarters with the authority to make exceptions in dealing with unusual problems.

Fitting a new type of research into the army system was bound to produce certain paradoxes. On the one hand, tremendous resources were placed at the disposal of the aviation psychologist. Thousands of cases were available. He could fly hundreds of miles in a few hours. On the other hand, unexpected difficulties were sometimes encountered.

In order to get the Pilot, Bombardier and Navigator Projects established quickly, it was necessary to set them up without any Manning Tables, asking the local Fields to absorb the additional, highly trained personnel by dipping down into the limited allotment of vacancies which they had available as a reserve for promotions. The administrative setup was one in which higher headquarters exercised direct control over the work but the local Base Unit was responsible for the promotions, and had to accept an increased burden without any increase in the means at its disposal. As a result, some men with the Ph. D. degree performed psychological research for considerable periods of time in the grade of private.[4] Though many of these men were able to overcome the handicap of low grade and assume officer responsibilities, the lack of a grade commensurate with their training tended to make it more difficult for them to be maximally effective in certain army situations. Despite the shortage of scientifically trained enlisted personnel, an average of from 10 to 20 hours per week of each man's time had to be lost in activities, such as KP, close-order drill and

[4] Fortunately, the need for officers to serve as clinical psychologists later enabled some of these men to obtain the direct commissions they deserved.

parades, which were unrelated to his specialty but took automatic precedence over research. Finally, the considerable rate of turnover of men who had acquired special experience interfered with the continuity of research studies.

Although, in general, excellent results were achieved by the administrative arrangements which had to be improvised during the emergency, it is believed that considerable thought needs to be devoted to the problem of making a still better place in the structure of the army for an increasing number of new types of research personnel.

Need for Psychologists to Learn About Flying and For Pilots to Learn About Psychology

In order to be able to apply psychological principles and methods to the problems of flying training, the Aviation Psychologists needed to acquire a thorough understanding of the particular conditions and techniques of this complex specialty. Since none of them was a pilot and since they were not authorized to receive formal flying training, the required intimate knowledge of the details of flying and of the vast and varied pilot training program had to be acquired gradually in devious ways, partly by receiving informal flying instruction, which was inefficiently unsystematic, and partly by working in close collaboration with specialists on pilot training. Fortunately, extremely competent and cooperative experts were available. The organizations supplying most of these experts are listed after the preface of this volume.

While the psychologists were learning about flying, Air Forces pilots were changing their ideas about psychology. At first, there was a tendency to think of the psychologist in terms of the popular stereotype of a mind reader, or one who delves into the vague underworld of mental abnormality. Some pilots had the uneasy feeling that they might be personally psychoanalyzed. They were surprised to learn that psychologists are interested in problems such as learning, tests and measurement, motor skill, fatigue, and hearing and vision, as well as fear and courage and normal and abnormal adjustment. There was a gradual change until psychologists were finally accepted by the Air Forces in the less colorful but more substantial role of scientists with special training in applying useful techniques of observation and measurement, statistics, and experiment to complex problems of human behavior.

CHAPTER TWO

Pilot Training in the Army Air Forces

Capt. Stanford C. Ericksen[6]

INTRODUCTION

The purpose of this chapter is to outline the organization of the Training Command and to describe the continually changing "laboratory" in which the Pilot Project conducted its research program. A bird's-eye view of the sequence of pilot training and the organization responsible for this task will give the reader a perspective for better understanding the research reports described in the following chapters. This chapter is not an historical summary nor is it evaluative in the sense of pointing out the good or weak aspects of the pilot training program. Attention is focused on those problems which have particular relevance for understanding the psychological research program.

SEQUENCE OF PILOT TRAINING

Description of Each Phase of Training

Preflight.—The purpose of this training was to begin the process of providing the cadets with the necessary technical information and military training required by an officer pilot. The cadets were arbitrarily assigned to their administrative units according to date of arrival or alphabetical order. In fact, assignment according to alphabetical order remained one of the more common bases of grouping throughout the rest of the cadets' training. These methods usually randomized the distribution of pilot aptitude, as reflected in the pilot stanine based on the battery of classification tests administered by aviation psychologists, but sometimes special selective factors resulted in far greater fluctuations than would be expected by chance in the ability of students in different squadrons and schools, or in successive classes in the entire command. Later in the war some effort was made to equalize the stanines of students sent from Preflight to the different Primary schools.

Primary.—This phase (sometimes referred to as Elementary training) was the first stage of military flying training.[7] The Primary

[6] Maj. I. F. Larkey, Chairman, Two-engine Advisory Training Board, Central Instructors School (Pilot), Randolph Field, Tex. and Sgt. Maurice F. Connery assisted in the preparation of material for this chapter.

[7] During the last half of 1943 and early 1944 aircrew candidates received up to 10 hours of dual flight training in light (65-75 h. p.) trainers while in the College Training Detachments. They did not solo.

training schools throughout the country were civilian operated schools whose facilities were leased by the Air Forces. However, the training curriculum, elimination policy, and the grading standards were formulated and supervised by the Training Command. A small military unit was established at each Primary school charged with the responsibility of supervising the military activities of the students. It was in this phase that the greatest elimination rate occurred, ranging as high as 50 percent for some classes at some schools. During their Primary training students usually received a total of 65 hours instruction, approximately 35 of which were dual and 30 solo.

Basic.—At the end of Primary training those students who successfully satisfied the requirements and appeared to have the necessary aptitude for flying were assigned to Basic schools. With but very few exceptions, the Basic schools were operated entirely by Air Forces personnel. The fundamental flight maneuvers were repeated in a more powerful and complicated airplane with new emphasis placed on maximum performance and precision execution. The cadets were given introductory training in night and formation flying and were taught the fundamental aspects of instrument flying.

Advanced.—At this level there were two types of specialized schools: Advanced single-engine and Advanced two-engine.

Advanced single-engine training followed the pattern of basic but in a faster plane with greater emphasis on formation flying, acrobatics, navigation, and the use of radio aids in instrument flying. Another new feature of the curriculum was the introduction to simulated fixed gunnery flying. The cadets flew camera missions and practiced this type of air-to-air and air-to-ground firing.

The Advanced two-engine curriculum was designed to train multi-engine pilots. The cadet was taught the peculiarities of multiengine flight with a good deal of emphasis on formation flying, night flying, navigation, and instrument flying. At the time of the peak load, approximately two-thirds of all Advanced school students were in the two-engine schools.

At the conclusion of the Advanced training phase the cadet became an officer in the Air Corps and was given the rating of pilot.

Instructors' School.—During the greater part of the war a considerable number of the recent graduates were retained in the Training Command as instructors for the Basic and Advanced schools since the expansion and rapid turn-over resulted in a chronic shortage of instructors. Immediately upon graduation these pilots were assigned to an instructor training squadron. For the most part these schools were conducted locally and the prospective instructor was given a week or two of standardized flight instruction by specially selected experienced instructors at the home station. Emphasis was placed on the uniform performance of the fundamental training maneuvers rather than on teaching techniques per se. A smaller proportion of the potential instructors were sent to central instructors schools which served the entire Command.

Transition.—The first intermediate step for rated pilots on their way to combat operation was a period of Transition training in combat type aircraft. The Transition schools were specialized in three groups: Fighters, Two-Engine, and Four-Engine (either B-17 or B-24). At this level emphasis was placed on familiarizing the pilot with the new tactical type plane with some practice in the fundamental maneuvers important in combat operations. For example, in the Fighter Transition schools about half the time was spent in learning and practicing fixed gunnery firing.

By the end of Transition training the pilots had received a total of approximately 40 weeks of intensive flying training designed to teach the individual student the skills and information necessary to make him a safe and competent pilot. At this point the pilot left the Training Command.

The Continental Air Forces.—The student-pilots were now ready for the further specialized training necessary in combat operations: skill in flying heavily loaded combat-type aircraft, use of special equipment, and proficiency in the fundamental maneuvers used in the overseas theaters. For the bomber pilot it was particularly important to add familiarization and integration with the other members of the combat crew. The four Continental Air Forces had the responsibility of conducting this phase of training prior to the assignment of the pilot and crew to the combat Air Forces. A more detailed description of the activities of the Continental Air Forces is given in the comprehensive report by the Aviation Psychologists assigned to these units.[1]

Combat Training Units.—After the pilot arrived at his overseas combat unit, continued specialized training was given in the procedures adapted to the particular conditions in that theater of operations.

A summary of the sequence of pilot training is presented graphically in figure 2.1.

ORGANIZATION OF THE TRAINING COMMAND

Headquarters and its Relation to Other Commands in the AAF

The AAF Flying Training Command was activated in Washington, D. C., in January 1942 and moved to Fort Worth, Tex. in July of the same year. This command was an autonomous Air Forces unit responsible only to Headquarters AAF in Washington, D. C. In July 1943 the Technical Training Command was combined with the Flying Training Command to form the AAF Training Command. The Training Command maintained liaison with the other major AAF

[1] Comprehensive Report No. 16. Psychological Research on Operational Training in the Continental Air Forces.

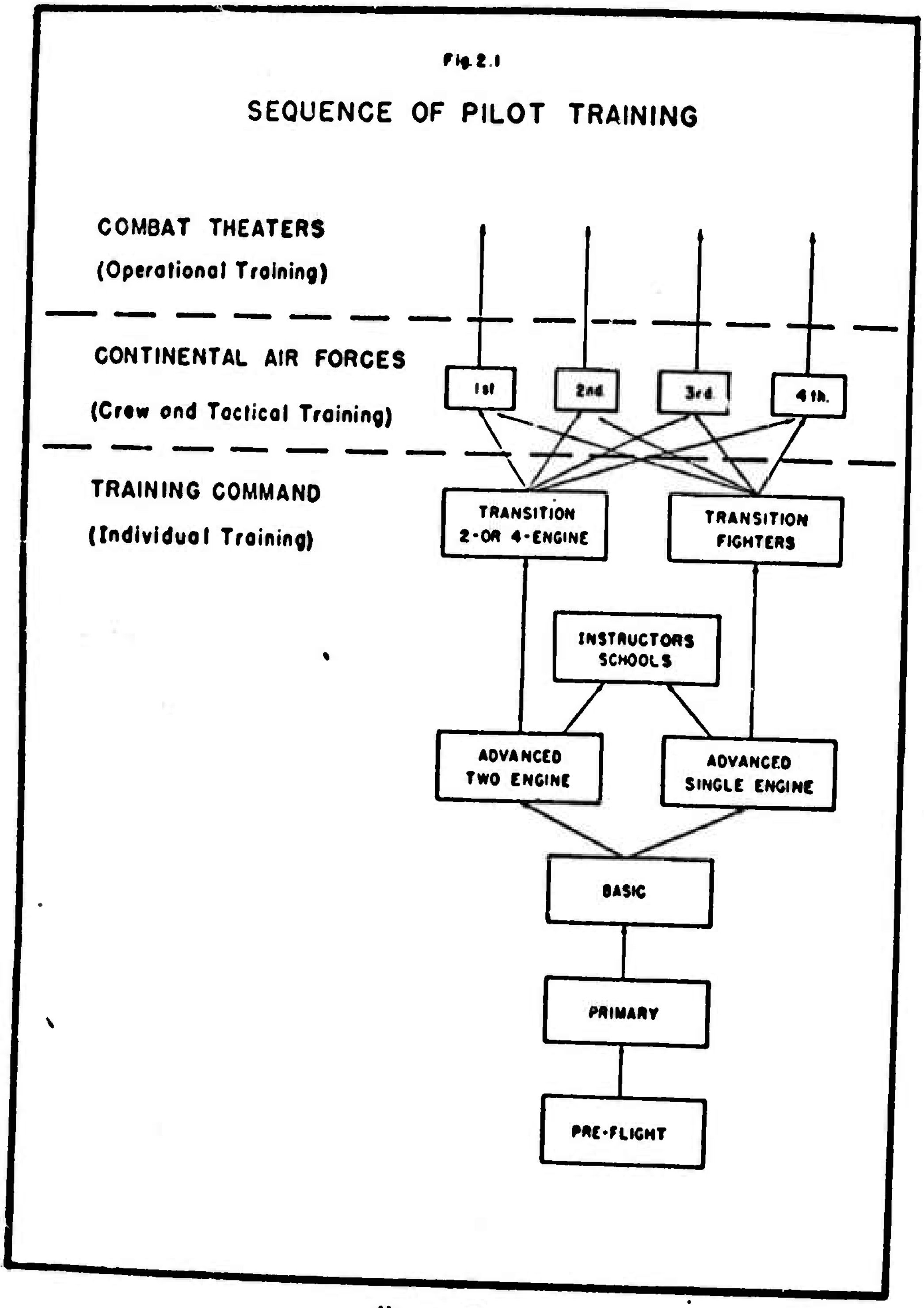


Fig 2.1
SEQUENCE OF PILOT TRAINING
COMBAT THEATERS
(Operational Training)
CONTINENTAL AIR FORCES
(Crew and Tactical Training)
1st
2nd
3rd
4th.
TRAINING COMMAND
(Individual Training)
TRANSITION 2-OR 4-ENGINE
TRANSITION FIGHTERS
INSTRUCTORS SCHOOLS
ADVANCED TWO ENGINE
ADVANCED SINGLE ENGINE
BASIC
PRIMARY
PRE-FLIGHT

FIGURE 2.1.

commands. With the return of pilots from combat tours, the Personnel Distribution Command became a source of supply for Training Command instructors and administrative officers.

Three Flying Training Commands

Before Training Command headquarters was established at Fort Worth, Tex., three regional flying training commands had been in operation, each being responsible directly to Headquarters AAF in Washington. The regional headquarters functioned throughout the war as administrative units supervising the training policies and procedures originating in Washington and later in the Fort Worth Headquarters for all of the training stations within their regional jurisdiction. This division of responsibility among the three regional flying training commands was necessary since no one headquarters could supervise adequately their operation and deal with the special problems of procurement and distribution of personnel and matériel for the large number of training stations devoted to pilot instruction. The three regional commands were: (*a*) Eastern Flying Training Command, headquarters at Maxwell Field, Ala.; (*b*) Central Flying Training Command, headquarters at Randolph Field, Tex.; and (*c*) Western Flying Training Command, headquarters at Santa Ana Army Air Base, Calif. While these names were changed from time to time, the above designations will be used throughout this report. The training program within the three flying training commands was essentially the same though there were always some differences which persisted in spite of efforts to standardize the pilot-training routine.

Figure 2.2 presents the areas covered by each of the flying training commands and the distribution of pilot-training schools within each. The distribution of schools is only representative of those in operation early in 1944 since there was a continuous process of closing some schools, activating new schools and changing the functions of others.

Flying Training Wings

As the training program expanded and became more specialized, a system of Wings was established within each of the three Flying Training Commands. A separate Wing was established for each level of pilot training. For example, one Wing supervised all Primary training in one command while other Wings operated at the Basic, Advanced, or Transition level. The Wings functioned as inspecting bodies charged with the responsibility of standardizing administrative, operational, and later, flying training procedures within their assigned area.

Pilot Training Schools

Figures 2.1 and 2.2 indicate the different types and the number of pilot training schools. Each of these schools was responsible directly to its regional command headquarters.

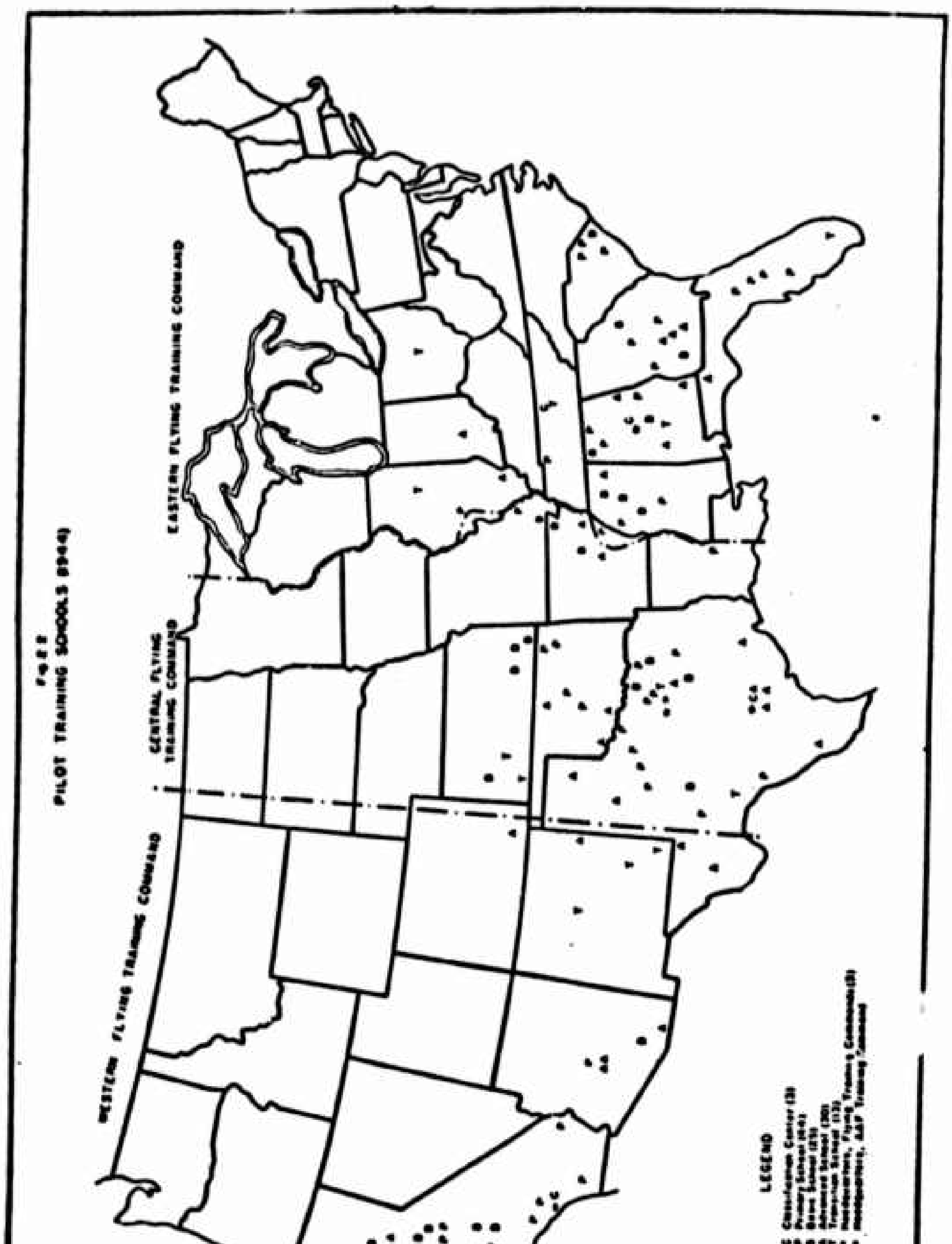

PILOT TRAINING SCHOOLS (1944)
WESTERN FLYING TRAINING COMMAND
CENTRAL FLYING TRAINING COMMAND
EASTERN FLYING TRAINING COMMAND
LEGEND

FIGURE 2.2.

Special Schools

Instructors schools were established for the major specialties of pilot training:

- Central Instructors School (Pilot) at Randolph Field, Tex.
- Central Instructors School (Instrument Pilot) at Bryan Army Air Field, Tex.
- Central Instructors School (Fixed Gunnery) at Matagorda, Tex.
- Four-engine Transition Instructors Schools at Lockbourne Army Air Base, Ohio for the TB-17 and at Smyrna Army Air Field, Tenn., for the TB-24.

DESCRIPTION OF THE STUDENT POPULATION

Successive Screening of Students in Pilot Training

AAF Qualifying Examination.—The first screening of candidates for pilot training was made by the AAF Qualifying Examination. This examination was taken by men still in civilian status and by enlisted men already in the army. It has been estimated that those who passed this examination represented the top twenty-five percent of high school graduates; the over-all results showed that about fifty percent of those who took the test failed to qualify as aviation students.[9]

The Pilot Stanine.—Those who passed this first hurdle were sent to a classification center where the medical and psychological selection and classification procedures were administered. Starting in December 1942, men with stanine scores of 1 and 2, the bottom 11 percent, were disqualified for pilot training. Gradually the minimum passing stanine score was raised until at the end of the war only men with the pilot stanine scores of 7, 8, and 9, the top 23 percent, were admitted to pilot training.

Preflight.—Very few cadets were eliminated in Preflight training until later in the war when the supply of pilot candidates exceeded the quota requirements. The most common basis for Preflight elimination, other than medical or own request, was failure in such courses as mathematics and code. Usually a student was given at least two chances to meet the minimum qualifications in these ground school subjects. This is a good example of the minimum weight given to ground school grades in the evaluation of student pilots.

Elimination of students during flight training.—Approximately 50 percent of the students who entered flight training were eliminated before completing Advanced school training. Over half of those eliminations occurred at the Primary level. The elimination rate varied considerably from school to school and from class to class; a

[9] A more detailed description of the nature of this examination, its use and results can be found in AAF Aviation Psychology Program Research Report No. 6, *The AAF Qualifying Examination.*

range of from 15 to 35 percent would be a fair estimate of the over-all elimination rate in Primary training.

Progressively fewer students were eliminated in Basic and Advanced training. Student officers who failed to meet the minimum proficiency requirements at the Instructor and Transition schools were usually transferred to some assignment which would make fewer demands upon their pilot skill.

A Homogeneous Population

Restricted range.—Insofar as the tests used in the original screening of pilot candidates correlated with success or failure in training, a restricted range of pilot aptitude was sent into flight training. In addition, the successive elimination of students in the different phases of training resulted in a fairly homogeneous population of students possessing a high average level of aptitude in advanced levels of training. As the need for pilots became less acute the restriction became greater both through the selection of candidates and in the graduation of students from the successive phases.

Implications for research.—This restricted range of pilot aptitude made the task of discriminating between *individual* students with objective or subjective measures of flying skill more difficult than it would have been with an unselected population. Since this factor will probably always exist in military aviation training, it is quite likely that higher proficiency test reliabilities would be obtained in conditions ordinarily found in civilian flying training.

In contrast to this difficulty for developing reliable measures of flying skill, the homogeneity of the population would have been a decided advantage in conducting experiments to compare the effects of different types of training or equipment since it would have reduced the standard error of the difference.

SCHEDULING TO MEET QUOTA REQUIREMENTS

Mass Production

The Training Command had to produce a large number of pilots in a short time. Planning was done in units of thousands of pilots. In order to meet the large requirements of the entire Air Forces, the Training Command set up a mass production assembly line system of pilot training. No provisions were made for individualized training except insofar as the instructors could vary their technique during the separate flights with each student. Everyone within a given school was given the same training with the same time limit for reaching the minimum level of proficiency in each sub-unit.

Rigid schedule.—Many technical schools in the army practiced the "hold-over" system of retaining slow students from class to class to bring them to the minimum proficiency standards for graduation.

But only in rare cases, for reasons such as sickness or emergency absence from the school, did the Training Command permit "hold-overs" during flight training. It was necessary to hold rigidly to the time limit allotted for each phase of training and no provision was made for the slow learner except elimination. Irrespective of weather or individual aptitude, each class had to be graduated on schedule and had also to meet a minimum standard of proficiency. This created a very difficult task for supervisory personnel who were held responsible for meeting the scheduled deadlines. One ornament almost universally found in offices of Directors of Flying was a chart showing the progress of the current classes toward meeting the goal of the prescribed number of hours of flying within the time limits allowed. Since "hours flown" was a more obvious record than level of student proficiency, it consequently received greater emphasis in arranging schedules and revising the curriculum.

Regular Rate of Class Advancement

During the early part of the war each phase of training required 8 weeks; this was later extended to 9 weeks and finally, in 1944, to a 10 weeks period.

In Preflight the students were assigned a class code which they kept throughout the rest of their training. Class 43-B, for instance, was the second graduating class in 1943. In order to utilize the training facilities most efficiently, two classes were in training at any one time at each school. For the first five weeks the students were lower classmen, then moved up to upper-class status when the new lower class entered the school. Classes were graduated from each phase of training every five weeks.

Quota Requirements *v.* the Individual Student

The quota system, which frequently reflected the somewhat unpredictable combat requirements, often put local supervisory personnel in a "straight-jacket" to provide the number of students called for by higher headquarters. For example, often times the Basic schools would not know until nearly the last week before graduation how many of their students were to be assigned to Advanced single-engine training and how many to Advanced two-engine schools. The result was to minimize weight given to the preference and qualifications of the individual student.

PROBLEM OF STANDARDIZING PILOT TRAINING

Factors Producing Variability

Rapid Expansion.—"From 1 October 1931 to 1 January 1939 not a single new training station had been established. On the latter date, the United States had but two air school training stations: Kelly and Randolph. It was therefore on these two fields alone that the great expansion program for pilot training was to be started".[10] These two schools were expanded into a pilot training organization consisting of well over 100 schools. Table 2.1 shows the increase in number of pilots graduated each year from 1937 through 1944. During these eight years there was an increase of more than forty-four *thousand* percent. In addition to handling this expanded load, it was necessary to assign many of the original, experienced pilots to tactical groups on the fighting fronts.

TABLE 2.1.—*Number of pilot graduates from 1937 through 1944*

Year	Graduates	Year	Graduates
1937	186	1941	7,244
1938	301	1942	28,782
1939	696	1943	63,399
1940	1,786	1944	82,487

In the early days of the expansion, the Commanding Officer of a new school was given an unprecedented load of students and told to go ahead to turn out the best pilots he could by whatever means and resources he could obtain. After each Commanding Officer was successful in solving his own problems in his own way, he was reluctant to change his training procedures without good and sufficient reasons. Thus, the autonomy which was necessary during the emergency expansion when no trained personnel could be spared for standardizing the different schools, resulted in the growth of different local traditions which had a strong tendency to persist.

Procurement difficulties resulted in differences between schools and the three regional commands. For a considerable time three different Primary trainers were in use throughout the Training Command and sometimes two different planes were used at the same school. As many as six different types of planes have been used at the Basic level, and the same kind of variation occured in Advanced two-engine schools. Even where the same type of plane was used, differences existed in such specific details as type of propeller, type of instrument hoods, and kind of interphone system. Though one type of equipment might have been recognized as being superior, procurement difficulties resulted in frequent delays in obtaining a supply which was adequate for the entire Training Command.

Geographical Differences

It can be noticed from figure 2.2 that most of the Training Command Stations were located in the southern and southwestern states with a good many more in California. Nevertheless, the climate and typical weather conditions varied a good deal among these scattered stations. The Training Command required the schools to maintain their quotas and time limits for training despite variable weather conditions.

[10] History Army Air Forces Central Flying Training Command, Vol. 1, 1 Jan. 1939-7 Dec. 1941, pp. 19-20.

During periods of poor flying weather, flight instruction was massed into those few favorable days which did occur. Differences in elevation, and hence in the lifting power of the air, from school to school introduced an additional variable into the performance of the airplane.

Temporal Changes in the Training Routine

The necessity of meeting the fluctuating demands of the combat units fighting on many diverse and rapidly changing fronts, required a continuously changing training program to provide the best qualified combat pilots. The pressure resulting from this type of variability was more apparent at the Advanced levels than in the Primary and Basic schools. An even more important factor producing changes was the great increase in the knowledge of how to fly and teach flying.

Examples of Variability

A good example of variability between schools appeared when the Pilot Project was conducting a large-scale study at all schools in the Training Command to determine the effect of an extra five weeks of training (see chapter 10). When the training supervisors from each of the two-engine schools were called together for a briefing conference, a good deal of discussion occurred as to the specific details of manner of performing the check-ride maneuvers. Most of these men seemed to be surprised at the extent of the differences between stations on some of the traditional and fundamental maneuvers. While these differences probably had no effect on the over-all effectiveness of the Advanced school students, they made it difficult to establish uniform testing conditions at all of the two-engine schools. For example, some schools used full-flaps and others half-flaps during landings. Some schools prohibited turning in the direction of the dead engine during single-engine procedure while others made such turns routinely. Some schools used a canvas hood during instrument maneuvers while others used colored glass shields and goggles. At some stations a good deal of practice was given on the power-off approach and landing maneuver while at others only demonstration rides were given.

Some schools gave instrument training in a separate instrument-training squadron while others believed that the most efficient method was to have the same instructor give both contact and instrument instruction. The distribution of instrument instruction also varied from school to school; some spread this training through the entire 10 weeks, while others concentrated instrument training during the first 5 weeks, or the middle 5 to 8 weeks of training.

Grading and standards for elimination have always been a difficult problem to standardize. One Commanding General was rather alarmed at the high elimination rate in the Primary schools in his regional command and ordered all Primary schools with elimination rates above the average of the command to reply by letter of indorsement giving full explanation. In another command, a directive was issued prohibiting any advanced school from graduating students "below average" in over-all pilot proficiency. Since the quota demands for the schools remained the same, the only alternative was the elimination of all "below average" grades and the consequent lessening of the discrimination shown by the flying ability ratings of the graduating students. In another instance, a directive was issued to the effect that no student would be given any idea of his progress during training. It was, of course, practically impossible to comply with this directive which was later rescinded.

Unifying Agencies

Increased Control by Headquarters.—The critical shortage of experienced training personnel needed for assignment to headquarters units required that nearly all of their time be spent with such problems as: procurement, creation of new schools, and preparation of administrative directives. As more qualified officers, particularly pilots familiar with the new training problems, became available for headquarters assignments, the Training Command increased its central, unifying control over the entire program.

Flying Training Wings.—The Flying Training Wings in each of the regional Flying Training Commands helped to standardize training by inspecting the administrative and operational procedures in all schools at a given level of training.

Instructors Schools.—At nearly all stations local instructors schools were established to help standardize the flying performance of the transfers and the new graduates being retained as instructors. During early 1942 instructors schools were established to serve each of the three flying training commands as an expansion of the program followed by the local stations. The first classes sent to these schools were carefully selected and included a large portion of supervisory personnel from all stations within the commands. Not only did this procedure provide for a rapid dissemination of the instructor school methods to men in key positions but it also added prestige value to pilots selected for training in subsequent classes.

In March 1943 the regional command schools were consolidated to form the Central Instructors School (Pilot) at Randolph Field, Tex. However, many of the local schools continued until the end of the war. The Central Instructor School (Pilot) drew personnel from all three commands and was staffed by an exceptionally able and experienced group of instructors. This new school continued to emphasize precision standardized contact flying rather than pedagogical technique.

CIS never had the authority to conduct inspecting tours through the Training Command to determine the extent of uniformity of teaching. Its influence was accomplished through its advisory function, preparation of manuals, and by its graduates who were sent to all schools in the Command.

The Central Instructors School (Instrument Pilot) was established at Bryan, Tex., for the purpose of training and standardizing instrument instructors. Only rated pilots were sent to this school, and after completing this course they returned to their home stations as instrument instructors or members of the instrument standardization board. It was the goal of the Training Command to have every person functioning in either of these capacities a Bryan graduate. In general, the Bryan instrument school performed the same function for instrument flying as the Central Instructors School at Randolph Field did for contact flying. The manuals for instrument training were prepared by Bryan personnel who were also responsible for the preparation of the various checks given at different times during instrument training.

Standardization Boards

Early in 1942 the Commanding Officers at most schools designated a small group, usually four or five, of the more experienced and expert instructors to be members of the Station Standardization Board. In the beginning these boards were identical with the local instructors schools. Gradually their function became more specialized and their primary responsibility became the making of periodic flight checks with both students and instructors to help maintain uniform methods of performing the fundamental training maneuvers.

Later Wing standardization boards were established to help reduce the differences between schools at the same level of training. Standardization boards (called Advisory Training Boards) were also established in the Central Instructors Schools and were perhaps the greatest single source of expert pilot advisors for the Pilot Project. They were particularly helpful in designing objective measures which could most readily be adapted to conditions in the various cooperating schools.

Training Manuals

The original technical manuals published by the Air Corps were poorly adapted for use in the expanded modern training program. The Central Instructors Schools were given the job of revising the prewar manuals in collaboration with such special groups as the Office of Flying Safety, the Visual Training Aids Department, and the Film Strip Unit. These new manuals were generally accepted as being an outstanding contribution to the training program and were used throughout the Training Command. The descriptions and explanations of the fundamental training maneuvers functioned as one of the strongest standardizing factors in the Training Command. The Ground Training Technical Advisory Department at Randolph Field played a similar role in the preparation and supervision of the ground school curriculum.

Gradual Diffusion of Experienced Training Command Personnel

In addition to the formal agencies operating to standardize the pilot training program, an additional factor played an increasingly important role as time progressed. The continuous process of transferring instructor and supervisory personnel from one station to another and to headquarters units tended to produce a synthesis of divergent opinions and methods of training.

Degree of Uniformity in Pilot Training at the End of the War

Toward the end of the war the pilot training program had reached its highest level of uniformity and efficiency of training. Wing members were familiar with the variable characteristics of each station, and the local standardization boards were able to make frequent flight checks with each instructor. A progressively larger proportion of the instructors were graduates from the Central Instructors Schools and there had been extensive diffusion of training personnel from station to station. Procurement bottlenecks were a diminishing source of difference between stations. Training manuals were generally accepted and used as guides by most instructors, and finally, the students themselves were a more highly selected and homogeneous group possessing a high level of flying aptitude.

All of these factors operating together resulted in considerable progress toward improving the efficiency and uniformity of training but differences still remained. Although it emphasized standardization, the Training Command recognized that a certain amount of variability was necessary for the growth of new and better ideas.

SUMMARY

This chapter presents a brief summary of those aspects of the organization of the Training Command and of the successive phases of training which are important as a background for understanding the research described in the following chapters.

The Training Command functioned directly under the Headquarters of the Army Air Forces. It was subdivided into three regional Flying Training Commands, which supervised pilot, bombardier, and navigator training. Within each of the Flying Training Commands a system of Wings was established to serve as inspecting and standardizing units for all pilot training schools.

Because of the pilot aptitude tests and the elimination of poor students in earlier phases of training, the population of pilot students represented a restricted range of high talent. The mass production assembly line system of pilot training, and the extreme pressure of war-time expansion limited the freedom of introducing modifications for purposes of research. The rapid growth of the pilot training program, its wide geographic distribution, the impact of new combat requirements, the discovery of new training techniques, and the procurement of different types of equipment produced variability between different schools at the same time and changes in the same school at different times. These variations had to be carefully watched as sources of constant errors in research. The standardization boards and other agencies, which were created to disseminate improved techniques and to reduce variability, were valuable sources of technical assistance in conducting research.

CHAPTER THREE

Analysis of the Pilot's Task

Capt. Richard P. Youtz and Capt. Stanford C. Ericksen

INTRODUCTION

The purpose of this chapter is to describe the task of learning how to become a military pilot. The first surveys of the task of learning to fly were made in order to provide a basis for developing pilot aptitude tests. In the first survey, written board reports were examined and the different types of reasons for eliminating students from training were tabulated; then experienced instructors and students who had just been eliminated were interviewed. Much later, some aviation psychologists were able to secure a certain amount of flying training. In connection with developing objective measures of flying skill, the members of the Pilot Project made detailed studies of relevant aspects of the pilot's task. For example, they performed informal experiments to determine the effect of various deviations from the prescribed technique of spin recovery upon the amount of altitude lost and the length of time required for recovery. It was necessary to perform such studies in order to determine which aspects of various maneuvers were most relevant and what the relationship was between the technique of the pilot and objective measures of different aspects of the performance of the airplane.

Flying is a complex skill. Our scientific language does not yet contain the words and concepts for summarizing briefly and accurately the demands of such a skill. Any attempt to describe this task, therefore, confronts the dilemma of either describing an exceedingly long series of specific details or making broad generalizations which are quite inexact. Until the fundamental knowledge of the structure of the human abilities involved in complex psychomotor skills has been considerably increased, psychologists conducting research on flying will probably need to acquire by direct experience a more intimate knowledge than can be presented in words.

Although the ultimate goal of all military flying training was operational flying, the descriptions in this chapter will be confined to the type of flying with which the authors are familiar, that done in the Training Command. It should be noted, however, that because the ultimate goal was combat, military flying training was different from civilian. While the civilian student pilot flew airplanes that were designed for safety, economy, and ease of operation, the military pilot began and continued his training on airplanes that were much more powerful, had a higher wing-loading, and were designed to prepare the student for the maximum-performance planes used in combat.

The civilian student pilot ordinarily began his training in an airplane of 60–75 hp., and one that was easy to land. The military student pilot started his training in an airplane of 220 hp., and one that was designed to train the student in the techniques of landing much more highly powered and maneuverable combat airplanes. The military student pilot was under pressure to learn rapidly and thoroughly in a relatively short period of time, while the civilian student might fly or not as he pleased and, before soloing or going on to any succeeding stage of performance, could take as much dual flight instruction as he could afford. Nevertheless, the fundamental elements of flying skill and the techniques and principles of take-offs and landings, climbs, glides, turns, and straight-and-level flight were basically similar for both civilian and military flying.

The main procedure in this chapter will be to select representative tasks in flying, to describe their salient points, and then to summarize the hypotheses concerning the necessary skills. The chapter is divided into two parts: I. The Fundamental Elements of Flying Skill; and II. Specialized Areas of Military Flying Training. Since the fundamental elements were taught almost entirely at the Primary level, the first section will deal with activities in Primary school. The specialized areas of military training are found at several higher levels of training, so that the second section will deal with areas covered in Basic training, in single-engine and two-engine Advanced training, and in Transition training.

Part I. The Fundamental Elements of Flying Skill

Capt. Richard P. Youtz[11]

THE GOAL OF AAF PRIMARY PILOT TRAINING

In Primary pilot training the student was first introduced to the airplane, took his initial flight, and was given training in all of the fundamentals of contact flying. Although it might be said that the goal of any flight is a safe take-off, effective performance of mission, and a safe return and landing at the desired field, in military flying much greater precision was required in the execution of all maneuvers. Students who would have made safe civilian pilots were eliminated if they did not exhibit the precision which would be necessary in flying close formation, making accurate bomb-runs, or aiming their airplane so that bullets from its fixed guns would hit the target.

[11] S/Sgt. Maurice F. Connery of the Pilot Project assisted in the preparation of material for Part I. Maj. H. H. Culler, Chairman of the Primary Advisory Training Board and Commander of the Primary Department of the Pilot Central Instructors School, and Maj. B. F. Meyer, member of the Primary Advisory Training Board, read and criticized Part I for technical correctness.

GENERAL DIFFICULTIES IN FLYING DERIVED FROM CHARACTERISTICS OF THE AIR

Compared with driving a car along a road, flying an airplane from take-off through landing presents a number of new problems which grow out of the characteristics of the air.

Air Is Transparent and Without Reference Points

Unlike automobile driving in which the driver can see the sides of the road and the line in the center of the road with relative ease, the pilot of the airplane must perform most of his maneuvers using only distant reference points, such as the horizon, or large features of the landscape far away. Because of the remoteness of these reference points the amount of relative movement is very small; the airplane seems to stand still and hang in the sky until the student has learned to make the necessary finer perceptions and to evaluate them in terms of the actual speed of the airplane. The importance of these finer perceptions of far-away reference points is strongly demonstrated to the student, particularly in cross-country flying and in simulated forced landings where the student must visualize his three-dimensional flight path through transparent space.

During take-offs and landings the reference points are nearby but the airplane is going so fast that the immediate surroundings are blurred. The pilot must choose reference points that are not so near as to be blurred or so far away as to be irrelevant to the take-off or landing.

The Air Is a Fluid, Elastic Medium

Since the air is fluid, it is subject to currents and turbulence. The airplane drifts with the currents or winds, and this drift must be taken into account in any maneuver related to reference points on the ground, such as take-off, landing, cross country, etc. Perceiving slight tendencies to drift, and differentiating the relative movements in the forward area of vision which are caused by drift from those caused by slight turns, are new and difficult types of discrimination for the average student. Since the air is turbulent, the airplane may be shaken about and thrown off course and the perception of drift must be made against a background of unsteady movement. Turbulence may also cause airsickness.

Since the air is elastic, the controls of an airplane have a less positive effect than the controls of an automobile in which the deflection of the steering wheel determines directly the radius of turn. Under certain conditions, there may be a temporal lag before the pressures on the controls of an airplane take effect or it may "mush" through the air in much the same way that a car makes a skidding turn on a slippery road.

An airplane is guided by control pressures which in turn cause the control surfaces of the ailerons, rudder, and elevators to exert pressures on the stream of air passing by them. This deflects the airplane in one direction or another. At high air speed a slight movement of the controls will produce a considerable effect; when the air speed is lower, as in landing, much more extensive movements are required to produce the same result. Although the extent of movement differs greatly, the amount of pressure exerted on the controls in producing these movements remains approximately the same. The student, therefore, must learn to control his aircraft in terms of pressures instead of movements. This change is disconcerting to some students.

Flying Is Paced

An airplane must maintain its speed or fall to the ground. In fog or darkness the airplane must continue at the same pace and in many situations the pilot is forced to make rapid decisions because of the necessity for maintaining speed. Safety in flight is usually obtained by flying "high and fast" rather than "low and slow." This is unlike automobile driving in which the vehicle may be slowed down or stopped at any time to give the driver a chance to consider what to do next. The pilot is always forced to pay attention to speed and add this variable to the complex of stimuli that demand his attention.

Steering in Flight Is in Three Dimensions

Besides steering the airplane to the right or to the left, as an automobile driver has to, the student pilot must point the nose up or down in order to maintain the desired altitude and must also keep the wings level or properly banked at all times. Thus the pilot has three kinds of steering to do compared with the automobile driver's one. Not only does the pilot have two more kinds of steering but the difficulties of steering are multiplied many times because of the interrelationship between the different kinds. In making turns it is necessary to bank the plane so that the wings are at an angle with the horizon. In doing this the sideward pressure on the stick must be matched with an appropriate pressure on the rudder in order to avoid skidding or slipping. Since banking the wings reduces the lift, it is also necessary to pull back on the stick, increasing the angle of attack in order to maintain a constant altitude.

CRUCIAL ASPECTS OF PRIMARY FLYING

This section will describe those maneuvers which best illustrate key difficulties in learning to fly. It is not the purpose of this section to give a complete description of all of the maneuvers in Primary training; it will be limited to those portions of the task which appear to be most important in differentiating the better from the poorer students.

Since approximately 90 percent of the students went to Primary school without previous flying experience, they had to become adjusted to the new sensations encountered in flight. Most of the students were well motivated by one or more of a number of factors such as interest in flying, a desire to contribute to the war effort, a desire for status as a commissioned pilot with its attendant increase in pay and privileges, or sometimes only a desire to obtain the better living conditions that prevailed in the Air Forces as compared with other branches of the service. Although they were well motivated, many students had an apprehension concerning flight. A majority of students were tense during their first few flights and this tension had to be overcome before they could coordinate as smoothly as necessary and could think clearly while flying.

As the first flight, each new student was given a 20-minute orientation tour of the area. Most of the flying was straight and level and the instructor pointed out landmarks in a calm tone. Even so, it was a tension-producing experience for the student to fly through the air with the roar of the engine in his ears. When the plane was banked, many new students leaned to one side trying to remain vertical to the ground instead of remaining in the same position relative to the airplane. In slight down-drafts the airplane tended to fall away from under the student and he felt a loss of support even though he was held firmly in his seat by the safety belt. The students had to become used to these sensations and had to gain confidence in the airplane and the instructor.

The apprehension concerning flight was in most cases soon overcome. However, new maneuvers were introduced rapidly and the student had to work hard to assimilate the material. Officers from Psychological Research Unit No. 1 who spent six weeks at an army Primary school taking flight instruction and observing the other students, reported:

"The rapid pace of the first 15 to 20 hours of instruction leaves one with the feeling of insufficient understanding of what one is doing and of insufficient practice in doing it. Forty minutes of dual flight daily (the scheduled amount) actually means 25 or 30 minutes in the air, and about 20 minutes outside the traffic pattern: very little time for practicing old maneuvers and learning the new ones which are introduced every two or three days. It is after about 15 such flights with days full of other things that one must solo."

Some students were so disturbed by flying and by the rapid pace of training that they requested elimination. Others became so tense that they were unable to make the necessary fine kinesthetic discriminations and could not coordinate properly. Still other students were particularly anxious during certain maneuvers so that they forgot what they had been told or were unable to think clearly, look around and take into account all of the necessary factors. In a few cases students were so uneasy about certain maneuvers that they did not practice them during solo flights and hence fell behind their classmates who did

703322—47——4

practice. Thus poor coordination, faulty memory, poor headwork, or slow progress were in some cases symptoms of emotional disturbance rather than lack of aptitude. On the other hand, students who because of low aptitude found it difficult to control the airplane, were likely to become emotionally upset.

The Four Fundamentals of Flight: Climbs, Glides, Turns, Straight-and-Level

It was frequently said that the four fundamental maneuvers were climbs, glides, turns, and straight-and-level flight, and that all other maneuvers were made up of combinations of these four. There were a few exceptions to this statement but a student who could perform these four basic maneuvers well was not likely to have trouble with other, more complex maneuvers. An important and representative situation where these four aspects of flying were all necessary was in the practice of take-off and landing. Therefore, a description of the take-off, traffic pattern, and landing gives a fair sample of the problems, perceptions, and skills necessary for the student pilot in these important maneuvers. The students had to become proficient in take-offs and landings before they could solo. A large proportion of eliminations occurred between the 10th and 17th hours of flying training because the students could not learn to do the take-offs and landings well enough to solo.

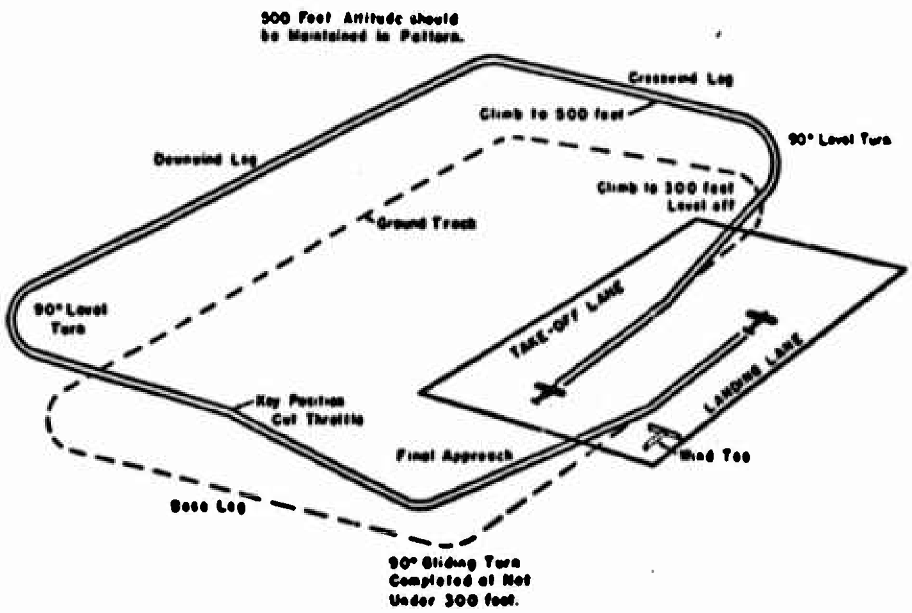
The rectangular traffic pattern used for take-offs and landings is illustrated in figure 3.1. In flying this pattern the airplane took off upwind, climbed to an altitude of 300 feet, leveled off, made a level 90° turn to the left, climbed on up to traffic altitude of 500 feet above the ground, then turned 90° to the left again outside the boundary of the field and flew downwind along its edge (downwind leg) until reaching a point beyond the downwind end of the field. Then another 90° turn to the left was made. The airplane then flew parallel to the end of the field (base leg) until time to turn left again and approach the field in a glide for the landing. In order to make the airplane perform this apparently simple rectangular pattern with the necessary degree of precision and safety, the pilot had to make rapidly a number of fine perceptions, make rapid judgments of relative speed and distance, and exert just the correct control pressures in the right sequence at the right time. This will be described in detail in the following paragraphs.

Take-off and climb to first turn. After the airplane had been lined up for the take-off and the proper checks had been made, the student advanced the throttle smoothly and continuously to the full open position. As the airplane gathered speed the student counteracted its tendency to turn to the left[12] by applying right rudder and

[12] This left-turning tendency of the plane during take-off was one of the manifestations of "torque." For definition of this and similar terms, see Glossary.

Fig. 3.1

TRAFFIC PATTERN USED FOR TAKE-OFF AND LANDING AT AAF PRIMARY SCHOOLS



When not remaining in the Traffic Pattern, the Student left the crosswind leg at a 45° angle and entered the middle third of the downwind leg at a 45° angle.

simultaneously judged the correct moment to allow the airplane to leave the ground.

After the airplane was off the ground 20 feet or so and had gained flying speed, the student throttled back slightly and held it in a climbing attitude as near to an airspeed of 70 m. p. h. as he could. All the while, he had to watch for other airplanes in the pattern and make sure that his climbing path was a projection of his take-off run. This last was particularly necessary because, if students did not fly parallel paths on take-off, there was real danger of mid-air collision. The straight path was difficult during climbs because "torque" produced a left-turning tendency during the climb as well as during the take-off and the airplane had a tendency to drift if the take-off was not directly into the wind. The direction of take-off and landing was determined in Primary school by the wind tee, which was an indicator that could be set on one of the eight points of the compass and was pointed as nearly upwind as possible. Since the tee had only eight settings, it was not always pointing directly upwind and the plane might have a tendency to drift sideways. This had to be counteracted by a slight turn into the wind, or "crab." Since very few Primary schools had runways for student use, the only effective way for the student to check on the direction of his climbing path was to look back at the field and see the wind tee. To do this he had to turn around and look long enough to detect any drifting tendency and still keep a sharp lookout for other nearby Primary trainers.

It was important that the student throttle back and maintain the plane at the most efficient climbing speed of 70 m. p. h. If the speed rose much above this the airplane would not be at its most efficient climbing speed and would not be able to reach the traffic altitude of 500 feet at the proper time. If the air speed fell too much below 70 m. p. h. there was danger of stalling the airplane at such a low altitude that it would crash.

It can be seen that the student could not spend much time watching the air speed indicator. It was necessary for him to train his perception so that he could estimate fairly well the air speed of the airplane. The student's cues for keeping the airplane at the correct airspeed were: the angle between the nose and the horizon; the whistle of the struts; the sound of the engine; the stiffness or "mushiness" of the controls; and his feeling of sustentation. In this situation, the perceptual aspects of the task of flying were considerably more difficult than the motor activities necessary to correct any deviations from the desired air speed. At a constant power-setting in the climb, the speed of the plane could be increased by a slight relaxation of back pressure on the stick. The crucial aspect was the perception of the need for such corrective actions.

The level 90° turn.—After climbing to an altitude of exactly 300 feet, the student leveled off and made a level, 90° turn to the left. In a turn, the stick and rudder pressures had to be coordinated so that the plane did not slip or skid during the turn. In a medium-bank turn to left the slight leftward pressure on the stick was coordinated with just enough pressure on the left rudder so that the plane came around smoothly without the nose going up or down on the horizon.

Too much rudder pressure would make the airplane skid around toward the outside of the turn; too little rudder pressure would not bring it around quickly enough and make it slip to the inside of the turn. This "coordination" was strongly emphasized by instructors principally because it indicated whether the student had the ability to perceive small irregularities in the airplane's flight.

The coordinations necessary for rolling into a turn and rolling back to level flight were important in Primary training because during his later training the student might have to fire at a target with machine guns that were fixed to the nose or wings of his airplane. In fixed gunnery of this sort excellent coordination was necessary so that the bullets would strike the place that the student was aiming at with his gun sights. A lack of coordination produced a slip or skid that would cause the bullets to go wide of the mark. Good coordination also had a more immediate application because a student who held too much bottom rudder in a gliding turn shortly before landing was likely to put the plane into an unexpected spin close to the ground with insufficient altitude for a safe recovery.

The principal cue for this coordination was the feeling of being swayed toward one side or the other in a slip or skid in much the same way that a person feels swayed or pulled toward the outside of a turn when riding in an automobile. When an airplane was in a perfectly coordinated turn, the pilot felt no tendency to sway to either side.

After the student had coordinated the stick and rudder pressures to produce the degree of bank that was necessary for a turn, the controls were neutralized. The airplane was then kept at the right speed and altitude principally with the elevators, which were controlled by back and forward pressures on the stick. Altitude control was more difficult in turns than in straight flight because the angle of bank reduced the effective lift and because the airplane would lose altitude by slipping if the bank was not properly coordinated with the rate of turn.

If the student did not hold enough back pressure on the stick, he would lose altitude rapidly. If the student held too much back pressure he would gain altitude, and with still more back pressure the airplane was likely to stall and fall irregularly out of the turn. Thus, stick pressure and altitude control in the turn were extremely important. As described more completely in Chapter 6 on the Development of Objective Measures of Flying Skill at the Primary Level, altitude control was found to be one of the best measures in differentiating between students with 15 hours of training and a greater number of hours of training.

When the student used too much back pressure he could tell that he was gaining altitude by the increased feeling of sustentation or lift. When the airplane approached a stall, there were a number of signs of the approaching stall. One of them was increasing looseness of the controls. Just before the stall occurred there was a characteristic shudder and "buffeting" as the air flow over the wings became irregular.

If the student acted quickly as soon as he felt the airplane shudder at the beginning of a stall, he could usually increase power and move the stick forward in time so that the airplane would not stall completely and fall out of the turn. This, however, required instant recognition of the shudder and immediate movement or pressure on the stick.

Crosswind leg, downwind leg, and base leg.—After taking off, climbing to 300 feet and making a level 90° turn to the left, the student flew the airplane crosswind until outside the boundary of the field, climbing to the traffic altitude of 500 feet. The climb was the same as the climb previously described, but in this case the student had to crab into the wind so that on this crosswind leg he would not be blown back over the field.

After completing the crosswind leg the student pilot turned the airplane 90° to the left and flew downwind outside the edge of the field on a path parallel but in the opposite direction to the one on which he took off and climbed. This was the downwind leg and in this leg the student flew straight and level, watched out for other airplanes and selected a landing spot. Straight-and-level flight does not consist of the absence of change but consists of a series of small corrections of altitude, direction, and ground track. Much as an automobile driver continuously makes small corrections with the wheel in order to keep his car in the correct lane on a road, the pilot of an airplane steers in three dimensions, corrects for possible drift, and keeps a sharp lookout in all directions since other airplanes may approach him from any angle or direction.

In straight-and-level flight the pilot's attention typically went through the following sequence: The pilot turned his head and looked to the rear at the left of the airplane, with his field of vision including the air above and below. He then looked out along the left wing to see whether it was level and in the correct position on the horizon. Following this he looked at the instrument panel and checked on airspeed, altitude, and engine rpm with occasional attention to fuel and temperature gauges. If, for example, all was in order except for altitude, he again looked out over the nose while making a slight correction in stick pressure. His glance then traveled around to the right, checking on the right wing, and then to the right rear, above and below. He then started back from the right rear to the right wing, to the instruments, and at that time checked the altimeter to see if the airplane was coming back to the correct altitude. The student did not keep watching the altimeter during the correction in altitude. He knew that if he did that, some other aspect of the airplane's flight would deviate from the desired.

Final Approach and landing.—The student flew on the downwind leg until he was about a quarter of a mile beyond the end of the landing field, then made a 90° turn to the left, and flew crosswind until ready to turn in toward the field for the final approach and landing. The position of the base leg was varied depending upon the wind velocity. With a higher wind the base leg was put close to the field and with no wind the base leg was approximately a quarter of a mile from the field. It was necessary for the student to crab into the wind on the base leg in order to maintain a ground track that was directly crosswind. On the base leg, the student picked out the spot on the field where he planned to land and at a place which was at a 45° angle from the desired landing spot he closed the throttle, held the airplane in a cruising attitude until it slowed to gliding speed and then nosed it down into a glide. Placing the base leg the proper distance from the field and cutting the throttle at the proper point required good judgment of spatial relations.

The desired gliding speed for the Primary trainer was 75 to 80 m. p. h., which gave a convenient rate of descent and a safe margin above stalling speed. In the glide, as in the climb, the student had to be careful not to let the air speed get too low because of the danger of a stall. If, on the other hand, the student let the air speed get too high, the airplane would not land upon reaching the ground, and the student would have to skim along the surface for an excessive distance. The student held the correct gliding attitude and air speed partly by reference to the air speed indicator but mostly by other cues, because in this case he also had to watch for other airplanes as well as to make plans for his landing. The cues for the correct gliding attitude and air speed were the position of the nose relative to the horizon, the sound of air hissing over the struts and fabric, and the feel of the controls. The ability to perceive and correctly interpret these cues was considered most important in students because it determined to a great extent the safety of the student during his final approach and landing and because a constant gliding speed was essential to good estimation of the height-distance relationship.

On the final approach to the landing field, the student had to estimate his point of landing and make any corrections necessary for "undershooting" or "overshooting." If he were going to "undershoot," i. e., land short of the field, he opened the throttle a little and the airplane gained enough speed to carry him into the field. If he were going to "overshoot," i. e., land beyond the middle of the field, it was necessary for him to open the throttle wide and gain altitude, fly over the field, and repeat the traffic pattern before making another try at landing.

The problem of estimating the probable point of landing was a complex one. If it is assumed that the airplane was in the desired 75–80 m. p. h. glide, the expected point of landing was the place on the ground toward which the glide path of the airplane was directed. However, this was not easily determined because the nose was not pointed directly at this place on the ground but was pointed at a spot farther on because of the nose-high gliding attitude of the airplane. Many students learned to estimate the probable spot of landing by observing during the glide the spot which had no tendency to move toward the edge of the visual field. Since the student was gliding toward the ground from above, all objects were rapidly becoming larger and those not on a projection of the glide path were moving out toward the edge of the visual field. Other students apparently did not use this kind of perception but made their judgments from the estimated velocity of the wind, the position of the base leg, and the altitude of the airplane just before the final turn onto the approach leg.

On the final approach as the airplane came near the ground, the student's problem was to decide at just what altitude above the ground to start increasing the back-pressure on the stick so that the airplane would not fly into the ground but would approach it more and more gradually. This change from the normal gliding attitude and air speed was called "breaking the glide." The air speed of the airplane was gradually reduced as it approached the gound and ideally the student brought the airplane closer and closer to a full stall as it approached the ground, until finally the airplane was stalled with all 3 wheels no more than 1 or 2 feet above the ground. Then, as the airplane stalled, it settled down to a perfect 3-point landing. In actual practice the perfect landing just described was the exception rather than the rule.

Since students often made errors and the air was usually moving and more or less gusty close to the ground most landings deviated from the perfect one to a greater or lesser degree.

One of the most difficult landing problems for students was the landing in which the student came down parallel with the tee, but the tee-setting was not exactly into the wind so that it blew a little across the path of the airplane. In order to keep from drifting, the student had to slip slightly into the wind, to crab slightly into the wind, or a combination of both. This was done during the approach glide and one of the greatest problems for the student was to determine just how much slip or crab to use. On the approach glide it was sometimes difficult for the student to discriminate between a slight turn and a small amount of drift. If he was able to do this and to crab just the right amount so that the ground track was parallel with the tee, he had to remove the crab at the last moment before landing or the airplane would land slightly sideways. If the airplane landed this way, there was danger that the landing gear would be damaged or that it would be difficult to avoid a ground-loop. On the other hand, if the crab into the cross wind was removed too long before the wheels actually touched the ground, the slight cross wind would then have time to drift the airplane away from its correct ground track and the airplane would be moving with a slight sideward component which would again tend to damage the landing gear or produce a ground-loop. Thus, it was necessary for the student to make accurate estimations of the degree of crab necessary to compensate for the slight cross wind, estimate his height above the ground, remove the crab neither too early nor too late in the landing sequence, and, in spite of the "mushiness" of the controls at this low speed, to turn the plane straight at the last moment and land. This was a difficult skill for most students, involving as it did accurate perceptions and manipulations, and was one of the maneuvers in which the student had to be proficient before he could solo.

After the student had broken the glide and the airplane was approaching the ground, the student's problem was to estimate the height of the wheels above the ground, to bring the airplane into a 3-point stalled position before letting it touch the ground, and also not to stall it so high that it would drop in with a jolt. The principal cue for the correct landing attitude of the airplane was the angle between the nose of the airplane and the horizon, or in areas where the horizon was indistinct or irregular, the angle between the longitudinal axis of the airplane and the surface of the ground. If this correct attitude were not attained, the wheels of the airplane were likely to strike the ground first, and produce a bounce in which the airplane rose into the air 5 to 15 feet. The student then had to act quickly, opening the throttle briefly so that the airplane would not drop into the ground from this increased height. Some students were "ground shy" and rarely bounced the wheels on the ground, but rather stalled the airplane too high and allowed it to drop several feet to the ground. This had to be avoided because of the danger of breaking the landing gear.

Estimating the distance of the wheels above the ground was one of the things difficult to learn. The student could not see the wheels, and the ground was going by at 60 to 70 m. p. h. The ground nearby was blurred because of the speed. The student had to look farther away so that objects would not be blurred but not so far away that changes in the visual angle subtended would be too small to make perception of height above the ground possible. At the same time that he was making these perceptions of height it was necessary for the student to hold the airplane straight and to keep the wings level, so that he had several aspects to watch at the same time. The degree of back pressure on the stick, for instance, was crucial. It had to be brought back slowly enough so that the airplane would not zoom back up into the air, but also rapidly enough so that the airplane would attain the correct stalling attitude and speed before touching the ground.

Even after the airplane was set down in a three-point attitude, the landing was not completed until it came to a full stop. Just after touching the ground the airplane was still going 50–55 m. p. h. The student's task was to keep rolling straight ahead until the airplane had slowed down to the taxiing speed of about 4 m. p. h. The design of the Primary training plane intentionally made this a difficult task so that the student would acquire a skill required later in tactical equipment. The front wheels on the Primary trainer were set fairly close together and well in front of the center of gravity so that there was an unstable equilibrium and without considerable care and skill the student was likely to "ground-loop" the plane. In a ground-loop the airplane turned sharply to one side or the other. Centrifugal force was then likely to tip it over, damaging a wing tip and sometimes throwing the airplane up onto its nose.

The students were advised by their instructors that the best way to prevent ground loops was to stop them before they started. This meant that the student had to observe very slight deviations of the nose to one side or the other of the straight path and make the appropriate corrective slight rudder pressures. If the nose of the airplane was allowed to swing more than slightly to one side or the other, the technique of correction was still more difficult. The student could not apply corrective pressure to the rudder and then hold that pressure until the nose of the airplane came back to the straight path. If the student did this, the airplane then went into a more violent turn in the opposite direction. The proper technique for stopping an incipient ground loop was to apply the proper amount of opposite rudder pressure for a very brief period of time, anticipating the effect so that the airplane swung back to the correct position but not beyond the correct position. The difficulty of these perceptual and motor skills was one of the reasons that students did not solo sooner. Fine adjustive reactions with the feet and legs were not part of the student's previous activities. They could be learned but it was difficult and took time.

The relative frequency of errors in the major areas of difficulty in landing an airplane was determined in a study done by personnel at Psychological Research Unit No. 3.[8] On 199 landings performed by 88 Primary students the major areas of difficulty and the percentage of times errors occurred in these areas were as follows:

a. Not stalling out correctly (76 percent).

b. Not keeping headed straight on the ground after landing (58 percent).

c. Not breaking the glide at the correct height (53 percent).

d. Not keeping headed straight on the approach leg (51 percent).

e. Not maintaining proper gliding speed in the approach (50 percent).

[8] This study was performed by Capt. Stuart W. Cook, T/Sgt. David H. Jenkins, Pvt. Harold H. Kelly and Pvt. Eli A. Lipman.

The take-off, traffic pattern, and landing have been considered in some detail because they contained a wide variety of problems representative of many phases of flying skill. For this same reason the maneuvers were particularly important in the training of the student and were among his principal problems.

To summarize this section, in the take-off, traffic pattern, and landing, the student had a variety of tasks which had to be performed at such a fast pace that before the habits became automatic it was somewhat like juggling too many balls in the air at the same time. It was easy for the student to become confused and "go to pieces," sometimes making mistakes on usually simple tasks. In the short space of 6 to 10 minutes the students had to do the following: from idling speed, advance the throttle smoothly and continuously to full throttle; hold the airplane straight during the take-off run, counteracting the torque-produced, left-turning tendencies; fly the airplane off the ground correctly, neither pulling it off too soon nor holding it on too long; after flying speed was reached, throttle back to a climbing power setting; adjust the throttle-setting and climbing attitude of the airplane so that the airplane climbed at an air speed of 70 m. p. h.; make sure that the climbing path was still parallel with the take-off path and the tee; keep a constant lookout for other airplanes which might also be taking off; watch the altimeter so that he could make a level 90° turn to the left at 300 feet above the ground; continue climbing on the crosswind leg until the traffic altitude of 500 feet had been reached; make another 90° turn to the left onto the downwind leg, keeping a close watch that he was at least 500 feet from any other airplane in traffic pattern; during the downwind leg maintain the correct altitude, heading, engine r. p. m. and airspeed, while taking into account the direction and velocity of the wind in planning how far out from the field to put the base leg and where he should land; make another 90° turn to the left onto the base leg, placing the base leg in such a position that the airplane could then land in the first third of the landing field; crab into the wind enough on the base leg so that the ground track was correctly crosswind; close the throttle at the correct place on the base leg so that landing in the desired place on the field would be possible; continue holding the airplane level after the throttle was cut until the speed had been reduced to the correct 75–80-m. p. h. gliding speed, then nose down into the proper gliding attitude; keep a sharp lookout for other airplanes in the pattern that might be landing; make the 90° left gliding turn toward the field so that the airplane would start its final approach in the desired landing lane; maintain the gliding path on the final approach parallel with the tee, even though the wind might be blowing from a little to one side; maintain the correct 75–80-m. p. h. gliding speed, using cues of "feel of the controls," sound of air hissing over struts and fabric, and attitude of the plane; estimate the probable point of landing and apply enough more throttle if it was not safely inside the field; glide on down, maintaining the correct air speed, until time to "break the glide" and begin the landing; bring the airplane down close to the ground, still maintaining crab if there was a slight crosswind, and taking off the crab at the last moment before the airplane was fully stalled and settled to the ground; and hold the stick full back after landing to keep the tail on the ground, while watching closely in order to prevent groundloops. When all of this had to be done rapidly and accurately during a crowded 6 to 10 minutes, it can be seen that in addition to the perceptual and motor skills involved, the student needed to be able to divide his attention, or at least make rapid shifts of attention with rapid decisions, while keeping a complex of other factors in mind.

Other Maneuvers Taught at Primary School

While many other maneuvers were taught during Primary training, they were for the most part made up of the same fundamental elements of flying skill that the student used in the take-off, traffic pattern, and landing. In aerobatics it was necessary for the student to maintain orientation, to have good timing, and a good "feel of the airplane." However, few students who were good in the other aspects of flying, were eliminated for lack of proficiency in aerobatics. One maneuver, the chandelle, which was a special kind of 180° steep, climbing turn, made more demands on the student than the ordinary 180° steep turn. It was particularly necessary for the student to have a good speed sense, to coordinate properly, and to be able to sense the approach of a stall. Because of this, some students who had only minimum proficiency on earlier maneuvers showed up their weaknesses more clearly in a chandelle and became candidates for elimination. However, there appeared to be no fundamentally new psychological demands on the student in these or other later maneuvers.

INVESTIGATIONS BASED ON ELIMINATION FROM PRIMARY PILOT TRAINING FOR LACK OF FLYING SKILL

The various units in the psychological program carried out a number of investigations which threw light on the pilot's task. Most of these were studies of the reasons for eliminations from Primary pilot training. One series of studies was concerned with the reasons given by flying instructors for eliminating Primary students. These reported reasons were classified and grouped, and a rating scale based on these groupings was studied in relation to subsequent elimination from Primary training. Another larger area of investigation was the one on classification tests, all of which were validated against graduation or elimination from Primary school to determine their value for prediction of success in pilot training. The work on the classification tests is reported in reports No. 4 and No. 5 of this series. They are not discussed in detail in this chapter. A third area concerned with the pilot's task was the development of objective measures of flying skill for the Primary level. This is described in Chapter 6.

Analysis of Reasons for Elimination from Pilot Training

When a student was eliminated from one of the pilot training schools, a summary was prepared of the reasons why he had failed to make successful progress in learning to fly. Field studies showed that these summaries usually were prepared with an eye toward justifying a decision, which had already been made, and staving off criticism in any future investigation. They almost always made a very strong case against the student. Nevertheless, the factors mentioned were probably related to the most common difficulties.

In a study prepared by the Psychological Branch, Research Division, Office of the Air Surgeon, an analysis of these reports was made in order to determine the nature and frequency of various reasons that were given as causes of elimination. A preliminary analysis was made of these reports for 300 students who were eliminated from flying training during the early part of the summer of 1941. On the basis of the categories determined in the preliminary study, a more complete analysis was made of 1,000 additional cadets who were eliminated during the summer and fall of the same year.

The frequency with which various reasons were given for elimination of the 1,000 cadets used in the major study is shown in table 3.1. The data are expressed as the percent of cases out of the total in which at least one comment was made in a given category. If the same reason for elimination was mentioned more than once for a given cadet it was tabulated only once, so that the data show only that a trait was mentioned, and not whether it was mentioned once or more than once for a given man.

In interpreting this table, the fact that a trait is frequently mentioned can safely be taken to mean that it was important, but any comparisons between the number of times different traits are mentioned must be made with reservation. The number of times a trait was mentioned is a function of the broadness of its definition (for example, "Progress in Developing Technique") and of the ease with which the instructor could diagnose it as an underlying cause of difficulty. The significant finding was that such a large variety of traits were mentioned relatively frequently. Frequent mention was made of traits which the investigators grouped into the following broad categories: Coordination and Technique, Alertness and Observation, Intelligence and Judgment, and Personality and Temperament.

As a result of this study a 20-item rating scale was made up on which the instructors in Primary school rated each of their students. This scale (Rating Scale for Aviation Cadets—Form C—Pilot), which had added a general over-all rating to make the number of items 21, was also used as a check list. For those students who were eliminated from pilot training, the flight instructors were directed to indicate each trait the lack of which was important as a cause of elimination. This was done in addition to the regular ratings. Students eliminated for physical deficiency or for administrative reasons were excluded from this study.

In a study performed by the Psychological Section, Office of the Surgeon, Headquarters, Flying Training Command, the frequency with which each of the twenty items was checked was obtained on 1,303 cadets eliminated from Primary pilot training.

TABLE 3.1—*The percentage of a group of one thousand aviation cadets eliminated from primary flying schools in the latter part of 1941 for which deficiencies in these categories were listed as a reason for elimination in the report of the faculty board*

Category	*Percent*
A. INTELLIGENCE AND JUDGMENT	68
1. *Judgment.*—Ability to make sound judgments and choices as to the best thing to do when he is faced with a practical problem in traffic, in making forced landings, and in similar situations	50
2. *Foresight and Planning.*—Ability to plan a series or sequence of maneuvers, plan ahead for landings, plan entry or exit from traffic, and foresee and avoid possible difficulties	38
3. *Memory.*—Ability to remember instructions from day to day, both general explanations and specific, detailed information	24
4. *Comprehension.*—Ability to understand and grasp the meaning of explanations, instructions, and demonstrations, either when they are given orally or in written form	17
B. ALERTNESS AND OBSERVATION	70
5. *Visualization of flight course.*—Ability to "get out of the cockpit" and fly the plane with reference to the horizon and reference points, as shown by the ability to handle ground pattern work, maintain constant altitude, control the direction of the plane, make turns of the desired amount, etc	36
6. *Estimation of speed and distance.*—Ability to make such estimates of speed, distance, and altitude as are required in flying a course, flying in formation, gliding, landing, etc	30
7. *Sense of sustentation.*—Ability to sense support or lack of support of the plane, and thus detect slips, skids, or the approach of a stall	24
8. *Division of attention.*—Ability to remain alert and observant of things around him while flying and at the same time attend to all the necessary details and carry on all the different activities necessary for precision flying	28
9. *Orientation.*—Ability to find his correct geographic position by the use of any available means, such as familiar reference points which are visible on the ground, identification of the area below as it is represented on charts or maps, etc	13
10. *Speed of decision and reaction.*—Ability to think quickly, to make rapid decisions, or to respond with speed and precision when the situation demands	15

TABLE 3.1—*The percentage of a group of one thousand aviation cadets eliminated from primary flying schools in the latter part of 1941 for which deficiences in these categories were listed as a reason for elimination in the report of the faculty board* —Continued

Category	*Percent*
C. COORDINATION AND TECHNIQUE	81
11. *Coordination.*—Ability to apply the correct pressures to the controls either in combination or in the proper sequence and with proper timing, as evidenced by the way in which he uses together all of the controls necessary in flying the plane	58
12. *Appropriateness of controls used.*—Knowledge of what control or combination of controls should be used at any particular time while flying and the ability to respond by operating these controls so as to achieve the desired result in relation to the attitude of the plane	21
13. *Feel of the controls.*—Ability to sense the responsiveness of the plane to control movement, or the effect being produced on the plane by various control pressures and by the ability to detect stiffness or "mushiness" of the controls	2
14. *Smoothness of control movement.*—Ability to operate the controls with smooth, even movements and good touch control, without evidence of roughness, heaviness, or jerkiness	22
15. *Progress in developing technique.*—Ability to learn rapidly the various coordinations of the controls and techniques necessary for flying the plane	54
D. PERSONALITY AND TEMPERAMENT	43
16. *Absence of tenseness.*—Freedom from undue tenseness or rigidness, ability to relax sufficiently while flying	22
17. *Absence of confusion and nervousness.*—Ability to remain cool and collected, and to think and act without interference from anxiety or emotion when faced with emergency or a difficult situation	12
18. *Absence of fear and apprehension.*—Presence of a desirable amount of self-confidence, courage, and aggressiveness, with the absence of timidity, cautiousness, or any strong fear of flying	18
19. *Suitable temperament.*—Possession of a stable and well-balanced temperament, with absence of careless, erratic and shiftless habits	9
20. *Motivation and attitudes.*—Strong interest in aviation and desire to be a military pilot, as shown by his effort, eagerness to learn, willingness to respect flying rules and regulations, and his determination to make good	6

A comparison of the frequency with which each item was checked on the rating-scale check-list with the frequency with which it was mentioned by instructors in the study previously described shows that 6 of the 20 traits had a frequency of 30 percent or greater in both of these studies.

In an earlier study performed by the Field Studies Unit of the Psychological Section, Office of the Surgeon, Hq. AAF Training Command, it was found that when instructors rated their students at the 8–10 hour level each of the above traits had a biserial correlation with graduation-elimination of 0.60 or better, with the exception of Coordination which had a biserial correlation of 0.55.

These six traits and their biserial correlations with Primary graduation-elimination were:

	r_{bis}
1. Judgment	0.66
2. Foresight and planning	.67
3. Visualization of flight course	.73
4. Estimation of speed and distance	.62
5. Coordination	.55
6. Progress in developing technique	.68

Reasons for elimination reported by eliminated cadets.—A total of 150 students who had been eliminated from pilot training were interviewed at Psychological Research Unit No. 3 in order to secure data on the students' impressions regarding factors involved in their failure in training. Although the factors cited by the students were undoubtedly not all of the ones involved in their failure, it seemed reasonable to assume that these factors were valid. That is, a cadet may have failed to recognize some of his problems or mistakes, but when he did recognize a weakness in himself or a problem, it seems likely that he was citing a valid reason or part of a reason for his failure. Of course, the possibility of a large amount of rationalization must be recognized.

In each interview an attempt was made to get an impression of the student's emotional reaction to his elimination. Although emotionally disturbed by their failures, the majority (60.6 percent) recognized that their mistakes or shortcomings were responsible for their elimination. However, 27.4 percent were strongly disappointed and showed resentment toward the persons and regulations involved in their elimination. It is interesting to note that 10 percent of the group were emotionally relieved to be eliminated. Nervousness and fear made flight training very unpleasant to these men, and they were glad to get out of it. The remaining 2 percent accepted their elimination without disappointment or any strong emotional reaction. They seemed to take an objective point of view regarding their disabilities.

Failure in training did not convince these men that they could not fly nor did it eliminate their desire to be pilots, for 86 percent stated that they still would like to take pilot training. The remaining 14 percent definitely were no longer interested in pilot training, although most of them wanted to be given some other assignment that would take them into the air.

In the following paragraphs the type of problem is given as the heading and the percentage of the 150 students giving this type of problem as one reason for their elimination, follows in parentheses. The various problems are presented and discussed in the order of percent frequency with which they were mentioned by the eliminated students, starting with the most frequent.

1. *Nervousness in the air (54 percent).*—This problem was admitted

by 54 percent of this group, although it is probable that there are many who were nervous and did not admit it. Nervousness was caused for 39 percent of the total group by fear of the air. Half of these men claimed they became adapted to the air in a few hours of flying time but that apprehension during the first stages of flight training interfered with their learning and probably resulted in slow progress from the start. The other half of the group admitted that they never succeeded in overcoming their fear and that the fear was a definite factor in their elimination. It is also possible that lack of proficiency was one of the causes of fear with some of these men. It is interesting to note that fear seemed to occur most commonly in landing and in maneuvers such as stalls in which the air suddenly ceased to support the airplane. In landings, it seems probable that visual cues of speed and visual reminders that the airplane was off the ground were much clearer than when the airplane was several hundred feet above the ground.

Fear of failure made 15 percent of the men nervous in the air. Some men were very anxious to be pilots but doubted that they were going to be successful in training. For others, the conflict was between a strong desire to prove that they were superior (or maintain their prestige in the group) and a fear or doubt that they were not going to succeed. Nine percent of the total group admitted that they were particularly nervous on check-rides.

2. *Slow progress (53 percent).*—With very few exceptions, the students seemed to lack insight into the causes of their slow progress, and the number of factors that could be involved were almost unlimited. It seems unlikely that a general slow rate of learning was characteristic of all these students' activities. It seems more probable that these students had trouble with one or two phases of their flight training. For example, if a student were poor on landing he might fail to solo within the required time and be eliminated for slow or inadequate progress, in spite of the fact that progress on other phases of flying was relatively satisfactory. "Slow progress" is a reason frequently given by instructors and students but too broad a category to be of much use in psychological analysis.

3. *Judgment of height and speed in landing (30 percent).*—Errors of judgment in landing resulted frequently in undershooting or overshooting the field, landing too fast, and stalling or pulling the stick back when too high off the ground. As described in an earlier section of this chapter the student had to make a number of complex perceptual judgments and fine motor adjustments in landing, which were undoubtedly too difficult or too numerous for a number of the students, particularly at the rate at which they were expected to learn.

4. *Lack of "feel of the ship" (27 percent).*—This involved inability to sense that the airplane was approaching a stall, or that it was losing or gaining altitude, or that it was turning when the student thought

it was flying straight, or that it was banked, or that it was banked to a greater or lesser degree than the student thought it was, or that the airplane was skidding or slipping. Probably kinesthetic and visceral sensitivity to movement and loss of support was important here in addition to the visual and auditory cues described earlier. Some students may have been deficient in these kinds of sensitivity or they may never have been in situations where they had to learn to pay attention to these kinds of sensations.

5. *Inappropriate attitudes (26 percent).*—This refers to any attitude which interfered with a cadet's progress in flight training. The principal attitudes under this heading are as follows, together with the percent of the total group giving these as possible reasons for elimination: hostility toward instructor (7 percent), lack of interest in flying (7 percent), lack of confidence (5 percent), overconfidence (4 percent), recklessness (3 percent), worry over personal problems (2 percent).

6. *Poor control of the airplane in landing roll (23 percent).*—As previously stated the airplane was still difficult to control after the wheels had been set on the ground in landing. The landing roll had to be kept straight and it was necessary to avoid ground loops. It is interesting to note that 7 percent of the total group reported that they ground-looped at least once during their pilot training before elimination.

7. *Instructional problems (20 percent).*—These students considered that their instructors were not good teachers. Students frequently made the statement "My instructor is an excellent flyer but he can't teach." Other criticisms which may or may not have been objectively correct but which were reported by cadets, were: failure of the instructor to teach the student a maneuver for which he was held responsible on check rides; assumption by the instructor that the student knew more than he did about flying, especially in cases where instructors were changed; failure of the instructor to point out to the student the errors he was making; insistence upon a rigid progress schedule without attempt to adjust the schedule to individual differences in learning rates or methods.

8. *Division of attention and memory (19 percent).*—The student had to attend to a variety of factors, including instruments, visual, kinesthetic, and auditory cues from the airplane itself, the directions received from the instructor, and stimuli outside the airplane, such as clouds, other airplanes, the ground, signals, etc. To overlook or misinterpret one aspect of this complex pattern might cause a serious error in handling the airplane.

9. *Stick and rudder control (17 percent).*—This was inability to handle the stick and rudder with precision and accuracy. This difficulty was encountered more frequently with the rudder than with the stick. From the comments of the students it appeared that this problem

occurred for many because they could not judge the position nor extent of movement of their arms and feet when forced to depend solely on kinesthetic and tactual cues. For example, several complained that they were inaccurate in handling the rudder because they could not see their feet, although the stick was within their visual field.

10. *Visualization of the airplane's orientation and behavior in space (16 percent).*—Some of the students who experienced this problem said that they could not translate the verbal directions of the instructor into a visual image as to what the airplane should do, or could not translate instrument readings into an image of what the airplane was doing, or could not perceive the relation between pressures on the controls and the behavior of the airplane in space.

11. *Judgment (15 percent).*—The more common types of errors of judgment listed by these investigators included: flying too close to other planes, turning while too near the ground, erring in selection of landing fields in forced landings, flying too low, flying over prohibited areas, and flying into or too close to clouds. In a few cases, recklessness or the desire for thrills was the factor behind poor judgment. Some of the men said flying at a higher altitude did not give them a feeling of speed nor provide as much of a thrill as flying near the ground.

12. *Motor coordination (13 percent).*—The difficulty here seemed to be an inability to move the arms and legs simultaneously in different directions, although what was called poor motor coordination was in some cases a lack of insight into the relation between pressure on the controls and the airplane's behavior or lack of perception of balance rather than poor motor coordination itself.

13. *Erratic performance (13 percent).*—These students said that they could do well on one day but very poorly the next day for no reason they could determine.

14. *Flight planning and traffic pattern flying (13 percent).*—Some of the students could not plan a flight from one point of space to another and were also unable to do this in the traffic pattern. For example, one student summarized his problem by saying, "My instructor would tell me when we were at 2,500 feet to go down to 1,000 feet and, when I reached that elevation, to be over a certain field. When I got down to 1,000 feet I was always somewhere a way off from where I was supposed to be." This seems to be very similar to the visualization problem, although of course the student might be able to visualize the flight path but because of the lack of skill in handling the airplane, be unable to follow out his plans.

15. *Mechanical flying (9 percent).*—This problem was frequently one of using the same amount of rudder or aileron, regardless of the degree of bank or turn desired, or flying according to verbal directions from the instructor rather than by the feel of the airplane or on the basis of understanding of the airplane's behavior.

16. *Inadequate wind-drift correction (9 percent).* In this, a student was unable to coordinate his degree of bank and the amount of heading change well enough so that he could compensate for the drift produced by a crosswind.

17. *Poor acrobatics (9 percent).*—This was apparently an inability to perform acrobatics because of deficient coordination and timing.

SUMMARY OF THE DEMANDS OF THE TASK

A statement of the psychological demands of the task of learning to fly is difficult because to date no well-agreed-upon group of functionally independent traits with clearly defined boundaries has been isolated in the structure of human personality. The terms identifying desirable characteristics of the student are used, therefore, not as names of elemental traits, but only as descriptive words.

Deficiencies in Any Specific Flying Skill May Have Multiple Causation

It is not usually possible to say that a given deficiency in flying skill is always caused by some one characteristic of the student. A particular difficulty may be caused by any one or more of a number of different reasons. For instance, a student who gets lost frequently may not merely have poor ability in orientation. One student may get lost because he has poor ability to recognize landmarks. Another, however, gets lost because he forgets to look around and notice landmarks; another, because he is unable to divide his attention and is so occupied by flying the airplane that he does not have time to notice landmarks; another may be distracted by airsickness; another may be preoccupied with family troubles; and still another lose his orientation because looking down at the ground makes him feel nervous, so he does not look often. A given kind of difficulty may have different causes with different students.

The Specific Demands of the Task of Learning To Fly

In the task of learning to fly, the demands upon the student may be grouped under the following heads: The student had to be well motivated to accomplish this difficult task; he had to conquer any fears or apprehensions and to become emotionally adjusted to his task; he had to be able to divide his attention, or shift it rapidly among a number of different problems and to keep track of a number of simultaneous and progressive changes in the actions of the plane; he had to make perceptual judgments of complex sorts to sensations from a number of different sense modalities; he had to have good muscular coordination; and he had to have good flying judgment. The following paragraphs will summarize the demands of the task.

Motivation and emotional adjustment.—While most of the students were motivated by a genuine interest in flying and a desire to aid the war effort as well as a desire to attain officer's status, a few liked everything about being a pilot except the flying. These few had difficulty in learning to fly because it required a high concentration of interest and effort on the part of a student.

The problem of emotional adjustment in the face of a fear-producing situation was an important one. Flying was regarded as dangerous and involved a number of fear-producing situations, such as loss of support, the loud roar of the engine, and the opportunity to look down at the ground from a considerable height. There were also a number of other fears such as fear of failure and fear of verbal punishment from the instructor. These various fears tended to produce tension on the controls, inability to concentrate on the task, and sometimes airsickness. These fears were often specific and did not appear as a general timidity factor in the personality. A student might be nervous and tense while flying certain maneuvers and quite unafraid in most situations on the ground; another might be nervous in social situations but fearless in the air. That nervousness, tenseness, and apprehension were real problems for the student was shown by the fact that more than half of a group of eliminated students who were interviewed admitted that they had been nervous in the air. An analysis of instructors' comments on more than a thousand eliminated students showed that for more than half of the students, nervousness, tenseness, or apprehension were mentioned by the instructor who recommended elimination. The student had to conquer his fears so that he could concentrate on the task at hand and not be emotionally disorganized.

Division of attention.—A student had to be able to keep track of a number of related and interacting activities going on at the same time. This was done either by division of attention or by shifting attention rapidly from one aspect of flight to another. The student had to have the ability to remain calm when, as occasionally happened, he was for a brief space unable to keep up with a rush of necessary judgments and corrections of the airplane's course. He had to do the most important corrections first and work as rapidly as he could to make all of the necessary decisions.

The student had to make rapid perceptual judgments.—In many flying situations the perceptual demands were the most crucial ones. This was confirmed by the validity of perceptual factors in classification tests, by students' reasons why they were eliminated, and by instructors' comments on reasons for elimination of students, as well as by participant-observer observations. The importance of perception was illustrated by the fact that a student with only a little flying time could perform quite difficult maneuvers if the instructor told him what to do and "talked him through" the maneuver.

The estimation of speed and distance was one of the most important pressures for smooth, coordinated flying. These airplanes had higher wing loadings and were more susceptible to stalls and spins when flown at slower airspeeds.

2. The required pilot procedures were more complex and presented particularly difficult problems during take-offs and landings. Combat aircraft also called for a greater understanding of the fundamental principles of aeroequipment, aerodynamics, engineering, and flight procedures.

3. The cockpit position of the pilot with respect to the airplane might be new, requiring different points of visual reference in both normal flight and in the landing attitudes.

Certain problems of transition were unique to the fighters and to the bomber aircraft. These are discussed in the following two sections.

Fighter Transition.—Combat fighter aircraft were fast, complex, had high wing loadings, and presented particularly difficult problems since there was frequently no room in the airplane for an instructor. This required the student pilot to fly solo from the very beginning. A large part of the fighter transition training was therefore done on the ground and became a problem of acquiring new knowledge and transfer of previously learned information to this new situation. The new fighter pilot needed sufficient confidence in his ability to make this transfer so that after only a few hours of ground instruction, he could climb in, take off, return and land safely. This component of self assurance, or its opposite, fear and lack of confidence, was a common variable among new fighter transition students.

After the pilot became airborne, he experienced for the first time the new flying characteristics of his fighter aircraft: The strong G-factor pressing him down into the seat during turns and dive recoveries obscured the normal cues of sustentation; the airplane was highly maneuverable and was particularly sensitive to control movements; it had a rapid rate of climb, fast take-off and landing speed, and imposed upon the pilot new stresses and strains inevitable during maximum performance flight maneuvers. Understanding the fundamental principles of aerodynamics and their application to the speed, power, and weight of his new airplane was one of the more important conditions determining the speed of transition and proficiency level of the fighter transition student.

Fighter pilots were frequently told that their airplane was nothing more than a platform for firing guns. This was to emphasize the importance of accurate fixed gunnery firing. Almost half the time in fighter Transition school was spent practicing the precise, coordinated flying necessary for accurate gunnery. The pilot was his own best instructor and had to recognize his own weaknesses and apply the appropriate corrections on the basis of previously learned flying techniques and information relative to air-to-air and air-to-ground fixed gunnery firing.

Bomber Transition.—The difficult transition problem to four-engine bombers is exemplified by the fact that nearly all of the 10 weeks in the four-engine Transition schools was required to bring the students to the minimum level of pilot proficiency required for the B–17 and B–24 airplanes. Only introductory training could be given in the maneuvers and use of special equipment important in combat operations.

The major problem facing the multiengine bomber pilot was the understanding and control of this very complicated and heavy piece of flying machinery. The more capable the student was in transferring his previously acquired information about aeroequipment, aerodynamics, and engineering, from the single- or two-engine level to these four-engine bombers, the faster he could attain proficiency as a four-engine bomber pilot.

The actual control of the four-engine plane in flight also called for a somewhat different technique by the pilot. It was more a mechanical control rather than flying by the "feel" of the plane as with smaller aircraft. Even under contact conditions the four-engine bomber pilot made frequent reference to his instrument panel since the attitude of the plane was not easily perceived directly, and on high altitude flights the ground and horizon cues were frequently obscured. It took considerable physical stamina and strength to fly the B–24, particularly on the long flights ordinarily made in large bombers. It was said that B–24 Transition training was the only school where the students did not gain weight.

Fear of the ship among new students was a persistent problem reported by four-engine instructors. This was particularly true in the earlier days of the expanded multiengine program when students were frequently sent to B–17 and B–24 schools direct from single-engine AT–6 schools. During the latter part of 1944 a preflight familiarization period of about five weeks was given the students during which time they flew as observers in the airplanes, took ground school courses, talked with other students and instructors and in general became accustomed to the new large airplane prior to actual flight instruction.

A good deal of emphasis was given, on the verbal level, to the pilot's role as "airplane commander." Since the student's time was almost entirely taken with learning to fly the airplane and since there was only one member, the crew chief, to command, little actual training in command was given at the Transition level.

Reconversion to Training Aircraft.—A transition problem in reverse was the reconversion of combat pilots to training type aircraft. This was primarily a relearning problem since these pilots had frequently forgotten, for example, that it was perfectly safe to land at the slow speeds characteristic of the lighter airplanes. With their relatively low-wing loadings, many maneuvers were possible and safe in training airplanes which would have been considered emergencies when flying under combat conditions. In many cases these returnees had never been taught, as cadets, the maximum performance possibilities of the training aircraft. They required a little confidence training before they felt safe turning into a dead engine or making single-engine and power-off approaches and landings in the two-engine training airplanes used in the Training Command.

On the other hand, the combat pilots had to relearn the fact that these trainers were not constructed to stand up under some of the aerial maneuvers commonly used in combat operations. Furthermore, they soon discovered, in the Central Instructors Schools for example, that the sometimes uncoordinated flight control of the airplane permissible in combat was not accepted in the Training Command as good flying technique. The returnees had to relearn the precision type of flying required of all pilot instructors.

The need for careful training in transition to lighter aircraft soon became apparent from the high rate of minor and serious accidents among the returned combat pilots when flying training aircraft without extensive, supervised retraining.

Summary of Essential Pilot Demands

Intellectual.—Rapid transition to new type aircraft, particularly combat airplanes, reflected the pilot's ability to transfer previously learned information to the larger and more complex airplanes. Since there was usually no room in the airplane for an instructor, the student fighter pilot had to make his own *correct* application of his knowledge of aerodynamics, aeroequipment, engineering, and flight procedures on his very first flight in the new airplane. The tremendously complex four-engine bombers required extensive pilot knowledge of equipment and engineering and detailed pilot procedures. The theoretical expectation would be that pilots who had learned to fly with a logical understanding of the aerodynamic principles involved would be better able to transfer to new airplanes than those who had merely acquired perceptual and motor skills by rote learning.

Motor.—It required considerable strength and stamina to fly multiengine bombers and to fly on long-range fighter missions. New airplanes required different motor techniques and physical control for both the precision flying and the maximum performance maneuvers characteristic of tactical airplane operation.

Perceptual.—Transition training frequently required adjustment to new perceptual cues for both the fighter and the bomber pilot. The fast take-off and landing speeds, new cockpit position of the pilot with respect to the airplane, and a more complex instrument panel, illustrate some of the new perceptual habits acquired in transition flying.

Personality and emotional.—The unfamiliar speed, power, size, weight, and complexity of new airplanes frequently presented a mental hazard to new students. Some students made a much more rapid emotional adjustment to this maze of new pilot skills. Traits of leadership and command ability were emphasized in selecting first pilots for multiengine bombers.

INSTRUMENT FLYING

Description of the Instrument Area

The development of objective measures of instrument flying skill provided the opportunity for personnel from the Psychological Research Project (Pilot) to observe many of the fundamental pilot traits and abilities in the important specialized area of instrument flying. The specific research results are presented in chapters 8, 9, and 12; the following analysis is expected to serve as an orientation for those chapters.

Definitions.—The purpose of instrument flying is to enable a pilot faced with bad weather to fly safely through it without the benefit of vision outside of his airplane and to allow him to come down through the clouds, break out of the overcast directly above the field and land. In addition to his knowledge of weather, the pilot is dependent on cues from the instrument panel and the radio aids for information which allows him to maintain his orientation and to plan his flight in a safe and efficient manner.

Importance.—Reports from all of the air forces emphasized the extreme importance of good instrument flying in combat operations. The tremendous mobility of the airplane as a weapon or transport is completely useless whenever it must be grounded because of bad weather. The necessities of war frequently dictated that missions fly through one or more weather fronts and possibly let down and land at a field that was overcast by bad weather; the development of the all-weather airplane created a demand for all-weather pilots. When the field was overcast and large numbers of airplanes were all returning at once from a mission, some of them running low on gas and with wounded aboard, the pilots who were "stacked" up waiting to let down had to hold their assigned positions exactly and those who were breaking out through the overcast could not afford to fail to locate the field on the first pass and go around again several times.

Instrument training and research will continue to receive increased attention in peacetime since it is through instrument flying that aviation can progress against the restraining influences of weather, poor visibility, long over-water hops, and night flying. Commercial aviation has always emphasized instrument flying since it is of primary importance in maintaining scheduled flights. The instrument panel is also used as the major reference for precision control of large multiengine aircraft where the pilot sits far forward and does not have direct view of the wings as an aid in controlling the attitude of the airplane with respect to the horizon.

Training.—At the Basic school the student was taught to perform under instrument conditions nearly all the maneuvers he had already learned under normal contact conditions: straight-and-level flight including changing air speed, turns, climbs and glides, stalls, recoveries from unusual positions, and emergency procedures. In the Advanced schools the student was introduced to the use of radio equipment as a means of performing cross-country navigation, location of the home radio station, and letting down for landing, all under instrument conditions. In Transition training all previous instrument training was reviewed and practiced in the new type airplane. Ground school courses in weather and instrument flying and navigation paralleled instrument flight training. Each rated pilot in the Training Command was required to practice instrument flying at least three hours each month and to pass an annual proficiency check to maintain his rating as an instrument pilot.

Analysis of Instrument Flying

Some of the maneuvers illustrating the more important skills required in successful instrument flying will be described in the following paragraphs.

Flight planning.—This phase illustrates the extensive and accurate information required by the instrument pilot. First he must choose the route and altitude which will have the fewest hazards such as mountains or icing conditions, and which will avoid the worst turbulence and headwinds, and will be facilitated by the most dependable navigational aids. Often a number of these factors must be balanced one against another. To do this involves a wide knowledge of meteorology, navigation, and the potentialities and limits of his airplane and its equipment. The development of printed tests to get at those aspects of this knowledge which can be measured on the ground is described in chapter 12.

Take-off.—The pilot must hold the correct heading with hair-line accuracy. It is apparent that any slight deviation, if not corrected immediately, will lead the airplane off the runway. In training the pilot must perform all actions incidental to the take-off under the hood and by reference to instruments alone. In an actual instrument take-off, the pilot can usually at least see the edge of the runway.

As soon as the airplane is airborne the motor components of flying remain much the same, but the pilot must use and interpret a new set of cues in order to know where he is going and what is happening to his airplane. The only source of these new cues is his instrument panel. Nearly all previous experiences of movement on the ground have been in only two dimensions using visual cues as the direct link between himself and the environment. Instrument flying contradicts these thoroughly established perceptual habits and requires the substitution of a highly *conceptualized* system of control. The pilot must learn the meaning of the instrument cues and substitute these for the more familiar ones in controlling the movement of his airplane in all three dimensions.

The following maneuver illustrates how the pilot makes this substitution of instrument cues for his own sensations of movement.

Straight-and-level course.—The pilot's problem is to hold a constant altitude, air speed, and heading. Since rather narrow limits of instrument deviations must be maintained this is a difficult task for the new students. The secret of successful performance lies in careful trimming and frequent cross-checking of the instruments. Each instrument has its own special advantages and disadvantages. The gyro- and magnetic compasses tell him in what direction he is going and whether or not he is turning. In this function they are supplemented by the rate-of-turn indicator. The magnetic compass has a tendency to swing and give a false indication of direction when the plane is turning, climbing, or diving. The gyrocompass is not subject to these errors but has a tendency to precess away from its original setting and hence must be periodically checked against the magnetic compass. The turn-and-bank indicator tells him whether he is slipping or skidding. The artificial horizon tells him at a glance the attitude of his airplane; whether his nose is high or low, whether his wings are level or banked.

From the preceding description, it should be apparent that the pilot must keep constantly looking from one instrument to another, cross-checking them and integrating the separate cues into a complete picture of what is happening to his airplane at any given moment. In fact, motion picture studies of eye movements made by the Navy Instrument School have indicated that a good pilot makes as many as 200 fixations per minute during precision instrument flying.

Turns.—Students were given a good deal of practice making turns to any predetermined heading. Usually these were "standard rate turns" meaning 3° of turn per second. This rate was used as a standard because it made timing of turns easier since 3° per second means 180° in one minute. It also was used to insure uniformity in the turns required from all pilots when making the let down through the overcast to complete the final low approach and landing. These turns required the student to cross-check the sweep secondhand on the clock against the gyrocompass as well as the ball-bank indicator. They demanded the coordinated use of rudder, ailerons, and elevators.

Frequently, the student was required to execute these turns without the use of the gyrocompass and the artificial horizon. This was done to insure the pilot's ability to continue normal flight in case these instruments should fail to operate correctly due to mechanical failure or because the plane had been thrown into an extreme or unusual attitude. The remaining instruments were referred to as "partial panel" or "restricted panel" and included the rate-of-turn indicator, ball-bank, air speed, altimeter, magnetic compass, and clock.

Steep turns with an angle of bank of 40° or more were also done under instrument conditions though here the main emphasis was on maintaining a safe air speed, a constant angle of bank, and a minimum loss of altitude rather than on coming out on a predetermined heading.

Since it is practically impossible to maintain constant altitude with a variable angle of bank, one of the more common errors was inadequate cross checking between the turn and bank and the altimeter. Other common errors were excessive use and over-control of the rudder, aileron, and elevators; and inaccurate anticipation of compass lag when rolling out of the turn.

Unusual positions.—One of the first things an instrument pilot had to learn was to trust his instrument panel cues despite the frequent conflicting physical sensations erroneously perceived while the plane was in flight. Unless the pilot kept a constant watch on the instrument panel, a slight deviation in bank might develop and not be felt, since it started gradually and involved no change above the threshold for sensing acceleration. This type of error could soon result in a tight diving spiral or, in the case of gradual loss of air speed, in a spin. These were persistent mental hazards which could only be overcome after the pilot gained confidence in his instruments and in his ability to execute a safe recovery from any unusual position of the aircraft.

The aim of all unusual position recoveries was to resume level flight at a safe air speed with a minimum loss of altitude. Recoveries were frequently made difficult by the extreme tenseness of the pilot. Often a pilot would think that he was using all of his strength to move the controls when in effect he was merely working against himself. This was particularly true in the use of the rudder, where a pilot's leg muscles might be so tense as to make it impossible for him to exert enough pressure on one pedal to overcome the pressure on the other pedal. In making these recoveries he had to learn to disregard his natural sensations of position because these were subject to dangerous illusions. For instance, if he straightened out after having been in a diving spiral to the left, he would have the strong illusion of turning to the right, and if he followed the natural tendency to correct for this, he would go back into a spiral again.

Use of radio aids in flight.—The maneuvers described so far did not involve the use of radio aids and were first taught to the student at the Basic level. At the Advanced level the student reviewed his "basic" maneuvers and was taught the procedures of radio navigation, range orientation, let down and low approach, and the use of other special equipment such as the radio compass.

It was only through the use of radio aids that pilots were able safely to take off, fly cross-country, and let down to land at the destination—all under instrument conditions. The "basic" maneuvers were primarily a system of training devices for teaching the pilot the essential components of an integrated instrument flight. This relationship is illustrated in the following maneuver.

Instrument let-down and low approach.—Most of the problems of instrument flying were concentrated in the Instrument Let-Down and Low-Approach maneuver. The purpose of this whole procedure was to enable the pilot to let down from his cruising altitude and to break out through the overcast at the right spot in order to land.

The successive steps in this important maneuver are illustrated diagrammatically in figure 3.2 and are briefly described in the text that follows. By reference to the appropriate tables, the pilot tuned in his radio to the station at his field destination. He then established his orientation in terms of the radio signals and flew to intersect one of the four radio beams. He approached the radio station by following this beam. If he flew too far to one side he heard a dot-dash (A); on the other side a dash-dot (N); when he was in the middle of the beam these signals merged into a constant tone signal, the beam. By a series of corrections he bracketed the beam and found the heading which took him straight down to the station.

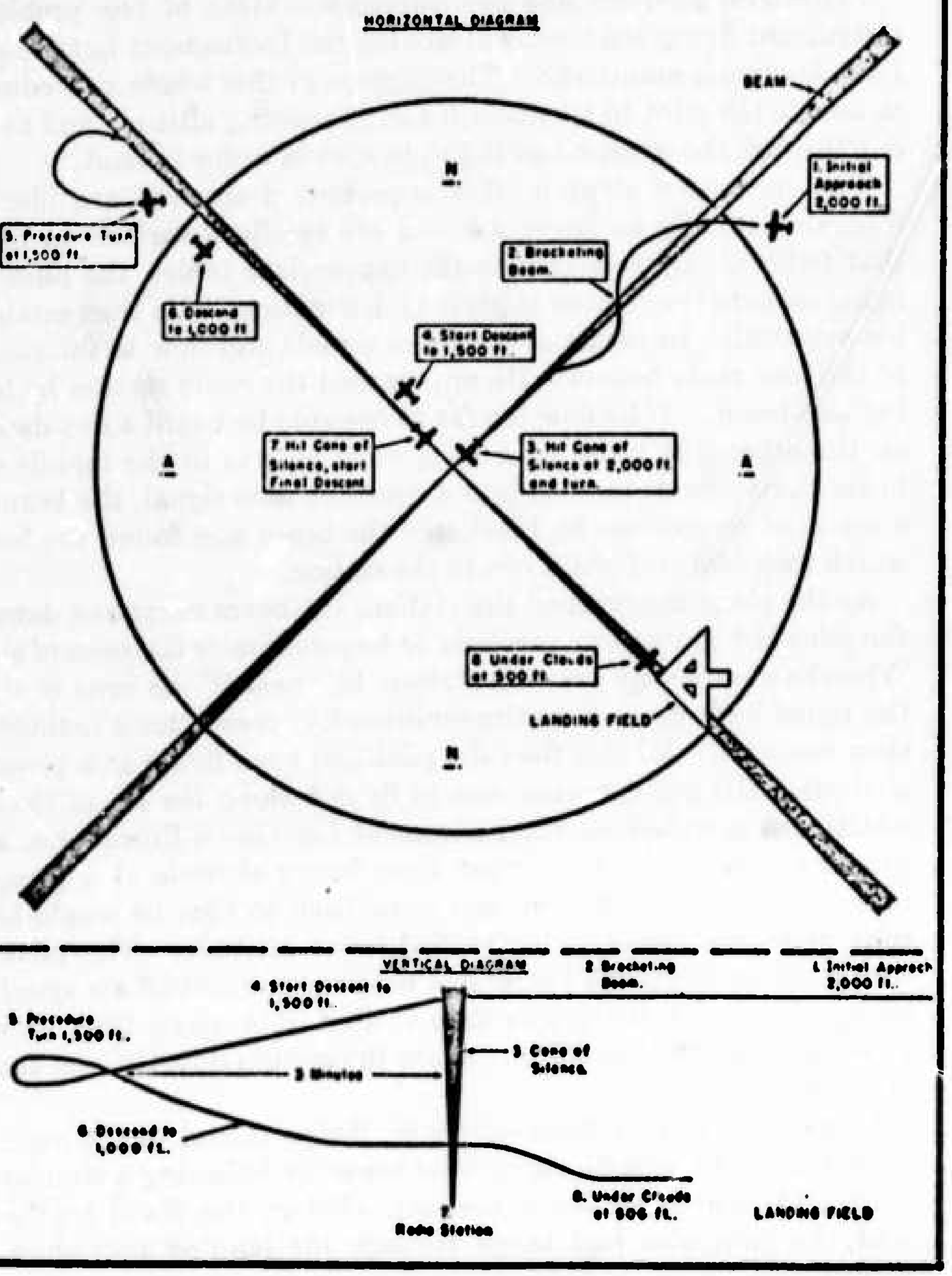
As the plane approached the station, the beam narrowed down and the pilot had to fly more precisely or he would miss the cone of silence. When he was directly over the station, he "heard" the cone of silence; the signal built up to a maximum intensity, ceased for a moment and then resumed. All this time the pilot had been flying at a prescribed altitude. His problem now was to fly out along the leg of the beam which was specified on his Instrument Let-Down Procedures, at the proper air speed, for the proper time, losing altitude at a prescribed rate, make a precision turn, and come back so that he would hit the cone of silence again at a specified lower altitude. The pilot then continued on down the beam at a constant prescribed air speed, still losing altitude at the proper rate so that at a given time he would arrive at an altitude of 500 feet above the ground directly over the edge of the field.

In addition to this flight sequence, the pilot had to maintain constant radio contact with the control tower by following a standardized voice and communication procedure. Before the plane reached the field the pilot also had to go through his landing procedures and checks.

Unless he performed all parts of the maneuver precisely, he was likely to come out in the wrong place and be unable to see the field. Then he had to climb back up according to a prescribed procedure and start all over again. At best, this would waste time and gasoline, and at worst, it might cause him to run into a mountain or some other obstruction. If the problem is complicated by placing many other

Fig. 3.2

DIAGRAM OF THE INSTRUMENT LET DOWN AND LOW APPROACH



airplanes in the vicinity (returning from a combat mission, running low on gas, anxious to land and spaced out on the beam at different altitudes as directed by the control tower) the need for precision flying and clear thinking is obvious.

The execution of this Instrument Let-Down and Low-Approach placed the pilot under marked tension, particularly if he was not familiar with this radio range, if the radio signals were not clear due to static interference or if he had not had recent practice in performing the maneuver. Though the pilot was applying his full power of concentration to the task at hand, a number of disturbing thoughts always managed to creep in to make him "sweat." Flying at 3 miles per minute, an error of 20 seconds or more could cause him to miss sight of the mile long runway. Failing to make the correct adjustment for atmospheric pressure on the altimeter could cause him to crash into obstacles.

All the skills necessary in instrument flying had to be called upon during this critical maneuver: The pilot had to remain relatively calm, relaxed, and clear headed, the successive steps in the procedure had to be followed in the right order and at the right time, the performance of each aspect had to be done with a high level of precision, perception of the changing radio cues and the instrument indications had to be immediate and accurate. While no single phase of this maneuver was unusually difficult or unfamiliar to the trained instrument pilot, the combination into a paced, rapidly moving, complex pattern of flight, placed a severe demand on his over-all pilot skill.

Summary of Essential Psychological Demands

The above description of some aspects of instrument flying has been aimed at emphasizing some of the psychological factors important in this specialized area. One convenient grouping suggests the following analysis:

Motor factor.—Most movements of the controls were small and intended to keep the plane within rather small and precise limits of deviation. Actually, these movements were not noticeably different from contact flying. The pilot was no longer flying with constant reference to the visible horizon and the ground but could refer only to instrument needles and dials whose absolute amount of movement was rather small and with a characteristic lag between the actual new movement of the plane and the indication on the instrument. The result of these conditions was a tendency for the student to work too hard by using excessive and over-controlled movements with consequent strain and physical fatigue.

Perceptual.—Precision instrument flying was as much a function of accurate perception as it was of coordinated motor control. In fact, the chief difference between contact and instrument flying was the substitution of new perceptual cues for old familiar ones. This, of course, is a difficult adjustment in any phase of human activity. The instrument pilot had to maintain a constant watch on his instrument panel, continuously checking the readings of one instrument against one or several others; he had to anticipate slight movements within the rather narrow perceptual field. This task became particularly difficult in rough air and was also conducive to vertigo and semi-hypnotic effects during long flights and at night.

703322—47——6

Intellectual.—Instrument flying was a continuous sequence of planning and following exact procedures. The importance of the knowledge and information element had been indicated in a preceding paragraph describing flight planning. Furthermore, it was a less direct, more conceptualized, and complex pilot task than contact flying. Instrument instructors frequently emphasized to their students that instrument flying is simple if they "just use their head." While this was a frequent admonishment in all phases of flying, it appeared to be especially appropriate in the instrument area.

The intellectual demands of the instrument pilot are substantiated by evidence from a study conducted at Psychological Research Unit No. 3. It was found that the navigator stanine, which has a greater intellectual component, significantly differentiated good and poor instrument students, but not good and poor contact students. This study was done at the Basic level; the difference might have been even greater at the Advanced schools where a more complicated type of instrument flying was taught.

Personality factors.—In contact flying the pilot had constant reference to ground and horizon and could alter his flying procedure at any time. In instrument flying, however, the pilot had less flexibility and was committed to following his course once the decision was made. He therefore had to have sufficient confidence in his knowledge and ability as an instrument pilot to follow this plan until an error was clearly indicated. Instrument instructors emphasized this trait in describing good and poor instrument students. Added to this condition was the fact that actual instrument flying was done under adverse flying conditions which in themselves added stress to the pilot's job.

NIGHT FLYING

Importance

Round-the-clock flying gave the Air Forces much greater flexibility and effectiveness both in combat and in routine flying. The individual pilot frequently found it necessary to fly at night or, if delayed by bad weather, mechanical difficulties, change of route, etc., to complete a planned daytime flight at night. Night flying like nighttime automobile driving was more dangerous than daytime operations and called for superior pilot skill and understanding of the task. Night flying training generally started at the Basic level and continued for the rest of the pilot's flying career.

Equipment

The P-61, the Black Widow, was the only AAF airplane specifically designed for night flying operations but most training and all tactical type aircraft carried the necessary equipment for night flying and landing. Navigation lights on the wing tips, dome, and tail made the airplane visible, runway lights pointed out the landing runway, landing lights on the airplane threw beams ahead and below, a floodlight at the landing end of the runway lit up that area, cockpit lights made the interior of the airplane visible when desired, and fluorescent lights could be turned on to illuminate the dials on the instrument panel. The radio equipment was frequently used, particularly during night cross-country flying to maintain orientation with respect to the different commercial and AAF radio stations.

Analysis of the Night Flying Task

Relation to instrument flying.—It was only on clear nights when ground cues could be grouped into an unmistakable pattern and the horizon was well defined, that night flying was similar to daylight contact flying. Otherwise, night flying had to be considered a special case of instrument flying. At night most new pilots were conscious of the ever-present hazard of loss of orientation and soon learned to depend on their instrument and radio cues as navigational guides. Most of the description given in the instrument flying section of this chapter is, therefore, applicable to night flying operations.

There were a few conditions characteristic of night flying which should be considered in addition to the conventional instrument problems.

Weather.—Night flying in weather that further reduces visibility was a source of confusion to the pilot. It differed from daytime weather flying since dangerous thundercloud formations were not readily discernible and might be entered unawares. The student pilot had to learn how to recognize these conditions and be on the alert for them. Icing was another danger which the student might not observe at night until his margin of safety was greatly reduced.

Perceptual confusion.—Any experienced pilot can tell how he has mistaken a star for a light beneath him and how he thought lights were moving past him, when actually he was turning about the lights. At night a pilot could easily get so confused that he actually did not know which way was up, or whether the aircraft was turning, diving, rolling, or climbing. With a few of these experiences, the pilots soon learned to use the instrument panel as the major reference.

Depth perception takes on new properties in night flying. A strong light on the ground will look closer than a dim light in the same place. Typical of night landing hazards is that of landing by floodlight. The inexperienced pilot is likely to attempt to land on the beam of light instead of on the runway below it.

The pilot must also be on guard against the factors producing vertigo and a type of semihypnotic condition which are not uncommon during night flying. Excessive concentration on the red tail light of a preceding plane in a formation flight, the effects of the constant beam signals in the ear, hyperattention to the dim instrument dials, are among the possible sources of pilot confusion which become particularly effective when the pilot is flying while sleepy or fatigued after several hours in the air.

Night vision.—Candidates for pilot training were not screened on the basis of night vision tests. However, a training program in night vision was developed in the Training Command involving a system of demonstration and exposition. The students were instructed how to adjust their eyes for night vision by remaining in a dark room or wearing red goggles and how best to see objects at night by using peripheral vision.

Since the instrument dials were not easily seen at night, a common error among students was to increase their margin of safety, for example, by flying at higher air speeds.

The superior night flyer was a pilot who understood the special conditions peculiar to night flying and how best to protect himself from these special hazards. The primary requisite was skill as an instrument pilot.

NAVIGATION

Introduction

Cross-country flying started in Primary school and continued through the remainder of training. The student pilot was constantly reminded always to maintain his orientation during both local and cross-country flights. The simplest type of navigation is called Pilotage and refers to control of the flight by direct reference to ground markings. Gradually, more complicated systems of navigation were introduced such as Dead Reckoning; calculations in terms of compass headings, and air speed, ground speed, etc. including corrections for wind, altitude, and compass deviations, and Radio Navigation. By far the most common method used by pilots was the combination of Pilotage and Dead Reckoning with radio aids being used during instrument conditions and as a double check on cross-country reference points.

Pilot navigation in the combat theaters was a much more specialized task which utilized devices and new equipment not available for Training Command instruction. No attempt is made to describe pilot navigation in the overseas areas.

Analysis of Pilot Navigation

Navigation flights were, to a large extent, the practical application of systems and methods of navigation learned on the ground. However, there were wide individual differences in the ability of pilots to apply this information during actual flight.

Planning the flight.—The pilot first consulted available information to determine the course and heading for the most efficient flight to his destination. The pilot had to consider such variables as: Speed; weather; type and range of the airplane and the navigational aids available in it. Good planning was dependent on the pilot's briefing or past experience and familiarity with the route to be followed but these factors could not compensate for a lack of specific knowledge and information essential to successful navigation.

Recognition of check points.—Since the pilot had so many other things to do besides just sitting up there watching the ground, it was important for him to recognize small, specific, identifying landmarks in a very limited period of time. It did not take long for a fast fighter airplane to pass over and leave a small country town but in those few seconds a good pilot navigator had to notice, for example, the angle at which the highway or railroad left the town and a small creek and the bluff near the outskirts to the north. He could then identify these features on his sectional map and locate his position with pin-point accuracy. The pilot who had to "buzz" the railroad station to identify the town probably lacked the fast perceptual ability necessary to recognize ground cues and find them on his map. Over large areas of the United States it was difficult to become reoriented by reference to ground cues unless the pilot possessed very keen perception of the few identifying check-points. The good navigator kept a constant, systematic check on his course.

Maintaining a continuous check on the pre-established course.—After correcting for magnetic variation, the compass heading and estimated time of arrival for the successive legs of the navigation flight were established before take-off. Some estimate was also made for necessary wind correction though this could not usually be determined with accuracy until the actual drift had been established during the flight. Ground air-speed corrections and new power settings also had to be made during the flight after the effect of the wind had been determined at the altitude selected for the flight. It was important that the pilot hold his heading, altitude, and air-speed deviations within as narrow limits as possible, in order to use them as the base line for his corrections.

Facility in making computations.—The speed of military aircraft required rapid and accurate computations of heading, air speed and rate of fuel consumption during flight. These computations had to be accomplished in addition to the usual activities required to maintain a safe and efficient flight. The pilot had available several computing devices, e. g., the E–6B computer, as navigational aids but these aids themselves required precise mental and motor application. It is obvious that a heading error of 2 or 3 degrees, traveling at over 200 m. p. h., if not corrected, would soon lead the pilot away from his check-points and final destination.

Use of radio aids.—Despite the most careful flight planning, many variable factors such as inaccurate forecast of weather conditions, wind direction and velocity, frequently introduced errors in the pilot's course. No predetermined compass heading could therefore be entirely depended upon to bring the pilot to his destination. The use of the radio range was the most common aid to navigation under instrument conditions. By tuning in on the correct frequency the pilot could proceed to the desired station by following the beam emitted by the station. The radio call signals from different stations were also used as intermediate checks en route. Recognition of the identifying station signals, fading or building of the quadrant radio signals (indicating approaching or leaving the station), and recognition of the cone of silence directly over the radio station, were constant auditory perceptual problems which had to be mastered by the pilot-navigator.

Summary of Esstential Pilot Demands

Successful pilot navigation depended to a great extent on such intellectual problems as flight planning, exact following of procedures, awareness of a multitude of conditions, and judgment and decision in the light of variable factors during the flight.

Perceptual skill was also necessary when recognizing and interpreting the ground check points and the radio cues. Precise motor control was important to maintain the course of the airplane within narrow limits of heading, air speed, and altitude. Speed was an important element in all phases of navigation since the orientation of the airplane changes so rapidly. Above all the pilot had to be careful and systematic enough to do a lot of small tasks correctly. He had to resist the temptation to become careless.

FORMATION FLYING

Definition of Formation Flying

Formation flying refers to the flight of two or more planes maneuvering together as a unit: Turning, climbing, and flying straight and level together. The purpose of formation flying is to combine the eyes, firepower, and manpower of the individual airplanes in such a manner as to achieve maximum concentration of power of attack and defense, and control of navigation. When flying in the basic formation, the pilot may be in either one of two positions: Lead pilot, with the responsibility of planning the flight in such a manner that the mission is accomplished without endangering the formation and minimizing the task of the individual pilots in holding the formation; or wingman, with the task of following the lead plane in the prescribed manner. The difficulty of this task varies considerably depending on the type of formation and the assigned position.

There are a number of different types of flight formations depending on the mission, type of plane, and flying conditions, but in all, the separate tasks of the lead pilot and the wingmen remain essentially the same. The following section describes the fundamental problems common to all types of formation flying, from the fighters to the four-engine bombers. Some typical formations are illustrated in figures 3.3, 3.4 and 3.5.

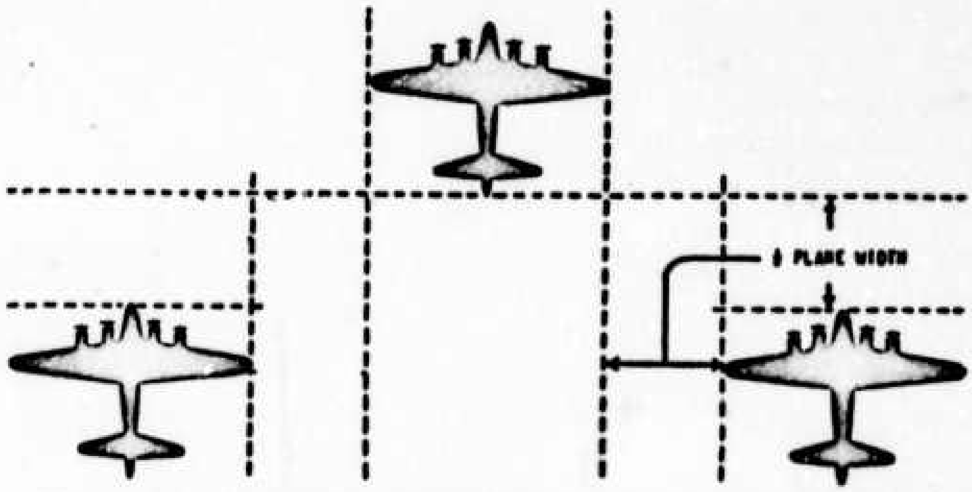


Figure 3.3.—Basic bombardment formation (single element "vee").

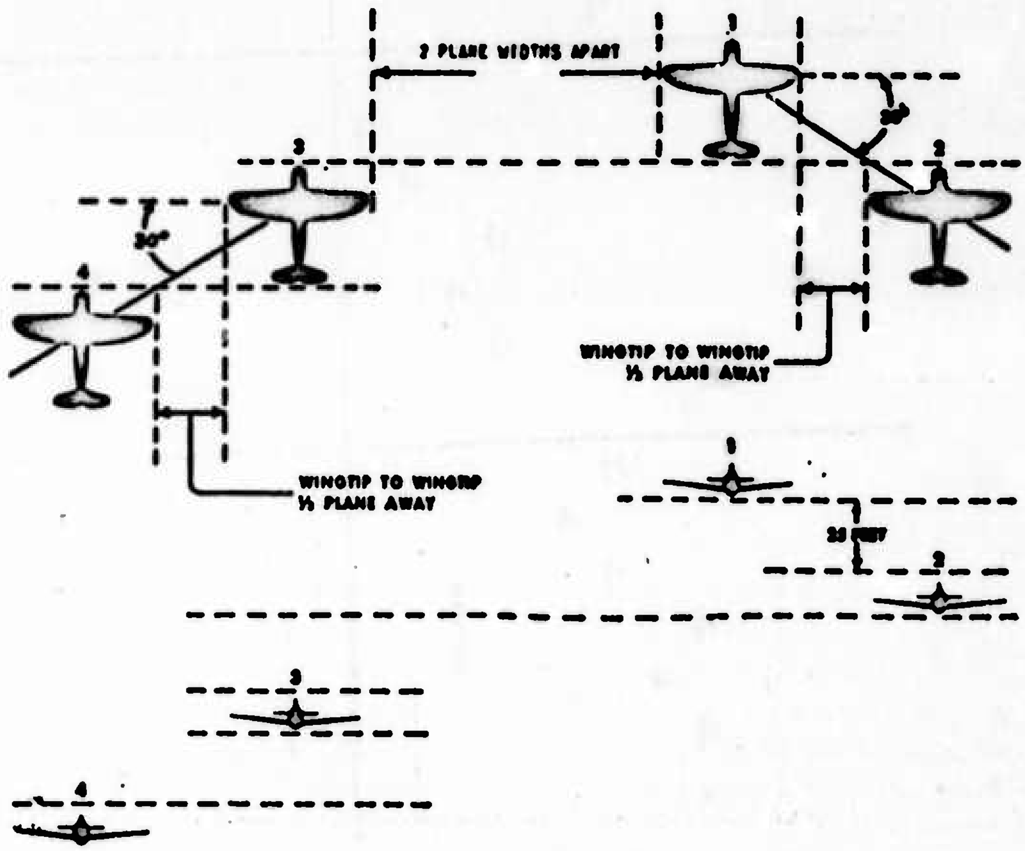


Figure 3.4.—Basic fighter formation.

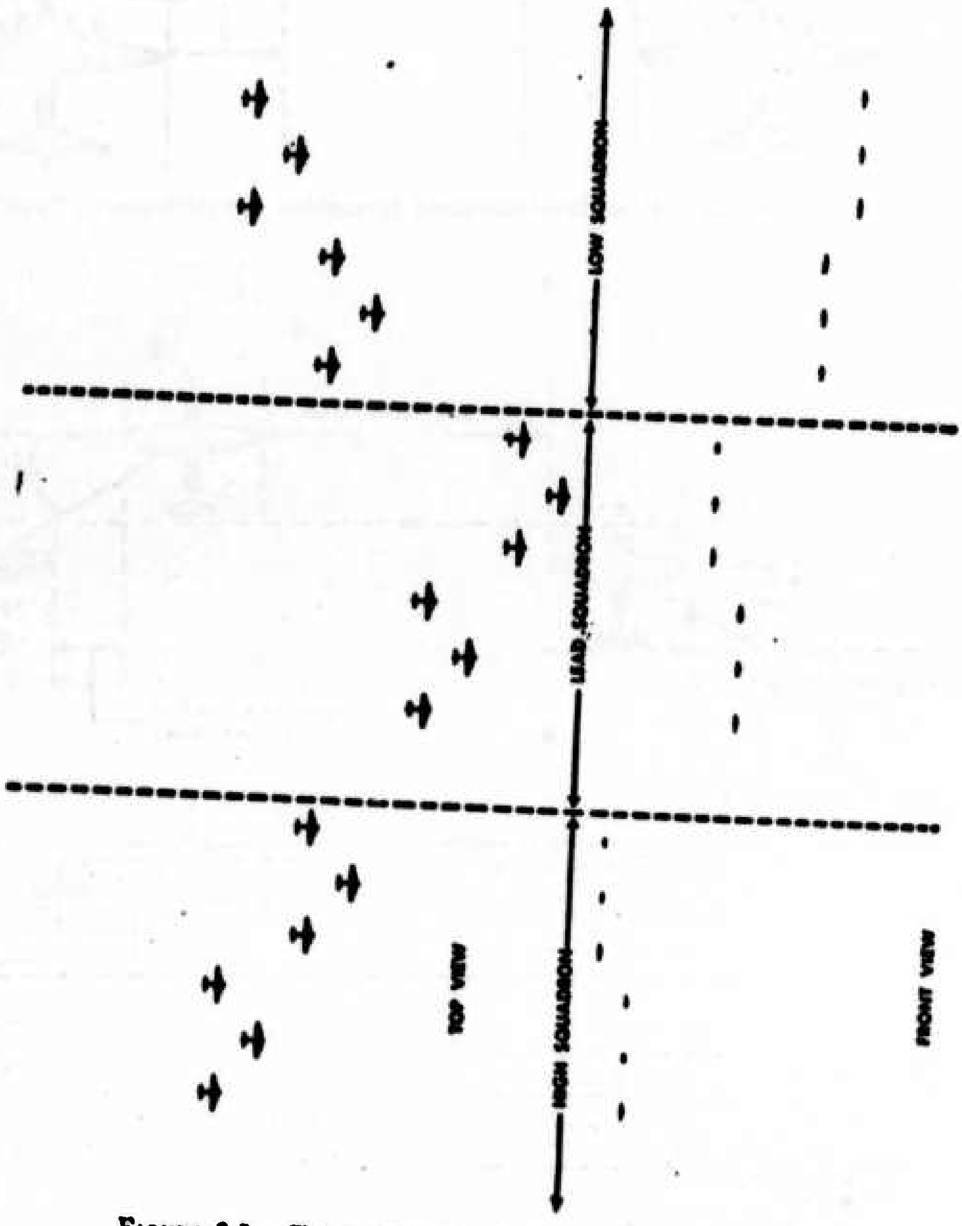


FIGURE 3.5.—Typical bombardment group formation.

Analysis of Formation Flying

The job of the wingman in formation flying was as pure a perceptual-motor task as could be found in any specialized area of military aviation. The wingman had to keep his eyes on the leader and follow every move, anticipating him where possible. When flying at speeds from 160 to 300 m. p. h., it did not take much of a lapse of attention to result in straggling or over-running and possible collision. This constant attention and physical adjustment to the effects of rough air or changing speeds and course of the lead airplane was very fatiguing. Wingmen did not like to fly with a new leader whose flying characteristics were unknown or abrupt, making it difficult to anticipate new movements. Obviously, the best leader was one who could fly a smooth, steady, consistent course, providing distinct signals when any change of positions was required.

The following word picture will illustrate the fundamental pilot skills required in good formation flying:

The problem of joining the formation was largely a matter of following established procedures. The time, altitude, and rendezvous point was established before the flight. The specific pattern of the formation determined the method of approach and joining. Once in the formation, the wingman established his power settings and trim adjustments to require a minimum of additional pilot control to maintain his assigned position. The distance between airplanes varied with the type of formation, mission, visibility and experience level of the pilots in the group but in most cases, the goal was a rather tight formation in which the spacing of airplanes would be at a minimum of one-half wing span distance and at a predetermined level. The good formation pilot would seem to "freeze" right in his niche behind his leading plane while a poor pilot was continually working, changing throttles and moving the rudder, aileron, and elevator controls to help him maintain his position. This excessive work resulted in an accordion effect with a consequent waste of fuel and made the task of smooth flying difficult for all the pilots in the formation.

If the student was not familiar with his plane, he was likely to make errors in amount of stick, rudder, or throttle action with the resulting ragged formation and also getting in the way of the following airplanes. If the student did not make accurate and immediate perceptions of deceleration and acceleration he was likely to lag or overshoot his position and crowd up dangerously close to the leading airplane. The specific perceptual-motor control varied with the pilot's position in the formation; if he was on the outside of a turn, he would need to add power and conversely the pilot inside had to decrease power.

Changes of position within the formation called for maximum skill in formation flying technique. In the first place, a confusion in interpreting the leader's signal would result in a very serious jumble and milling around by all the airplanes. When a formation of nine or more airplanes (as many as several hundred in combat operations) started to revise their relative positions, each pilot had to be able to follow the exact procedure and reform in his new position with a minimum loss of time and gasoline.

For example, when changing from the right- to the left-wing position in the B-24, the cross-under and cross-over techrique was frequently used. If the pilot was on the right, he first lowered the nose of his airplane and then reduced power to compensate for the gain in air speed caused by the new nose-low attitude. He then skidded to the left until under his approximate new position when he immediately added power to regain the altitude lost while crossing under. Common errors were to reduce power too late or too slowly thus crowding ahead of the lead airplane while crossing under, and banking instead of skidding when crossing under thus sending his airplane too far out of the correct lateral position.

The difficulty of performing all formation flying maneuvers increased with the altitude. At high altitude, the airplane was sluggish and responded more slowly when power was added. This placed an even greater premium on the ability of the pilot to anticipate every variation from his desired position since his recovery would be much more difficult and slow. The pilot's knowledge of the aerodynamic and engineering factors involved in high altitude flight made his task both easier and safer.

Summary of Essential Pilot Demands

Except for understanding the signals and procedures for joining, leaving, and changing the formation, the pilot's task was primarily perceptual-motor in character. Minimum throttle changes, smooth, coordinated control action and properly trimmed flight were essential to prevent undue pilot fatigue on long flights as well as to conserve fuel and maintain maximum safety during tight formation flying. The pilot had to possess the ability to maintain close perceptual attention over long periods of time and to catch incipient signs of lag or acceleration. Regardless of the pilot's skill, tight formation flying varied considerably depending upon the altitude, consistency of the lead pilot, mechanical trim and power of the airplane, turbulence, and visibility.

FIXED GUNNERY

The analysis of the task of fixed gunnery is given in chapter 11.

Students were impressed with the maxim that their fighter airplane is nothing more than a rifle barrel with wings. Extensive training was required before the student possessed the fine degree of precise coordination essential for accurate fixed gunnery.

The pilot's task in fixed gunnery firing involved two major types of skill: perceptual skill in evaluating the aim; and flying skill to enable the pilot to fly the airplane in such a manner as to obtain and keep the correct aim while flying a smoothly coordinated course.

THE COMBAT PILOT

When the pilot completed the last phase of instruction in the Training Command, he was ready to learn to become a combat pilot. This was a type of pilot specialty that could only be taught by units in much closer proximity to combat requirements than the Training Command. Aviation psychologists had the opportunity to observe this final and crucial phase of training in the four continental air forces and in several combat air forces in the overseas theaters. The analysis of the task of the combat pilot is presented in the following three reports: No. 14, Psychological Research on Problems of Redistribution; No. 16, Psychological Research on Operational Training in the Continental Air Forces; and No. 17, Psychological Research in the Theaters of War.

CHAPTER FOUR

The Problem of Measuring Flying Proficiency

Maj. Neal E. Miller

FUNCTIONS OF FLYING GRADES

Selection and Elimination

One of the functions of measures of flying proficiency is to select students on the basis of the aptitude which they display. One of the most important types of selection is deciding which students should be allowed to continue flying and which should be eliminated for flying deficiency. Because flying training was expensive and because more than a third of the students who started were eliminated during training, it was important during the war to try to discover and weed out the poor students as soon as possible. In addition to selecting the poor students for elimination, measures of proficiency were sometimes used to select the best students for training in certain specialties such as "night fighter".

Whenever measures of proficiency are being used to select students, either for elimination or for some special assignment, their function is to predict what the student will do in the future. Thus, a high grade at the end of Primary is a prediction that the student will do well in Basic school and eventually in combat; a low grade should mean that he will do poorly.

Validation of Aptitude Tests

Another function of measures of proficiency is to determine the relationship between scores on pilot selection tests given before training and the aptitude which students show during flying training. This function is, of course, closely related to the one which has just been described. Any program of aptitude testing is dependent at a number of crucial points upon the accuracy with which subsequent performance can be measured.

Graduation on a Proficiency Basis and Diagnosis of Weaknesses

A somewhat different function of measures of proficiency is to determine how much students have learned in order to see whether or not they are ready to move on to the next stage of training. If the student is being trained in a number of things at once, the measure may be used to diagnose specific strengths and weaknesses so that time and effort may be shifted from the strong areas to the weak ones where it is most needed. Since most aspects of flying training proceeded on a fixed schedule rather than a proficiency basis, this function was much less important in the wartime program of flying training than the one of selecting students for elimination. The instructors, however, did make some use of the daily and check-ride grades as a basis for diagnosing specific weaknesses and deciding where to concentrate their efforts.

Evaluating Training Methods

Another function of proficiency measures is to determine whether students trained one way are better than those trained in another. Since there was not time for many training experiments, this function was not emphasized during the war. Any improvement in the means of measuring the results of experiments is a fundamental contribution to training research.

It can be seen that the first two uses of proficiency measures involve aptitude while the second two involve learning. To the extent that the students with more aptitude learn more rapidly, these two types of functions are basically similar; to the extent that some students may catch on slowly at first but eventually reach a high level of performance, these two types of functions are basically different.

DIFFERENT TECHNIQUES OF MEASUREMENT

The techniques of measuring flying proficiency may be divided into two main categories: Subjective and objective. The subjective category may be sub-divided into general subjective evaluation and the rated work-sample technique, and the objective category into objective observation and mechanical recording. The divisions between these categories are not absolute; there are cases in which one shades off into the other. Each of the different methods will now be discussed in more detail.

Subjective Evaluation

General Subjective Evaluation.—In this technique, the student is given an over-all rating on some relatively broad quality such as "general pilot ability." This is the type of grading system which was in the most common use in the Training Command. The categories were A, B, C, D, E, or F. An attempt was made to anchor these categories by means of descriptions such as the following one for the grade of D which is given from an instructor's manual for Primary Schools:

> This student is slow to take instruction. You find that you must constantly repeat yourself, *but he finally gets it.* His progress is characterized by inconsistency. You find yourself wondering how to make him understand. "How am I going to get this across?"

Briefly, he has faults of sound *judgment*, but faults that you believe are correctable—not likely to be habitual or dangerous. His *attitude* must still be good. Any habitual indifference calls automatically for an E or F. His *progress*, though slow must still be fairly sure. His *technique* is not good—but not dangerously bad.

Another of the ways to define such categories is the percentile or normal curve system, in which, for example, the upper 10 percent is given A, the next 20 percent B, the middle 40 percent C, etc. This was not used in the Training Command.

Rated Work-Sample.—In this technique, the performance is fractionated into a series of aspects on part tasks, each of which is given a subjective rating. The tasks which the rater should observe are clearly specified. For example, the student may be rated Excellent, Good, Fair, or 1, 2, 3, 4 on a specific item such as the use of throttle to check approaching stall in slow flying. The ratings may be relatively undefined or an attempt may be made to anchor them to specific descriptions. The rated work sample may take the form of a check-list in which a number of different aspects of performance are rated on a two-point, or yes-no scale.

The rated work sample has the advantage of increasing the likelihood that all students will be measured on the same tasks and of specifying the aspects of each task which should be measured. Some research from the Royal Air Force illustrates the type of variability which can occur when the task is not specified. The Empire Central Flying School had 30 flying instructors give check rides to students. They found that these 30 instructors used 25 different exercises, 8 or 9 per instructor, and made 240 different comments in evaluating their students. As a result, they embarked on a program of standardizing the maneuvers which should be used in a check ride, the aspects of these maneuvers which should be observed, and the way in which they should be rated. In this way they produced a rated work-sample type of check ride.

In addition to specifying what should be observed, the rated work-sample method may involve assigning the relative weights to be given to performance on each part of the task. The total score is then an objective addition of the weights given to the various subjective ratings rather than an over-all subjective combination. In the best scales of this kind, the aspects of behavior which are to be rated and the relative weights to be given each of them are determined empirically on the basis of validation experiments.

The British, Canadians, and Australians devoted considerable effort to refining this method of grading flying skill. They found, however, that even with this kind of scale it was necessary to give the check riders extensive training in order to get them to agree on common standards. Refresher courses were necessary in order to maintain agreement. Because agreement was produced by having the different members of the group compare their grades with each other, instead of by comparing their grades with a permanent independent standard, there was no guarantee that the standards of the group as a whole may not have gradually crept up or down over a period of time. Furthermore, the standard agreed upon was not described so that it could be understood and interpreted by anyone who had not taken the special training.

Objective Measurement

In objective measurement absolute (in addition to relative) agreement among observers is achieved by using a permanent independent standard. Instead of using subjective standards such as "poor" or "too slow" which are dependent upon the individual making the judgment, the performance is defined in terms of a permanent standard such as a reference point on the plane or an instrument reading, which is relatively independent of the specific observer. This makes it possible for the results to be interpreted in the same way by everyone. The instrument check specified in AAF Regulation 50-3 represents an important step toward objectivity. It suggests certain limits of heading, airspeed, and altitude which the candidate should be able to maintain.

A convenient way of determining how well absolute agreement has been achieved is to have two or more independent judges use the measure to make observations of the same thing at the same time. In doing this, a distinction must be made between two types of agreement: Relative and absolute agreement. In relative agreement the two judges place the individuals measured in the same rank-order; they agree on who is at the top and the bottom of the class but may not agree on whether the bottom student should be given an "E" or an "F." In absolute agreement, the two judges agree on the grades to be given to each individual and, of course, this means that the rank-orders are the same.

The difference between absolute and relative agreement assumes special importance whenever different students are graded by different observers. If the check-riders from different squadrons or schools were asked to grade the same group of pupils, they might agree fairly well on who should be at the top and bottom of the class but not on the absolute grades to be given to these students. When these check-riders are dealing with different sets of students, the ones who are fortunate enough to get the check-rider who has lower absolute standards are more likely to get a high grade and thus will have a better standing in the ranking of the combined group than those who happen to be checked by a pilot with higher standards. To the extent that absolute agreement can be achieved, it will tend to prevent wasteful differences among the elimination rates of various schools.

In testing for absolute agreement, it should be noted that if all of the students in any experiment are observed by the same two judges, a high correlation between these two judges will by itself mean only that relative agreement has been established. In order to establish absolute agreement it will be necessary to show also that the average of the grades assigned by the two judges and the range (or standard deviation) of these grades are the same. When the results of a number of judges paired at random are combined, so that some students are measured by one pair of judges and others are measured by a different pair, variation in standards will reduce the correlation. Thus, if observations of a number of judges paired at random have been combined on the same scatter diagram, a high correlation indicates absolute as well as relative agreement.

The two main types of objective measurement will now be discussed.

Objective Observation.—In this type of measurement the check pilot makes an objective observation, such as the number of feet altitude lost in a turn or whether or not the cabane strut is parallel to the horizon, and then records his observation by some method such as writing it down or making a check mark on a printed facsimile of the instrument dial. The advantage of this method is that it can be used with the equipment available on the average army airplane. The disadvantage is that the phenomenon observed is transitory and the check-rider must "catch it on the fly" the first time or lose it forever.

Mechanical or Photographic Recording.—This method has the advantage of producing a permanent or semipermanent record. Devices such as the work-adder and slip-skid recorder, which will be described later, add up the student's errors during a given period of time so that they can be read from a dial which remains stationary at the end of the period of recording. These are semipermanent methods. They place much less of a strain on the check pilot, who can make his readings at leisure, observing one reading at a time and checking it, after the maneuver has been performed. Photographs of the instrument panel are an example of permanent recording. It is possible to grade recorded data subjectively, as the Committee on the Selection and Training of Aircraft Pilots did in having a number of observers evaluate motion picture records of a flight. Recorded data are, however, especially adapted to objective scoring.

Necessary Conditions for Good Measurement

Task Clearly Defined to Student.—The student must be told how the various measures of a given maneuver will be scored and weighted. This is necessary because flying is a complex task in which the student can improve one aspect of his performance at the expense of another by shifting his attention. He must be told, for example, the relative penalties for a 5° deviation of heading and a 50-foot one in altitude. If the task is not clearly defined, two students with the same ability

703322—47——7

may get different scores because they have different opinions concerning what they are supposed to do.

One solution to this problem is to devise situations in which the application of a number of complex skills produces a simple, single result. Fixed gunnery (described in chapter 11) is a good example; the student uses all of his flying skill to achieve one single result: hitting the target.

Administration Under Standard Conditions.—Unless all students are measured under standard conditions, the scores will measure the difficulty of these conditions as well as the skill of the student. Differences in the conditions of testing are an especially important source of error in attempting to measure flying skill.

One of the sources of difficulty is the fact that airplanes are much more complex and difficult to standardize than the average laboratory apparatus test. Cars of the same year and model may perform differently, especially if they have been subject to various amounts of use. Similarly, airplane engines come out of the factories in somewhat different condition and change considerably as they receive different amounts of wear. Furthermore, the flying characteristics of the airplane are influenced by a host of other factors, such as the rigging, the weight of gasoline in the tanks, etc. It is difficult to standardize all the characteristics of a piece of machinery as complicated as an airplane.

The fact that the skies are not a very stable laboratory in which to conduct measurements is another important source of difficulty. The velocity, gustiness, and direction of the wind change from day to day, season to season, and place to place. Similar changes occur in visibility, turbulence, and even in the density of the air. These factors exert a considerable influence upon the performance of the student. It is difficult to measure and specify all of them. The Pilot Project encountered at least as much difficulty in dealing with the effects of different conditions of test administration as it did in achieving absolute agreement between observers measuring the same thing at the same time.

The problem is to increase the ratio of the effects produced by differences among students to those produced by differences among uncontrolled conditions. One way of doing this is to devise situations which magnify individual differences. It is sometimes possible (for example, in the rudder exercise stall to be described in chapter 6) to make the task so difficult that the performance of the poor student differs greatly from that of the good one. Another method has been to attempt to devise measures which are less sensitive to the effects of disturbing factors. For instance, maintaining that air speed which is most efficient for climbing is less influenced by up- and downdrafts than is the actual amount of altitude gained per unit time. Still another approach is to standardize planes and instruments and only administer check rides under ideal conditions of wind, turbulence, visibility, etc. Sometimes, especially in instrument check rides, it is possible to go to an altitude at which the effects of turbulence are minimized. For many other types of measurement the standardization of all important conditions is extremely difficult in a large-scale training program.

A different approach is to repeat the measures under different conditions so that the variable errors produced by them will be balanced out and the constant factor of student skill will emerge from the average. The student can be given a number of different check rides, or an objective measure may be repeated as a part of the daily grading system throughout a given phase of training. It should be noted, however, that replication will not get rid of the constant errors introduced by the fact that conditions such as turbulence may be generally worse in some sections of the country or during certain seasons of the year. The work on fixed gunnery described in chapter 11 shows that it is difficult to devise formulae to correct for such factors. These constant differences in the conditions will produce the same effect as would a difference in the absolute standards of the observers.

ATTRIBUTES OF A GOOD MEASURE

Permanence and Reliability

If a quality is extremely transient, it is usually not important to measure it. A transient characteristic cannot predict future success. The permanence of a characteristic can be measured by repeating the test after a period of time. If the characteristic is in the process of being learned through a course of training, it is not necessary that the student remain at the same absolute level; it is only necessary that he maintain approximately the same rank order relative to other students who are going through the same course of training. If the test and retest are given by different observers in different airplanes, a good correlation between them indicates not only that the measure has a certain degree of permanence, but also that there is reasonably good agreement between the two observers and that changes in conditions do not unduly influence it. The longer the interval between the test and retest, the more severe is the test of permanence and reliability.

Relevance

It may be possible to measure a relatively permanent characteristic of the individual which, however, is irrelevant. Three methods have been employed to test the relevance of the measures developed by the Pilot Project:

Professional Judgment of Experts.—Before measures were subjected to further tests, they were criticized by expert pilots. Only those considered to measure qualities relevant to the task of the combat pilot were saved for further work. In developing the printed tests of flying information which are described in chapter 12, it was not possible to get any further criterion of combat relevance. It was necessary to use items which had been experimentally demonstrated to give reliable measures of types of information which experts returning from combat believed were important. Though this procedure is not nearly as safe as validation, it has definite practical value when suitable criterion groups are not available.

Validation Against Groups Differing in Flying Ability.—In this procedure each measure was tested for its ability to discriminate between students who showed different amounts of skill at the end of the same amount of flying training. In some studies, the measures were correlated with flying grades. In other studies, extreme groups composed of the best and poorest students in a class were selected on the basis of grades or by having each instructor rank his students. The check rides were given only to the students in these extreme groups and the ability of each measure to differentiate between these two groups was determined. The advantage of this method lay in the fact that even though the grading system had low reliability, there was likely to be little overlapping between the true abilities of the extreme groups. The final method was to determine the ability of each measure to predict which student would subsequently be eliminated for flying deficiency. This method yielded a rigorous test of both the permanence and the relevance of the measure. All three of the methods were based upon the same reasonable assumption: namely, that the system of grading which was in common use dealt with relevant aspects of flying skill and needed improvement only in its reliability and standardization.

Validation Against Groups Differing In Amounts of Flying Training.—In this procedure a test was made of the ability of a measure to discriminate between two groups of students who had the same aptitude but differed in the number of hours of flying training received. An alternative procedure was to compare measurements on the same group during earlier and later stages of its training. This procedure was based on the reasonable assumption that the curriculum taught the students relevant types of flying skill so that the measures which discriminated best between the two groups with different amounts of training were the most relevant.

In using this criterion the investigators kept in mind the fact that some items of achievement might be so specific and easily learned that the student's performance would be determined almost entirely by the sequence in which these items were taught in the particular curriculum to which he was exposed. (To give an extreme example, it might be possible to secure a perfect discrimination between two groups by teaching only one of them to understand an unusual technical term.) Such nonpower type of items would be expected to be unstable and transient. An attempt was made to distinguish them from the items, measuring more fundamental and slowly acquired types of achievement, which represented milestones on the course of learning. Wherever two groups with different amounts of flying training were used to validate measures of flying skill, the investigators from the Pilot Project made certain that the less-experienced group had a good chance to become familiar with the maneuver to be measured. The assumption was that, if the less-experienced group had already had a chance to pick up all the quickly acquired aspects of performance, the superiority of the more experienced group would necessarily be limited to those more fundamental skills the gradual acquisition of which is an important feature of learning.

It will be noted that the two experimental procedures just described test for different kinds of relevance: The first procedure tests ability to discriminate differences in aptitude; the second, ability to discriminate differences in learning. If the students with the same experience but less aptitude make the same types of errors as those with normal aptitude but less experience, the two procedures will be equivalent. To the extent that students with low aptitude and students who are inexperienced characteristically make different kinds of errors, the two methods will not be equivalent and it will be desirable to select two somewhat different arrays of measures for the two different uses. Measures selected by the first procedure will be most relevant in deciding which students should be eliminated and in validating aptitude tests; those selected by the second will be most useful in graduating students on a proficiency basis, diagnosing their strengths and weaknesses, and in evaluating the results of training experiments.

Comprehensiveness

Measures may be relevant, but still partly inadequate because they cover only a few of the important factors involved in flying skill. If a suitable over-all criterion is available, it is possible to determine the comprehensiveness of a given set of measures by seeing how well the combined score of all of the measures is able to predict the criterion. Since part of the failure to predict perfectly may be due to the unreliability of the criterion rather than to a lack of comprehensiveness of the group of measures being tested, it is necessary to have some independent way of evaluating the reliability of the criterion. If the reliability of the criterion is known, it is possible to use the formula for correction for attenuation to estimate how well the measures would predict a perfectly reliable criterion.[18] When no comprehensive criterion is available and one is relying upon expert judgment, it is especially important to examine the comprehensiveness of the group of measures to be used.

[18] See Guilford, J. P., *Psychometric Methods*, McGraw Hill Book Co., New York and London, 1936, p. 367, formula 155.

If the measures are going to be routinely used to evaluate the results of training, the instructors will tend to modify their teaching in such a way as to train their students to get the best grades on these measures. Any continued use of a system of testing is bound to exert a certain amount of influence on the curriculum. If the system of testing is one-sided, the curriculum will be distorted; if the measures are sufficiently comprehensive, they will tend to keep the curriculum properly balanced. For this reason particular attention was paid to the comprehensiveness of the measures developed by the Pilot Project.

Specificity

Although the battery of measures should be comprehensive, it is desirable to have the individual items specific so that they can be used to diagnose the student's strengths and weaknesses and as differential criteria for validating aptitude tests. One of the defects of subjective measures is their susceptibility to "halo effect." The rater is likely to be influenced by his general impression so that he tends to ignore the good points of poor students and the weak points of good ones.

measuring more fundamental and slowly acquired types of achievement, which represented milestones on the course of learning. Wherever two groups with different amounts of flying training were used to validate measures of flying skill, the investigators from the Pilot Project made certain that the less-experienced group had a good chance to become familiar with the maneuver to be measured. The assumption was that, if the less-experienced group had already had a chance to pick up all the quickly acquired aspects of performance, the superiority of the more experienced group would necessarily be limited to those more fundamental skills the gradual acquisition of which is an important feature of learning.

It will be noted that the two experimental procedures just described test for different kinds of relevance: The first procedure tests ability to discriminate differences in aptitude; the second, ability to discriminate differences in learning. If the students with the same experience but less aptitude make the same types of errors as those with normal aptitude but less experience, the two procedures will be equivalent. To the extent that students with low aptitude and students who are inexperienced characteristically make different kinds of errors, the two methods will not be equivalent and it will be desirable to select two somewhat different arrays of measures for the two different uses. Measures selected by the first procedure will be most relevant in deciding which students should be eliminated and in validating aptitude tests; those selected by the second will be most useful in graduating students on a proficiency basis, diagnosing their strengths and weaknesses, and in evaluating the results of training experiments.

Comprehensiveness

Measures may be relevant, but still partly inadequate because they cover only a few of the important factors involved in flying skill. If a suitable over-all criterion is available, it is possible to determine the comprehensiveness of a given set of measures by seeing how well the combined score of all of the measures is able to predict the criterion. Since part of the failure to predict perfectly may be due to the unreliability of the criterion rather than to a lack of comprehensiveness of the group of measures being tested, it is necessary to have some independent way of evaluating the reliability of the criterion. If the reliability of the criterion is known, it is possible to use the formula for correction for attenuation to estimate how well the measures would predict a perfectly reliable criterion.[16] When no comprehensive criterion is available and one is relying upon expert judgment, it is especially important to examine the comprehensiveness of the group of measures to be used.

[16] See Guilford, J. P., *Psychometric Methods*, McGraw Hill Book Co., New York and London, 1936, p. 367, formula 155.

If the measures are going to be routinely used to evaluate the results of training, the instructors will tend to modify their teaching in such a way as to train their students to get the best grades on these measures. Any continued use of a system of testing is bound to exert a certain amount of influence on the curriculum. If the system of testing is one-sided, the curriculum will be distorted; if the measures are sufficiently comprehensive, they will tend to keep the curriculum properly balanced. For this reason particular attention was paid to the comprehensiveness of the measures developed by the Pilot Project.

Specificity

Although the battery of measures should be comprehensive, it is desirable to have the individual items specific so that they can be used to diagnose the student's strengths and weaknesses and as differential criteria for validating aptitude tests. One of the defects of subjective measures is their susceptibility to "halo effect." The rater is likely to be influenced by his general impression so that he tends to ignore the good points of poor students and the weak points of good ones.

CHAPTER FIVE

Studies of Subjective Measures of Flying Proficiency

Capt. Avrum H. Ben-Avi[16]

INTRODUCTION

Since more than one-third of the students starting the expensive process of flying training were eliminated for flying deficiency, it was important to have an accurate way of measuring their ability so that good students would not be eliminated or poor ones allowed to continue. The grading system developed for this purpose was similar at the various levels of training in the Training Command. In general, three types of flying grades were given: Daily grades, progress or elimination check-ride grades, and final proficiency ratings. All three types of grades were subjective, in terms of categories such as A, B, C, D, E, and F, which were based upon the experience and judgment of the instructor or check pilot.

The daily grades were based on the instructor's observations during routine training flights. The instructor either flew with the student or, in some of the more advanced stages of training, flew alongside of him in another airplane. Check-ride grades were based upon observations made by an experienced check pilot during an official check ride. In some cases, the check rider was required to include a specified list of maneuvers; in other cases, he was allowed complete freedom to select any maneuvers which had been included in the student's training up to that point. At the completion of each phase of training, the student pilot was rated on various aspects of his performance and these ratings were forwarded to the next school.

Since the various subjective flying grades used in the training program were the most readily available criteria against which to validate aptitude tests for the selection of pilots, various organizations in the Aviation Psychology Program made different studies of these grades. The studies relevant to each aspect of the grading system are summarized in the following sections.[17]

[16] S/Sgt. Charles P. Gershenson assisted in the preparation of the manuscript for this chapter.

[17] Unless otherwise specified, studies analyzing daily grades and check-ride grades were performed at Psychological Research Unit 2. This series of projects was supervised and conducted by Maj. M. P. Crawford, Capt. J. T. Dailey, Lt. J. T. Cowles, Lt. R. J. Keller, T/Sgt. E. N. Schrader, S/Sgt. H. F. Schmonsees, Sgt. R. W. Unger.

GRADUATION-ELIMINATION AS A CRITERION

The chief criterion used in evaluating the selection and classification tests developed in the Aviation Psychology Program was graduation or elimination from pilot training. This was also one of the criteria which the Pilot Project used in evaluating objective measures of flying skill. In using this criterion, the few students eliminated for reasons other than flying deficiency were discarded. This criterion was selected for two related reasons: (1) Its obvious practical importance, since eliminees represented a considerable waste of time and money; (2) the fact that because of its practical importance, the schools expended the maximum effort of making this decision as carefully as they could.

That this criterion was far from ideal, however, will become evident as the various measures upon which it is based are discussed in subsequent sections of this chapter. Whenever this criterion is used to evaluate the subjective measures described in this chapter, it is important to remember that, since these measures entered into the process of making the pass-fail decision, it is by no means a completely independent criterion for these measures. Indeed, in the case of the check ride, the measure was the direct basis for passing or failing the student.

A dichotomous criterion of pass-fail restricts the precision of prediction. Also the very nature of the pass-fail criterion tends to interfere with its evaluation since the failures are not available for further study. In addition to these briefly mentioned qualities of this criterion, was a third, that of variability which is considered in greater detail below.

Variability in Elimination Rate

There was a great deal of evidence that different commands and schools had varying standards for grading their students. Evidence for this is reflected by the dissimilar elimination rates among schools and commands. Table 5.1 abstracted from a study by Lt. Col. J. P. Guilford at Psychological Section, Office of the Surgeon, Hq. Army Air Forces Training Command, illustrates the variability of elimination rates.

The following conclusions were drawn from this study: (1) Elimination rates ranged from 10 percent to 60 percent in the 269 class groups, with dispersions two to three times those to be expected by random sampling. Even when differences due to variations in average level of aptitude (measured by pilot stanine) were allowed for, large differences remained; (2) the greatest differences were found between commands; (3) in some commands the greater differences were between successive classes, and in others, between schools in the same class; (4) schools did not recognize differences between class groups nearly as well as they did between individuals in the same class.

TABLE 5.1.—*Distribution of the elimination rates in primary school classes 43–G, 43–H, 43–I, 43–J, and 43–K*

Elimination rate (percent)	Number of school-class groups[1]			All commands
	EFTC	CFTC	WFTC	
-10	....	....	1	1
11-15	1	....	6	7
16-20	4	....	7	11
21-25	6	6	11	23
26-30	19	12	8	39
31-35	25	22	18	65
36-40	15	27	10	52
41-45	15	12	4	31
46-50	15	11	3	29
51-55	6	3	....	9
56-60	1	1	....	2
N	107	94	68	269
M	36	37	29	34.7
σ	9.4	7.6	9.5	9.5
σ_p[2]	3.6	3.5	3.0	3.4

[1] The elimination rate for each class at each school was computed and tabulated separately. Thus, in WFTC one class at a school had an elimination rate from 6-10 percent, six had rates from 11-15 percent, etc.

[2] The standard error of the percentage: computed in each case on the hypotheses of an over-all elimination rate of 34.7 percent; the size of the samples was taken as the mean size of class, that is, 176, 196, and 225 for the three commands, respectively, and 195 for the entire Training Command.

EFTC, CFTC, and WFTC refer to Eastern, Central, and Western Flying Training Commands, respectively.

Other psychological units conducted studies which demonstrated the existence of similar variability in the elimination rates of different schools. In the Western Flying Training Command, Psychological Research Unit No. 3 demonstrated that for Classes 43–A through D, the elimination rate for flying deficiency at one school was as low as 14 percent, while that of another was as high as 42 percent. Psychological Research Unit No. 2 reported similar results for classes 43–D to 43–I in Central Flying Training Command. There the elimination rate for flying deficiency was 22 percent for the lowest school, and 41 percent for the highest school. Each of these elimination rates was based on approximately 1,000 cases; the differences were far greater than those expected by chance.

All results indicated that there were substantial differences in the standards of the various schools. After making a field trip to study this problem, two officers of Psychological Research Unit No. 2 concluded that the supervisory personnel and instructors at one school were convinced that it was their duty to eliminate any students whom they felt would not develop into excellent pilots; while those in another school believed that it was their duty to graduate any students whom they considered reasonably capable of succeeding in later stages of training.

Considerable effort was devoted to the standardization of elimination policies, especially during the latter part of the war. When students were assigned to schools, consideration was given their aptitude scores in an attempt to equate classes for aptitude. Standardization boards were organized in each of the training commands which were responsible for visiting the various schools and checking on the standards for elimination. These efforts, however, never completely achieved the goal of establishing uniform and standardized procedures.

Reliability of the Pass-Fail Criterion

The reliability of the pass-fail criterion could not be evaluated directly in the normal training situation, since the failures were not available for any second appraisal. The experiment in which the Pilot Project planned to have students retained who would otherwise have been eliminated, could not be carried out with meaningful results because of the disruption of training in the period following shortly after the end of the war. One may deduce, however, that the pass-fail criterion has a certain amount of reliability, since it can be predicted by the pilot stanine. From data which are available it is possible to estimate the minimum reliability of the criterion.

On an experimental group of over 1,000 students who were not selected on the basis of their aptitude test scores, the correlation between the pilot stanine and the pass-fail criterion for complete pilot training was 0.66. This is the figure which is relevant in evaluating the stanine as a method of selecting students for pilot training.

Under ordinary circumstances, however, only the students with the higher stanines were sent to training. Thus, the group normally in pilot training was highly selected. Once these students were selected, the problem for training research was to secure reliable measures on this group in spite of the restricted range of talent. The relevant figure for estimating the reliability of the pass-fail criterion, therefore, is the correlation of the pilot stanine with pass-fail for the group of high-stanine students sent to training. On one large sample this correlation was 0.40.[18] Similarly, the reliability of the pilot stanine was 0.94 for the relatively unselected population which is relevant to evaluating the stanine as a selective device, and may be estimated as 0.84 for the restricted group retained for training.[19] By using the values of 0.40 and 0.84 in the formula for correcting a correlation coefficient for attenuation, it can be estimated that the reliability of the pass-fail criterion could not have been below 0.19.[20] This is a minimum estimate of reliability because it assumes that if it were not for the unreliability of the stanine and criterion, the stanine would be a perfect predictor.

Another very rough approximation of the reliability of the pass-fail criterion may be obtained in a different way. On a sample of 2,905

[18] This figure is based on 33,788 students in classes 44-J through 45-F tested by the November 1943 battery.

[19] The value of 0.94 is a corrected (Spearman-Brown) odd-even reliability of the aggregate weighted scores for a sample of 1,000 cases tested from 30 January 1944 to 14 February 1944. This study was directed by Maj. P. H. DuBois and Lt. A. F. Jenness at Medical and Psychological Examining Unit No. 7. The other figure was estimated by using the formula for correcting a reliability coefficient for heterogeneity which is discussed in Kelley, T. L., *Statistical Method*, pages 221-23. The standard deviations of the stanines in the restricted and unrestricted samples were assumed to be the same as those in the sample cited in the preceding footnote.

[20] Peters and Van Voorhis, *Statistical Procedures and Their Mathematical Bases*, Formula 121, page 203.

students, ratings of general pilot ability made at the Primary level of training correlated 0.27 with later ratings made at Basic. These two ratings were independent. They were made at each school by the same personnel who were responsible for passing or failing the student. Because they were less important than the decision to pass or fail the student, these ratings were probably made with somewhat less care. Assuming that the ratings at the two levels of training were of approximately equal reliability, the reliability of each must have been at least 0.27. Any differences in the skills required at the two levels of training or changes occurring in the students during the 10-week period between ratings would lower the correlation between them. In this sense it is a conservative estimate of their reliability.

It is also an estimate for a population with a restricted range because it does not involve the students who failed in Primary and hence never went to Basic. This restricted population is the relevant one for estimating the reliability of flying grades at the Basic level. The estimate for the reliability at the Primary level, however, should be corrected in order to apply to the less restricted range involved there. For 500 students in Pilot Classes 44-H, I, and J at Jones Field, Bonham, Tex., the standard deviation of the flying grades of students graduating from Primary to Basic was found to be 0.70 of that of the entire class before the weaker students were failed. Using this figure in Kelley's formula for correction for heterogeneity, it may be estimated that the reliability of flying grades on the entire range of the Primary population is somewhere in the vicinity of 0.65. Since this is an indirect estimate based on a number of possibly inexact assumptions, it must be taken as a very rough approximation.

PROGRESS CHECK RIDES

Introduction

The Progress Checks were given routinely to the students at the completion of the main stages of training throughout the course of flying training. In addition, any time the student's daily grades were not satisfactory, he was given a progress check ride.

Progress checks were given by special check pilots, who usually referred to the student's daily grade slips, and/or discussed the student's weaknesses with his instructor before the check, so that the check ride would be devoted to the critical areas of the student's performance. A few check pilots preferred to know nothing about the student's performance prior to the check ride. In almost every instance, however, the check pilot discussed his opinion of the student with the instructor following the check ride in order to compare his evaluations with those of the pilot who had observed the student over a longer period.

If the check pilot believed that the student had promise, he was returned for further training with another instructor. If not, he was given another ride by a second check pilot. If the two check pilots agreed that the student should not be allowed to continue in training, he was eliminated. For the poorest students, the second check ride was more or less a matter of form; on the other hand, some students received as many as a dozen check rides before finally being eliminated or graduated from the school. In a typical Primary school investigated by Psychological Research Unit No. 2 it was found that the 836 students in classes 44-H to 44-K were given an average of 4.1 check rides by pilots other than their own instructors.

In evaluating these studies it should be remembered that successive evaluations in check rides at the same school are not independent of each other or of instructors' grades.

Relationship to Pass Fail

Attempts at determining the relationship of mean check-ride grades with graduation-elimination have resulted in biserial correlation coefficients of about 1.00, when Primary pass fail was used as the dichotomous variable. These results substantiate the statement that elimination from Primary school was based directly upon the results of check-ride grades.

Relationship to Success in Subsequent Phases of Training

The ability of grades at one level of training to predict subsequent performance was an important practical test of a measure's validity. It was also a rigorous test since it was reasonably certain that the two estimates of performance at the different levels of training were independently made.

A study conducted at Psychological Research Unit No. 2, determined the ability of average check-ride grades for the Primary stage of training to predict various criteria of performance at the Basic and Advanced stages. The results from this study, presented in table 5.2, show a low positive relationship between average check-ride grades at Primary school, and measures of proficiency at Basic and Advanced schools.

In another study, Psychological Research Unit No. 2 found a correlation of 0.26 between Primary and Basic average check-ride grades for 390 cases. It is of interest to note, in table 5.2, that the pilot stanine predicted success in Basic at least as well as did the average of the check-ride grades in Primary.

The low reliability of the subjective grades probably contributed to these low correlations. It is also possible that students who were slow in catching on at first did better during the later stages of training; that different levels of training demanded different qualities or abilities from the students; or that some students changed their motivation or attitude. Whatever the reason for these low correlations, the fact that grades at one level of training did not predict grades at the next level of training is of considerable practical importance in that it suggests that quite a few of the students eliminated in Primary might have received satisfactory grades had they been allowed to proceed to Basic.

TABLE 5.2.—*Correlation of average check-ride grades with subsequent success* [1]

	Prim.	Bas.	Adv.[2]	N_t	P_s	M_t	M_s	M_f	S. D. $_t$
Average check-ride grades (Primary)	....	0.16	0.11	252	0.59	5.02	5.14	4.84	1.75
Average check-ride grades (Basic)	0.16	....	.38	252	.59	4.98	5.42	4.34	1.77
Upper-lower (Advanced)	.11	.38	....	252	.50	....	....	....	....
Pilot stanine	.19	.19	[3] .24	252	.59	6.56	6.80	6.22	1.48
Pilot stanine plus experience credit	.20	.19	.27	252	.59	6.74	7.00	6.37	1.56

[1] Based on an average of approximately 4 check rides per student; the borderline cases received more check rides and the extreme ones fewer. All average check-ride grades were converted into normalized scores by schools. If the differences between schools had not been eliminated in this manner, the correlations would have been still lower.

[2] Final statements of instructors on student proficiency were dichotomized into an upper and lower group. Data from class 44-H, Frederick.

[3] All correlation coefficients are uncorrected Pearsonian product-moment, except for correlations with the Advanced variable, which are biserial coefficients.

Reliability

Method.—Two attempts were made in a study conducted at Psychological Research Unit No. 2, to arrive at estimates of the reliability of check-ride grades. One method used was the analysis of variance technique developed by Hoyt,[21] modified for this specific situation.[22] The second estimate of reliability was the correlation between mean check-ride grades assigned by Army and civilian check riders and corrected for double length (Spearman-Brown). The procedure in use at that time provided for elimination check rides to be administered by Army fliers, while regular check rides could be given by civilian pilots.

Population.—The population used consisted of class 44-H, Garner Field, Uvalde, Tex., and classes 44-H through K, Jones Field, Bonham, Tex.

Conclusion.—Both techniques yielded similar correlations ranging from 0.72 to 0.89 in different schools and different classes. It should be pointed out, however, that these correlations yield maximum rather than true estimates of reliability since the Army check riders were influenced by the results of the civilian check rides and both check riders were influenced by daily grades and by the instructor's opinion.

What little evidence is available indicates that when adequate precautions are taken to see that the subjective evaluations by check riders are made independently, the reliability is somewhat lower than the figures quoted above. In a study conducted by the Pilot Project[23] 46 lower phase Primary cadets in class 46-A were given an experimental check ride at Moton Field, Ala. The correlation between two

[21] Hoyt, C. J., "Test Reliability Estimated by Analysis of Variance," *Psychometrika*, 1941, 6 153-160.

[22] The modification consisted of taking account of the unequal number of check rides assigned students in obtaining the "within group" sums of squares. This analysis was the work of Lt. R. J. Keller.

[23] This study was supervised by Lt. Ralph Showalter and Sgt. John R. Rohrs.

individual check riders' subjective evaluations was 0.35. When this figure is stepped up by the Spearman-Brown formula to predict a reliability for a combination of 4 check rides comparable with the reliability quoted above for an average of 4.1 check rides, the reliability becomes 0.69. Since this study, which is the only one available, involved such a small number of cases, the results must be taken only as a first, rough approximation.

OFFICERS PROFICIENCY CARD—AFTRC FORM 2

Introduction

The function and purpose of the Officers Proficiency Card was described as follows: "The Proficiency Card is designed to establish a uniform system for rating the demonstrated proficiency of each individual student in each successive phase of aircrew training received at schools under the jurisdiction of this command, and also to provide the organization of first assignment with an individual report on the training and proficiency of each student for use as a guide to facilitate training and assignment."[24] According to the procedures recommended, this one form was to contain a complete and cumulative record of the student's history in the Training Command. At each phase, the student was rated for certain military and personal characteristics as well as upon the pertinent aspects of his training course, on a three-point scale as follows: AA for above average, A for average, and BA for below average. A copy of this form appears on pages 381 and 382.

Discussion

In almost all studies dealing with the Officers Proficiency Card AFTRC Form 2 certain factors were observed which are relevant to its use and interpretation. First was the manner in which the ratings on the card were made. The Psychological Section, Hq. Second Air Force, found that the filling out of this form was usually regarded as an unnecessary burden. The feeling that it constituted an administrative chore, the results of which would never be utilized was expressed, and in many instances was well founded. It was more the exception than the rule for instructors or supervisors to consult these cards for information relevant to the student's previous training or past performance.

The attitude on the part of those responsible for making the ratings was reflected in a number of different ways. A survey conducted by the Psychological Section, Office of the Surgeon, Hq. Second Air Force, at ten Operational Training Units disclosed that in none of these were completed forms available for more than 60 percent of the personnel. The ratings also showed considerable "halo" effect,

[24] T. C. Memorandum 15-2, Proficiency Card (AFTRC Form No. 2), Hq. AAF Training Command, 28 February 1944.

demonstrated by the fact that the ratings for an individual were much more similar than the wide range of traits and performances being rated would suggest. There was also a serious restriction in the range of ratings within groups. It was repeatedly established that the majority of students within a class received the same rating. Substantial numbers of records were encountered on which entries were incomplete or improperly entered. Entries from different schools had such different distributions that, when studies were based upon populations drawn from more than one school, the grades had to be normalized to eliminate school differences.

In spite of these limitations, Psychological Research Unit No. 3 in working with the AFTRC Form 2 found for classes 44-G, 44-H, and 44-J totalling approximately 6,000 cases, that ratings of General Pilot Ability at the Primary level correlated as highly with the pilot stanine as did graduation-elimination.[25] These ratings were also positively and significantly correlated with other measures of performance in the Primary phase of pilot training. The other measures were flying grades, amount of dual instruction prior to initial solo flight, and a three-category rating by instructors. In a later study, the same unit found that the correlation between the pilot stanine and the rating on General Pilot Ability in Basic schools was 0.38 for 1,597 students.[26]

Prediction of Subsequent Performance

The following table summarizes the results of experiments designed to evaluate the predictive power of the rating of General Pilot Ability on the Proficiency Card. It can be seen that all these correlations are low, ranging from 0.23 to 0.34. They are of about the same order as those for average check-ride grades.

The Psychological Section, Office of the Surgeon, Hq. Second Air Force, investigated the relationship between Training Command ratings on the AFTRC Form 2 and similar ratings of proficiency in operational training. The results of this study indicated that individuals given a high rating in the AAF Training Command tended to do better in operational training than did individuals given poor ratings. Correlations between estimates of proficiency in the Training Command and in the Second Air Force were between 0.30 and 0.40. There was a definite tendency for correlations between ratings in the two commands to be higher with the more specific and clearly defined ratings in the Second Air Force. It was also apparent that, in general, ratings made at Basic and Advanced schools correlated higher with operational criteria than did ratings in Primary.

The finding that the more specific, clearly defined ratings were better was also evident in a study by Psychological Research Project

[25] This study was done by Capt. R. H. Henneman and Sgt. S. H. Marsh.
[26] Capt. R. H. Henneman and Sgt. R. C. Anderson conducted this study.

(Pilot). It was demonstrated that ratings of flying ability were somewhat more predictive of success in Central Instructors School than were the ratings on personal qualities. It was also found that certain of the specific flying ratings (such as the one on formation flying), were about as predictive of success at Central Instructors School as the rating of General Pilot Ability. The simple sum of the best specific ratings on the form gave a considerably better prediction of success at Central Instructors School than did the rating of General Pilot Ability. Since the best ratings were selected on the basis of the same sample used to validate the prediction of the combined score, this prediction will be expected to shrink somewhat on subsequent cross validation.

Conclusions

The ability of the grades recorded on the Proficiency Card to measure proficiency and predict future performance was reduced by the lack of uniformity in the methods used to procure these grades, irregularity and omissions in making the ratings, and the fact that in many cases, the individuals who made the ratings on this card did not have any clear idea of their possible importance. In general, the more specific and better defined ratings were more closely related to success in flying training. The simple sum of these ratings seems to be better able to predict success than the over-all ratings of General Pilot Ability. Proficiency Card ratings were more highly correlated with other measures of proficiency at the same level, than with grades at subsequent phases of training. This was probably because the other measures at the same level influenced the ratings which were placed on the card.

TABLE 5.3.—*Correlation of the rating for general pilot ability on the proficiency card with subsequent measures of general pilot ability*

Levels of training	Measure predicted	*N*	r[2]	Source
Primary v. Basic	Proficiency card [1]	2005	0.27	PRP(P).
Primary v. Basic	Proficiency card	681	[3].34	PRU #3.
Primary v. Basic	Final grade	681	[3].25	PRU #3.
Basic v. CIS	Pass-fail	683	.31	PRP(P).
Basic v. Advanced	Proficiency card	2025	.21	PRP(P).
Advanced v. Transition	Transition record	852	.26	PRP(P).
Advanced v. CIS	Pass-fail	683	.23	PRP(P).

[1] Proficiency card grades are all ratings of General Pilot Ability.
[2] Correlations which are either uncorrected tetrachoric correlations or have been corrected for grouping into broad categories according to the procedure described in Peters and Van Voorhis *Statistical Procedures and Their Mathematical Bases*. (pp. 393-399.)
[3] Differences between schools have been eliminated by normalizing data, by analyzing each pair of schools (one at each level) separately and combining correlations by z-transformation, or by using single pairs of schools.
PRP(P) is Psychological Research Project (Pilot).
PRU #3 is Psychological Research Unit No. 3.

ANALYSIS OF STAGE-GRADES

Introduction

In Primary schools the students were given certain flying exercises called stages. Grades of performance on these stages were entered in each student's daily log book. In one study of stage grades, the population consisted of approximately 500 students in classes 44-H I, and J from the elementary school at Jones Field, Bonham, Tex. The variables studied are listed below. For all except the last two, a letter grade, A, B, C, D, or E was assigned. In computation, these were weighted 5, 4, 3, 2, 1, respectively.

Supervised solo.—This is an average of three grades assigned by the student's own instructor on the basis of his first three solo flights. These involve taking off, circling the field and landing with no traffic.

Solo stage.—This is the same as the supervised solo, except that it is performed in traffic and a single grade is assigned by the instructor in charge of the stage. All of the stage grades are assigned in this manner with the instructors taking turns as "stage commanders." Instructors do not have their own students for stages, except for the supervised solo stage.

90-degree stage.—This involves accuracy landings from a 90-degree approach in traffic. These accuracy landings involve landing as closely as possible to a prominent white line across the runway area. Both accuracy and technique are considered in assigning grades.

180-degree stage.—This is the same as the 90-degree stage except that a 180-degree approach is employed.

Power stage.—This involves *accuracy landings*, as defined above, made from a power approach.

Cross-country stage.—This involves flying a triangular course of about 300 miles. Grades are given on the basis of the student's over-all performance during the flight.

Hurdle stage grade.—The student tries to land his airplane in a small circle from a straight power approach over a 6 to 8 foot hurdle placed just short of the landing circle. The grade is an over-all rating on both accuracy and technique.

Hurdle stage score.—This score is an objective score based on the number of times the cadet succeeds in landing inside the landing circle. It is assigned at the same time and by the same grader as the hurdle stage grade. Scores may range from 6 to 35.

Hours dual before solo.—This variable is the total number of hours of dual instruction before the first supervised solo. Since this value should bear an inverse relationship to rapidity of progress in learning, it was decided to correlate it with the other measures in the study. Because of the nature of the variable, all correlations with it have been reversed in sign to show positive association with desirable performance. Since, by directive, all cadets must have at least 8 hours of dual before solo, this variable is highly skewed, although the function it measures is probably much more nearly normally distributed. Thus, it was decided to dichotomize hours dual before solo at the median and compute biserial correlations against this dichotomy.

Correlation with Pass-Fail

In table 5.4 given below are the correlations of certain of the stage grades with pass-fail.

Intercorrelations

In table 5.5 are given the intercorrelations of the stage grades and other related variables.

Conclusions

In general, the intercorrelations between stage grades are very low, even though some of the stages represented very similar maneuvers. This seems to indicate that the stage grades were either lacking in reliability or measured highly specific variables. In either case they appear to be of little value as criteria of pilot proficiency. The pilot stanine predicted pass-fail better than did any of the stage grades. The stage grades did not correlate appreciably with the pilot stanine, average check-ride grade, pass-fail or even with each other.

ANALYSIS OF DAILY GRADES

The daily grade sheet consisted of a list of maneuvers and exercises which were taught in Primary, along with categories for evaluating the student's Attitude, Judgment, Progress, and Technique. The student's performance for each maneuver was routinely evaluated during each flight on a scale from A to F. Each grade below C was amplified by a written statement describing his weakness. The analyses presented below were conducted on both the grades and the comments. Copies of the Primary Daily Grade slip forms appear in the appendix.

TABLE 5.4.—*Correlations of hours dual before solo, average supervised solo grade, solo stage grade, mean check-ride grade, and pilot stanine with primary pass-fail*

	Class	N_t	P_s	M_s	M_e	M_t	SD_t	r_{bis}	Average r [1]
Supervised solo	44-H	156	0.90	3.44	3.13	3.41	0.57	0.28	....
	44-I	189	.87	3.22	3.12	3.21	.64	.08	....
	44-J	139	.89	3.62	3.33	3.59	.57	.26	0.21
Solo stage	44-H	154	.92	3.06	2.92	3.05	.60	.11	....
	44-I	188	.87	3.01	2.92	3.00	.65	.08	....
	44-J	133	.89	3.25	2.64	3.19	.64	.49	.24
Hours dual before solo [2]	44-H	163	.91	......	......	......	......	[3].08	....
	44-I	190	.86	......	......	......	......	[3].15	....
	44-J	139	.89	......	......	......	......	[3].28	.17
Mean check-ride grade	44-H	162	.91	2.65	1.29	2.53	.70	[4].97	....
	44-I	190	.86	2.59	1.15	2.30	.85	.91	....
	44-J	139	.89	2.52	1.00	2.36	.80	.99	.97
Pilot stanine	44-H	163	.91	6.45	5.53	6.36	1.49	[5].31	....
	44-I	190	.86	6.49	5.81	6.40	1.27	.29	....
	44-J	139	.89	6.99	5.80	6.86	1.30	.48	.36

[1] Combined by Fisher's z-transformation.
[2] Dichotomized at median.
[3] Tetrachoric coefficient obtained by computing four-fold point correlation and correcting for unequal dichotomy of criterion variable.
[4] These high correlations confirm the observation that the check-ride grades are the basis for passing or failing students.
[5] Tested under both July and November 43 Directives. No correction for restriction in range attempted for class 44-H only.

TABLE 5.5.—*Intercorrelations of stage grades, check-ride grades, and stanines for class 44-H, Jones Field* [1]

	N	1	2	3	4	5	6	7	8	9	10
1. Supervised solo	141	......	0.36	0.15	−0.01	0.24	0.07	0.15	0.15	0.19	0.15
2. Solo stage	141	0.36	......	.07	.12	.11	.00	.02	.07	.04	.12
3. 90° stage	143	.15	.07	......	.16	−.04	−.18	.09	.07	−.03	.05
4. 180° stage	147	−.01	.12	.16	......	.07	.02	.08	.06	.12	.03
5. Power stage	147	.24	.11	−.04	.07	......	.14	−.02	.04	.14	.07
6. Cross country	80	.07	.00	−.18	.02	.14	......	−.14	−.03	.10	.11
7. Hurdle stage grade	115	.15	.02	.09	.08	−.02	−.14	......	.82	−.05	.11
8. Hurdle stage score [2]	147	.15	.07	.07	.06	.04	−.03	.82	......	.00	.07
9. Pilot stanine	148	.19	.04	−.06	.12	.14	.10	−.05	.00	......	.20
10. Average check ride	148	.15	.12	.05	.05	.07	.11	.11	.07	.20	......

[1] Only students who graduated from Primary were included in these computations.
[2] Sign reversed to indicate association of good performances.

TABLE 5.6.—*Correlation between the average grade during the first four weeks of primary training and that during the last five weeks* [1]

Samples of Class 44–J (N=137) and 44–K (N=145)

Item	44-J r'_{II}	44-K r'_{II}	Combined [2] r'_{II}	r_{II}
1. Straight and level [3]	0.27	0.24	0.26	0.41
2. Turns [3]	.34	.26	.30	.46
3. Stalls	.52	.64	.59	.74
4. Spins	.39	.56	.48	.65
5. Landings	.46	.52	.49	.66
6. Forced landings	.61	.70	.66	.79
7. Attitude	.71	.77	.74	.85
8. Traffic	.53	.63	.59	.74
9. Relaxation	.67	.68	.68	.81
10. Progress	.57	.67	.62	.77
11. Judgment	.66	.72	.69	.82
12. Daily grade	.60	.74	.68	.81

[1] Since both sets of grades were assigned by the same instructor, the "halo effect" probably makes these correlations spuriously high as estimates of split-half reliability.
[2] By use of Fisher's z-transformation.
[3] For these two items the low coefficients appear due to the fact that a maximum of one or two grades were assigned in these maneuvers during the advanced stage.

Intercorrelations.—The intercorrelations of the daily grades were also determined, simply by correlating the mean grades for each maneuver or exercise with every other. These data are given in table 5.7.

TABLE 5.7.—*Intercorrelations of average check-ride grade and daily grades*

Based on 343 Primary Students in Classes 44–J and K

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Average check-ride [1]		0.66	0.69	0.77	0.75	0.69	0.74	0.23	0.69	0.56	0.77	0.73	0.77
2. Straight and level	0.66		.84	.70	.71	.73	.68	.18	.73	.69	.76	.68	.75
3. Turns	.69	.84		.81	.78	.77	.78	.30	.76	.69	.85	.79	.84
4. Stalls	.77	.70	.81		.83	.79	.83	.33	.77	.66	.85	.81	.86
5. Spins	.75	.71	.78	.83		.79	.80	.27	.79	.65	.80	.77	.80
6. Landings	.69	.73	.77	.79	.79		.79	.24	.82	.66	.78	.73	.77
7. Forced landings	.74	.68	.78	.83	.80	.79		.26	.80	.63	.84	.87	.87
8. Attitude	.23	.18	.30	.33	.27	.24	.26		.24	.26	.29	.32	.30
9. Traffic	.69	.73	.76	.77	.79	.82	.80	.24		.65	.73	.76	.79
10. Relaxation	.56	.69	.69	.66	.65	.66	.62	.26	.65		.69	.63	.67
11. Progress	.77	.76	.85	.85	.80	.78	.84	.29	.73	.69		.88	.96
12. Judgment	.73	.68	.79	.81	.77	.73	.87	.32	.76	.63	.88		.93
13. Daily grade	.77	.75	.84	.86	.80	.77	.87	.30	.79	.67	.96	.93	

[1] See text for description of ways in which the instructor who gives the daily grades may influence the student's average check-ride grade.

Studies of the Daily Grade

First to be considered will be studies concerned with an evaluation of the actual grades entered by the instructors on the daily-grade slips.

Reliability.—The reliability of the daily grade was estimated by correlating the mean daily grade on each maneuver for the elementary phase of Primary (the first four weeks of training) with the mean grade for the same maneuver for the accuracy-and-acrobatics phase (the last five weeks). The statistics for this analysis are given in table 5.6. A survey of table 5.6 reveals that the correlations are high for most single items and for the total daily grades. These correlations, however, give a spuriously high estimate of reliability, since the grades which were correlated were not made independently by different instructors, but were for the most part all assigned by the same individual. The fact that an instructor's grade at one time agreed with his grade at another may not mean that he was able to score the student's flying accurately, but only that he formed a fixed opinion which tended to persist from one day to another. Such general opinions, called "halo effects," have been found to be common in almost every type of subjective rating.

In this table it can be seen that, with the exception of the grades for Attitude, all of the daily grades are highly intercorrelated. This characteristic is undesirable when the goal is differential information concerning the elements of a grading system for flying instruction. The high correlations substantiate the existence of "halo" mentioned above, though, of course, some correlation between these grades would be expected, since they were measures of functions considered to be related. In table 5.8 are presented the results of a similar analysis conducted on a different population, using a larger number of maneuvers than in the study mentioned above. While with this second sample the intercorrelations in general were somewhat lower than those presented in table 5.7, they continued to be relatively high.

Correlation with pass-fail.—The summary statistics for the correlations of the daily grades with pass-fail are given in table 5.9. It will be remembered that in many instances the daily grades determined whether or not the student would be sent up for an elimination check ride. The check pilot frequently discussed the student with his instructor before the check ride in order to learn what weaknesses to look for, and before making a final decision, almost invariably compared his appraisal based on one ride with the instructor's opinion which was based on many more rides. In the light of these conditions it is not surprising that the correlation between the instructor's daily grades and the check pilot's recommendation for pass-fail is high. The highest correlations with pass-fail are for grades on stalls, spins, landings, and traffic; the lowest are on attitude, relaxation, and judgment. It is interesting to note that although the personnel at schools all emphasize the importance of judgment, their grades on

Table 5.8.—*Intercorrelations of maneuver grades in primary training*

[Based on 303 Students in Classes 45-A and 45-B, Jones Field]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Straight and level		66	53	49	58	55	51	53	39	54	53	46	53	43	36	27	58	58	49
2. Turns	66		81	59	67	66	65	60	36	56	69	60	64	51	47	18	60	63	66
3. Coordination exercises	53	81		52	60	58	59	53	34	51	58	52	61	48	42	7	49	53	61
4. Rectangular courses	49	59	52		66	48	54	49	28	45	56	46	53	52	42	21	54	50	57
5. 8's across road	58	67	60	66		61	61	55	30	48	63	63	64	60	58	17	56	60	65
6. Stalls	55	66	58	48	61		79	63	28	50	59	64	69	55	52	21	55	58	67
7. Rudder exercises	51	65	59	54	61	79		65	24	48	58	66	70	59	52	21	56	56	67
8. Spins	53	60	53	49	55	63	65		24	42	56	60	61	53	50	19	57	52	61
9. Taxiing	39	36	34	28	30	28	24	24		45	39	33	35	24	26	21	37	41	29
10. Take-offs	54	56	51	45	48	50	48	42	45		63	51	56	40	40	14	56	53	45
11. Landings	53	69	58	56	63	59	58	56	39	63		63	64	55	54	26	59	62	63
12. Forced landings	46	60	52	46	63	64	66	60	33	51	63		63	59	58	25	56	59	76
13. Steep turns	53	64	61	53	64	69	70	61	35	56	64	63		58	54	20	58	59	68
14. 8's along road	43	51	48	52	60	55	59	53	24	40	55	59	58		70	17	53	55	61
15. Elementary 8's	36	47	42	42	58	52	53	50	28	40	54	56	54	70		15	47	49	56
16. Attitude	27	18	7	21	17	21	21	19	21	14	26	25	20	17	15		29	27	27
17. Traffic	58	60	49	54	58	55	56	57	37	56	59	56	58	53	47	29		62	63
18. Relaxation	58	63	53	50	60	58	56	52	41	53	62	59	59	55	49	27	62		65
19. Judgment	49	66	61	57	65	67	67	61	29	45	63	76	68	61	56	27	63	65	

this factor are in the group which has the least relationship to whether they pass or fail the student.

Factor analysis.—A factor analysis of the data given in table 5.7 was performed. Table 5.10 showing the Centroid and Rotated Loadings for the maneuver grades is given below.

In spite of the high intercorrelations of the maneuvers and exercises three factors which appeared fairly significant were isolated. These were:

Rotated Factor 1.—This factor had highest loadings for spins, landings, traffic, and forced landings, and had the lowest loadings for attitude, straight and level, relaxation, and turns.

Rotated Factor 2.—This factor had the highest loadings for judgment, and forced landings. The extremely low loadings for landings and traffic in this factor are worthy of note.

Rotated Factor 3.—Although most of the grades had high loadings on this factor, both straight and level and turns had the highest loadings and were nearly pure measures of this factor. This factor appeared to account in large measure for the "halo effect" which might be inferred from the high intercorrelations among the daily grades.

TABLE 5.9.—*Correlations of daily grades with pass-fail in primary training* [1]

[44-J (N=137; P_s=.85) 44-K (N=206; $_s$=.83)]

	M_s	M_f	SD_t	r_{bis}	M_s	M_f	SD_t	r_{bis}	z—average 44-J & K
1. Straight and level	3.60	2.83	0.56	0.71	3.45	2.50	0.58	0.92	0.86
2. Turns	3.21	2.42	.57	.77	3.11	2.08	.63	.91	.87
3. Stalls	3.01	2.07	.62	.83	2.91	1.58	.75	.99	.97
4. Spins	3.15	2.11	.61	.93	3.17	1.85	.75	.98	.97
5. Landings	3.36	2.29	.69	.85	3.21	1.89	.76	.97	.94
6. Forced landings	2.95	2.03	.68	.74	2.94	1.64	.77	.94	.89
7. Attitude	4.14	3.86	.44	.35	4.19	3.94	.49	.29	.32
8. Traffic	3.64	2.68	.54	.98	3.53	2.15	.78	.98	.98
9. Relaxation	3.64	2.71	.64	.79	3.66	2.80	.65	.74	.76
10. Progress	2.88	1.91	.61	.86	2.80	1.73	.68	.88	.87
11. Judgment	2.98	2.11	.64	.74	2.91	1.92	.69	.82	.79
12. Daily grade	2.78	1.86	.63	.79	2.68	1.62	.67	.88	.85

[1] See text for description of ways in which the instructor who gives the daily grades may influence whether the student passes or fails.

TABLE 5.10.—*Factor analysis of maneuver grades in primary training*

Item	Centroid loadings			Rotated loadings [1]			
	I	II	III	I	II	III	h^2
1. Average check ride	0.81	0.15	0.03	0.66	0.22	0.51	0.67
2. Straight and level	.83	−.24	.18	.28	.24	.80	.78
3. Turns	.90	−.20	.07	.40	.18	.81	.85
4. Stalls	.91	.09	.08	.59	.31	.64	.85
5. Spins	.89	.09	−.09	.64	.16	.60	.80
6. Landings	.87	.06	−.18	.66	.06	.59	.79
7. Forced landings	.90	.23	.09	.66	.38	.54	.87
8. Attitude	.32	−.10	.21	.03	.21	.35	.17
9. Traffic	.87	.08	−.17	.68	.09	.58	.81
10. Relaxation	.75	−.21	.10	.29	.16	.70	.60
11. Progress	.93	.04	.12	.54	.33	.70	.89
12. Judgment	.90	.11	.24	.51	.46	.63	.87

[1] Code: I has been tentatively called perceptual judgment; II headwork; and III motor technique.

Analysis of Comments on the Daily Grade Slips

The comments which instructors entered on the grade slips to amplify each grade C and below were also analyzed by Psychological Research Unit No. 2, in the hope of establishing differential criteria of flying ability. The purpose of the grade slip comments was to provide a cumulative record of a student's progress throughout the course. For the guidance of the instructor a standardized list of "Suggested Grade-Slip Entries" was provided for all schools in Central Flying Training Command by the AAF 31st Flying Training Wing, in charge of Primary pilot training. The entries on this list were grouped according to maneuver and included virtually all common deficiencies and mistakes. Examination of the grade-slip entries at several schools disclosed that most of the grade-slip comments followed closely the suggested list.

Method of coding.—It was apparent that grade-slip comments could not be employed individually because there were too many different remarks employed (three to four hundred). In addition, most individual comments were too rarely used to yield a reliable score. Since it seemed probable that the bulk of the comments could be accounted for in terms of a relatively few independent factors, the problem became one of combining various comments into categories relatively homogeneous internally and independent of one another. The first step was to develop an 81-item code which partly on the basis of association with certain maneuvers and partly on an *a priori basis* was regrouped under one of the following categories: motor technique, perception, motivation, headwork or emotional difficulties. The final code was developed in terms of a breakdown of these items, made in accordance with their relation to graduation-elimination. The first 10 categories in table 5.11 are the ones which were formed on the basis of the findings from a preliminary study. The results of this preliminary study are presented in tables A5.1 and A5.2 in the appendix.

The number of comments under each code or category was divided by the number of grade slips for the individual. Thus the final category score represented the average number of comments per grade slip for a given category of comments.

No weighting was made of the comments on the basis of content, intensity, or the letter grade which it accompanied.

Results.—In table 5.11 are presented the results of this later study. These data indicated that the category scores derived from grade slip comments offered possible differential criteria; the intercorrelations among the comments were not as high as those among the letter grades on the grade slips.

The category scores predicted success in Primary pilot training, the most promising yielding a correlation of over 0.40. There was considerable variation, however, in the size of the correlations for the three populations studied. All of these correlations are, of course, somewhat spuriously high since the instructor who wrote the comments on the grade slips also made the recommendation that the poor students should be sent up for elimination check rides and usually discussed the students with the check pilot before and after he flew with them. The correlations with success in the next phase of training, Basic, are free from this factor and are considerably lower. When the result of the three populations studied are combined, the highest correlations with success in Basic, are in the neighborhood of 0.25. It will be remembered that the average check-ride grades and the proficiency card ratings in Primary also were found to have about this same degree of relationship to success in Basic.

Factor analysis.[17]—Intercorrelations between the grade slips category scores, Primary check ride, selected tests, and the pilot stanine for the November 1943 Aviation Cadet Classification Battery are given in the appendix in table A5.3. These intercorrelations were subjected to a centroid factor analysis and 11 rotated factors were isolated. The rotated factor loadings are given in table 5.12 below. The factors which they isolated were described by them in the following way:

Rotated Factor I: The pilot stanine was included in the matrix of variables because of interest in discovering the loadings of the stanine in any new factors which might be discovered. It was realized that, in general, the practice of introducing such a *linearly dependent* variable into a matrix transforms specific and error variance into common variance. This results, theoretically, in the necessity of extracting an additional centroid for each other variable. It was believed, however, that it was sufficient for the purposes of this preliminary study to have but one factor best defined by the linearly dependent variable. Rotated Factor I is this factor. It emerged very clearly in the analysis.

Rotated Factor II.—This is obviously a "Halo" Factor and has no significant loadings for any variables not derived from the instructor's comments.

Rotated Factor III.—This is the Numerical Factor commonly found in analyses of the Classification Battery.

Rotated Factor IV.—This factor is the Reasoning Factor found in other analyses.

Rotated Factor V.—Spatial Relations Factor. (Previously isolated.)

Rotated Factor VI.—Verbal Factor. (Previously isolated.)

Rotated Factor VII.—Perceptual Factor. (Previously isolated.)

Rotated Factor VIII.—Mechanical Experience Factor. (Previously isolated.)

---

[17] This factor analysis was performed by Lt. John I. Lacey and Sgt. Harold H. Singer.

TABLE 5.11.—*Intercorrelations of categories of primary grade slip comments and their relationship to measures of proficiency in basic*

Primary Pilot Classes at Stamford and Bonham

[44-H (Jul. 43 Battery N=158), 44-H (Nov. 43 Battery N=169), and 44-K (Nov. 43 Battery N=318)]

	A[1]	B	C	D	E	F	G	H	I	J	ABC	DE	GH	Primary average check-ride	Primary P-F[2]	Basic average check-ride	Basic instrument check-ride
A. *Motor technique*—Low correlation[3] with pass-fail	......	38	54	32	39	19	53	49	22	40	81	44	59	47	56	24	−02
	......	28	51	28	36	35	49	52	27	46	81	42	59	27	51	29	15
	......	35	54	36	31	25	49	50	16	42	81	42	58	32	44	13	01
B. *Motor technique*—Negative correlation with pass-fail	38	......	34	13	23	09	30	26	03	19	74	22	32	07	−09	02	−18
	28	......	23	28	24	30	20	33	09	21	66	34	31	05	−17	−01	06
	35	......	37	17	21	04	35	21	18	23	73	24	32	08	07	04	09
C. *Motor technique*—High correlation with pass-fail	54	34	......	31	39	22	47	54	37	40	80	43	58	38	52	23	−04
	51	23	......	24	43	34	39	43	18	28	77	44	48	20	22	24	−02
	54	27	......	39	38	20	48	50	37	50	81	48	58	25	43	09	07
D. *Perception*—Negative	32	13	31	......	33	16	31	46	31	24	32	81	44	09	22	14	19
	28	28	24	......	18	19	17	53	33	16	36	76	41	04	−17	11	01
	36	17	39	......	25	12	40	38	26	27	39	79	46	22	34	06	−03
E. *Perception*—Positive	39	23	39	33	......	55	31	45	21	23	43	82	44	41	57	25	04
	36	24	43	18	......	48	15	46	30	10	46	77	36	14	11	12	11
	31	21	38	25	......	26	23	38	42	32	38	81	36	36	57	05	03
F. *Motivation*	19	09	22	16	55	......	36	43	47	09	21	44	45	37	34	28	31
	35	30	34	19	48	......	06	41	61	29	44	44	29	45	42	30	−06
	25	04	20	12	26	......	24	44	38	21	21	24	39	32	50	13	12
G. *Headwork*—Low	53	30	47	31	31	36	......	55	10	34	55	38	88	32	31	19	06
	49	20	39	17	15	68	......	47	20	35	48	21	86	25	33	18	20
	49	35	48	40	23	24	......	46	20	38	56	39	85	24	39	23	10
H. *Headwork*—High	49	26	54	46	45	43	55	......	52	42	55	56	88	45	55	35	19
	51	33	43	53	46	41	47	......	54	38	57	64	86	26	22	19	14
	50	21	50	38	38	44	46	......	35	43	52	48	85	37	66	13	19
I. *Emotional difficulties*	22	03	37	31	21	47	10	52	......	34	26	32	36	17	20	15	04
	27	09	18	33	30	61	20	54	......	21	24	41	43	38	32	20	21
	16	18	37	26	42	38	20	35	......	30	30	42	32	19	28	−07	13
J. *Rough, abrupt, tense use of controls*	40	19	40	24	23	09	34	42	34	......	42	29	43	36	45	25	−08
	46	21	28	16	10	29	35	38	21	......	43	17	42	35	40	26	01
	42	23	50	27	32	21	38	43	30	......	49	37	48	32	53	09	11
A+B+C—Motor technique	81	74	80	32	43	21	55	55	26	42	......	46	63	38	42	30	−11
	81	66	77	36	46	44	48	57	24	43	......	53	62	24	26	24	09
	81	73	81	39	38	21	56	52	30	49	......	49	63	28	39	34	07
D+E—Perception	44	22	43	81	82	44	38	56	32	29	46	......	54	31	49	−09	14
	42	34	44	76	77	44	21	64	41	17	53	......	50	12	−04	15	08
	42	24	48	79	81	24	39	48	42	37	49	......	51	36	56	07	−00

See footnotes at end of table.

TABLE 5.11.—*Intercorrelations of categories of primary grade slip comments and their relationship to measures of proficiency in basic*—Cont'd

Primary Pilot Classes at Stamford and Bonham

[44-H (Jul. 43 Battery N=158), 44-H (Nov. 43 Battery N=169), and 44-K (Nov. 43 Battery N=348)]

	A	B	C	D	E	F	G	H	I	J	ABC	DE	GH	Primary average check-ride	Primary basic average		Basic instrument check ride
															P-F	Check ride	
G+H—Headwork	59	32	58	44	44	45	88	88	36	43	63	54	......	44	48	31	14
	59	31	48	41	36	29	86	86	43	42	62	50	......	30	32	21	21
	58	32	58	46	34	39	85	85	32	48	63	51	......	35	61	21	17
Primary average check ride	47	07	38	09	41	37	32	45	17	36	38	31	44	......	1.05	28	12
	27	05	20	04	14	45	25	26	38	35	24	12	30	......	95	19	12
	32	08	25	22	36	32	24	37	19	32	28	36	35	......	86	30	−14
Basic average check ride	24	02	23	14	25	28	19	35	15	25	20	−09	31	28	......	......	09
	29	−01	24	11	12	30	18	19	20	28	24	15	21	19	......	......	34
	13	04	09	06	05	13	23	13	−07	09	34	07	21	30	......	......	−21
Basic instrument check ride	−02	−18	−04	19	04	31	06	19	04	−08	−11	14	14	12	......	09	......
	15	06	−02	01	11	−06	29	14	21	01	09	08	21	12	......	34	......
	01	09	07	−03	03	12	10	19	13	11	07	−00	17	−14	......	−21	......
Primary pass-fail	56	−09	52	22	57	34	31	55	20	45	42	49	48	1.05	......	......	......
	51	−17	22	−17	11	42	33	22	32	40	26	−04	32	95	......	......	......
	44	07	43	24	57	50	39	66	28	53	39	56	61	86	......	......	......

[1] The top coefficient in each group is for 44-H (Jul. 43-Battery); the middle one for 44-H (Nov. 43 Battery) and the bottom one for 44-K (Nov. 43 Battery). Decimal points have been omitted.

[2] Biserial correlation of comments with Primary pass-fail corrected assuming unrestricted standard deviation of 1.87 for pilot stanine (without experience credit).

[3] Comments within each category regrouped on the basis of correlation with pass-fail as shown in preliminary study.

Rotated Factor IX.—This factor has not been isolated in previous analyses and has been tentatively labeled as the Motor Tenseness Factor. It is interesting to note that the Rudder Control Test (CM120B) has the highest loading for any test on this factor and also has its own highest loading on the factor. Also, Primary average check ride has its highest loading on the Motor Tenseness Factor. The factor is best defined by the comments in the categories of "Motor Technique" and "Rough, Abrupt, or Tense Use of Controls."

TABLE 5.12.—*Rotated factor loadings of grade slip comments*

[Based on intercorrelations presented in table A5.3]

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	[illegible]
1. Motor technique-grads and elims	03	57	01	09	12	01	−04	12	45	−04	−11	59
2. Motor technique-grads	−03	48	01	−03	−03	06	−03	01	09	−04	−14	27
3. Motor technique-elims	−01	63	12	03	01	14	09	14	29	05	−04	56
4. Perception-grads and elims	−14	52	−10	05	07	05	−04	20	00	07	17	39
5. Perception-elims	−02	40	19	10	−03	−10	13	16	17	33	−07	40
6. Motivation-elims	12	25	00	−04	00	−05	00	15	17	61	−05	52
7. Headwork-grads and elims	−04	52	−10	06	29	20	−06	04	31	−05	07	53
8. Headwork-elims	07	58	05	23	19	−04	−04	15	26	26	16	62
9. Emotional difficulties-elims	00	33	03	00	16	01	00	19	03	54	11	49
10. Rough, abrupt, or tense controls	01	38	−05	02	08	06	03	08	45	15	00	40
11. Elementary check-ride	16	12	−08	18	18	05	03	09	42	34	00	42
12. Biographical data-navigator	−07	−06	23	−11	−13	26	13	09	14	02	16	23
13. Biographical data-pilot	23	−02	−07	−12	−13	07	14	39	10	08	11	30
14. Spatial orientation I	04	−03	−05	02	35	12	65	02	11	−05	05	63
15. Spatial orientation II	10	−01	08	21	27	11	55	14	−01	−02	−07	47
16. Reading comprehension	00	−05	20	35	21	61	−03	01	00	−01	01	60
17. Dial and table reading	00	05	44	11	40	20	31	−15	04	−02	11	55
18. Finger dexterity	−11	−11	09	−08	37	−08	−04	20	05	−09	00	24
19. Two-hand coordination	00	−02	−06	12	56	−15	−10	42	−02	06	−10	56
20. Discrimination reaction time	05	−09	15	11	47	06	14	−01	12	12	17	35
21. Complex coordination	20	−02	16	−04	57	−04	03	21	11	−09	16	49
22. Mechanical principles	04	−04	00	46	25	25	10	41	12	02	−04	54
23. Instrument comprehension II	17	04	16	03	48	26	31	00	17	00	−10	49
24. General information	27	−02	00	37	12	38	−01	27	12	12	−16	60
25. Rudder control	16	−01	01	18	12	−15	04	09	28	−09	04	19
26. Rotary pursuit	02	05	−05	−14	40	01	−05	39	06	−09	−11	37
27. Mathematics B	03	00	42	43	12	39	08	00	06	00	−01	55
28. Pilot stanine	63	−02	05	13	58	17	22	53	10	−01	14	1.14

Rotated Factor X.—This is another new factor and is best defined by its high loadings for the *Motivation* and *Emotional Difficulties* categories. It should be noted that Primary average check ride has its second highest loading on this factor while none of the Classification Tests have any appreciable loadings on it.

Rotated Factor XI.—Residual Factor.

As can be seen from these results, the use of the category scores in a factor analysis appears to yield a possibility of new insights into the nature of both the over-all pilot criterion and the Classification Tests themselves. Apparently, two sizable independent components of Primary pilot proficiency have been isolated in Rotated Factors IX and X. Classification Tests have very low loadings on these factors, so these factors may be regarded as promising areas for new test development. Since proper combinations of the category variables in table 5.13 would apparently yield relatively homogeneous measures of the *Motor Tenseness* and *Motivation* Factors, these combinations of the category variables should be used as the criteria against which to validate and refine experimental tests designed to measure the two new factors.

PILOT RATING SCALE, FORM C

The Pilot Rating Scale was developed originally in the Office of the Air Surgeon, Hq. Army Air Forces, to provide specific criteria of different aspects of pilot performance for use in validating classification tests. The items for the scale were selected after an analysis of the faculty board proceedings for 1,000 cadets who failed Primary pilot training. The scale passed through two revisions, and a large number of students were rated on it.

Correlation with stanines and Primary pass-fail.—Table 5.13 presents the intercorrelations between instructors' ratings made after 8–10 hours of training, pilot stanines, and graduation-elimination.[*]

TABLE 5.13.—*Correlation of ratings on pilot rating scale (Form C) with pilot stanine and with primary pass-fail*

[For 1900 Cadets Without Previous Flying Experience in Class 43-C, EFTC]

	1	2	3	*M*	*SD*	M_g	M_e
1. Ratings	....	0.28	0.81	82.98	17.66	89.47	65.27
2. Stanine	0.28	....	.33	5.46	1.76	5.73	4.73
3. Pass-fail	.81	.33	....	(P_g = .73)	....	....	....

It can be seen from these data that ratings given by instructors after 8-10 hours of training were more closely related to graduation-elimination in primary than were pilot stanine scores. However, consideration should be given the fact that for about half the eliminees the ratings were made at the time of their elimination, and that in all cases the ratings were made by the same instructors who were influential in deciding whether or not a student should be graduated or eliminated.

It is also evident that the stanines appeared to be somewhat more closely related to graduation than they were to instructors' ratings.

Prediction of performance in basic.—In another study by the Psychological Section, Office of the Surgeon, Hq., AAF Training Command, ratings made in primary training were correlated with those in basic for 435 students in class 43-D. The scores used for this cor-

[*] These data were published by Hq. AFTAS on the basis of material submitted by Psychological Research Unit No. 1.

relation were the sum of all items on the Pilot Rating Scale (Form C). The results are presented in table 5.14.

TABLE 5.14.—*Correlations of ratings on pilot rating scale (Form C) at primary with pilot stanine and with ratings at basic*

	N	*M*	*SD*	Correlation with the basic rating [2]
First primary rating [1]	435	91.52	13.27	0.25
Second primary rating [2]	391	96.30	14.00	.33
Pilot stanine	435	6.08	1.97	.31
Final basic rating (all cases) [2]	435	84.56	23.84	....
Final basic rating (only cases with final primary rating)	391	88.47	21.51	....

[1] Sum of all items on scale administered at 8-10 hours of training.
[2] Sum of all items on scale administered at graduation or elimination.

Like all of the other correlations between different phases of training, the coefficients are low. Ratings made at the 8–10 hour level and at graduation or elimination from Primary correlated 0.25 and 0.33 respectively with those made at graduation or elimination from Basic.

For purposes of comparison, the correlation of the pilot stanine with the ratings in Basic is included in table 5.14. It can be seen that the stanines, which were based upon aptitude tests administered prior to entrance into preflight training gave as accurate a prediction of flying proficiency in Basic as did instructors' ratings based on the *whole course of Primary flying training.*

Differential criterion.—If correlations of tests with measures of specific types of pilot performance were available, it might be possible to identify independent aspects of flying performance and develop improved methods of predicting each of them. Psychological Research Unit No. 2 investigated the possibility of using the Pilot Rating Scale in this manner.

Intercorrelations of the items in Form A of the Pilot Rating Scale were computed on 369 students in class 42–K who were rated between the eighth and tenth hours of Primary training. These intercorrelations are presented in table A5.4 of the appendix. They were high. It was not possible to determine how much of these high intercorrelations was due to "halo effect" and how much was the result of true association between the variables.

On the basis of an inspection of the intercorrelations, the items were arranged into clusters, shown with their intercorrelations in table 5.15.

In general the intercorrelation among clusters of items is fairly high being somewhat lower with items in Cluster VI (Attitude).

In table 5.16 are presented the correlations of the Pilot Rating Scale (form C) with Primary pass fail. From this table it is apparent that all the items on the rating scale show a substantial correlation with pass fail in Primary school. This is not surprising since the instructor who made the ratings also played an important role in determining whether the student passed or failed and since a number of the ratings were made after the instructor knew that his student had been eliminated. It may be noted that the average correlations of items in Cluster I (Headwork), and II (Perception) are greater than those of the other clusters. Cluster VI (Attitude) showed the lowest predictive efficiency.

Conclusions.—Ratings made in Primary flying training did not predict subsequent success in Basic any better than did stanines based on aptitude tests given before the student entered preflight training. Ratings on the individual items showed high intercorrelations, indicative of considerable "halo effect" and these characteristics made this scale of little use as a source of differential criteria. On the basis of these facts it was decided to discontinue work on the Pilot Rating Scale.

TABLE 5.15.—*Average intercorrelation among clusters of items on pilot rating scale (form A)*

[Based on 369 students in class 42-K rated between the 8th and 10th hours of primary training]

	I	II	III	IV	V	VI
Cluster I (Headwork)[1]	0.67	0.63	0.54	0.59	0.60	0.49
Cluster II (Perception)	.63	.76	.52	.66	.55	.40
Cluster III (Orientation)	.54	.52	([2])	.47	.51	.39
Cluster IV (Coordination)	.59	.66	.47	.76	.50	.39
Cluster V (Emotion)	.60	.55	.51	.50	.72	.48
Cluster VI (Attitude)	.49	.40	.39	.39	.48	.61

[1] The next table, 5.16, gives the list of items included in each cluster and Appendix table A5.4 gives the complete matrix of intercorrelations from which the within-cluster (e. g., I v. I) and between-cluster (e. g. I v. II) averages were computed.
[2] Since there was only one item in this cluster, no within-cluster correlation could be computed.

INVESTIGATION OF PILOT TRANSITION TRAINING RECORD

The Pilot Project [29] investigated the Transition Training Record which was available in four-engine schools, for possible use as a source of proficiency data to employ in selecting lead-crew pilots, as a criterion against which to evaluate classification tests, and as a guide in the construction of objective measures of flying skill. The pilot Transition Training Record was made up of nine major divisions designed to include all phases of transition training: day transition, night transition, instrument flying, formation flying, altitude flying, day navigation, night navigation, additional flying instruction, and ground instruction. Each major section was subdivided into items as illustrated below:

Formation Flying:

a. Assembly from single plane take-off.

b. Relative position in 3-plane V.

Straight and level.

Turns—10°, 20°, and 30° bank.

[29] Most of this work was done by Sgt. John R. Rohrs.

c. Relative position in 3-plane echelon.

Straight and level.

Turn-away from echelon.

d. Landing

Assuming proper spacing from 3-plane.

V and echelon.

In all there were 102 such items. On each item the students were rated by their instructors on a three-category scale: "good," "fair," and "poor." For the purposes of this study only the 82 items were used which were rated on at least 90 percent of the transition records.

In order to have a criterion against which to evaluate the transition record, the supervisory personnel in each squadron of the school were asked to divide their students into the best quarter, the second best quarter, the third best quarter, and the poorest quarter. Ratings by the supervisory personnel were based on an analysis of the daily grade slips and conferences with the instructors.

TABLE 5.16.—*Correlation of items on pilot rating scale form C with primary pass-fail* [1]

Class 44-A $N_t=487$ $P_s=.71$

Class 44-B $N_t=990$ $P_s=.76$

Key	Class 44-A				Class 44-B					
	M_s	M_e	$S.D._t$	r_{bis}	M_s	M_e	$S.D._t$	r_{bis}	r_a [3]	
1. Judgment (I)[2]	4.34	2.97	1.17	0.70	4.15	2.97	1.03	0.67	0.68	1
2. Foresight and planning (I)	4.23	2.85	1.20	.69	4.03	2.94	1.08	.59	.64	2
3. Memory (I)	4.46	3.06	1.24	.68	4.32	3.03	1.17	.64	.66	3
4. Comprehension (I)	4.46	3.21	1.13	.66	4.40	3.47	1.06	.52	.59	4
5. Visualization of course (I)	4.19	2.79	1.19	.71	4.11	2.97	1.09	.61	.66	5
6. Estimation of speed and distance (I)	4.15	2.92	1.16	.64	4.07	2.99	1.05	.61	.62	6
7. Sense of sustentation (II)	4.28	2.88	1.18	.71	4.15	2.98	1.12	.61	.66	7
8. Division of attention (I)	4.19	2.91	1.19	.65	4.04	2.99	1.10	.56	.60	8
9. Orientation (III)	4.38	3.48	1.08	.50	4.35	3.48	1.07	.50	.50	9
10. Speed of decision and reaction (I)	4.17	2.92	1.17	.65	4.04	2.99	1.08	.57	.61	10
11. Coordination (IV)	4.25	3.12	1.09	.62	4.10	3.18	1.00	.54	.58	11
12. Appropriateness of controls used (IV)	4.31	3.11	1.07	.66	4.16	3.36	0.91	.50	.59	12
13. Feel of controls (II)	4.28	2.96	1.14	.70	4.12	3.07	1.10	.61	.66	13
14. Smoothness of control movement	4.18	3.08	1.15	.57	3.96	3.14	1.06	.45	.51	14
15. Progress in developing technique (I)	4.31	2.28	1.23	.73	4.09	2.88	1.11	.64	.69	15
16. Absence of tenseness (V)	4.57	3.36	1.22	.59	4.37	2.37	1.13	.52	.56	16
17. Absence of confusion and nervousness (V)	4.54	3.35	1.17	.61	4.41	3.41	1.09	.54	.57	17
18. Absence of fear and apprehension (V)	4.84	3.82	1.11	.55	4.68	3.88	1.03	.45	.50	18
19. Suitable temperament (VI)	4.84	3.84	1.11	.55	4.73	3.92	1.01	.47	.51	19
20. Motivation and attitude (VI)	5.15	4.41	1.10	.41	5.08	4.39	1.00	.41	.41	20
21. Probable success	4.50	2.99	1.20	.76	4.36	3.07	1.11	.68	.72	21

[1] Students were rated on the Pilot Rating Scale between the fourth and twentieth hours of Primary training. The ratings and pass-fail were not independent measures since the instructors who made the ratings also played an important role in determining whether or not the student was passed or failed; furthermore, a number of the ratings were made after the instructor knew that his student had been eliminated.

[2] Key to cluster in which each item was included:

I = Headwork IV = Coordination

II = Perception V = Emotion

III = Orientation VI = Attitude

[3] r_a = Average correlations for both samples obtained through Fisher's z transformation technique.

A total score for the transition record was secured by assigning a weight of plus 1 to a rating of "good," zero to "fair," and minus 1 to "poor." The weights on all items were algebraically summed. For

237 students in classes 44–E and F the correlation between this total score of the Transition Record and the squadron rankings was only 0.31.[30] It was not possible for both of these measures to be good, since the correlation between them was low.

On the assumption that the better of these measures would be more likely to correlate with other factors, the relationship between each of them and the unaugmented pilot stanine was determined. For 181 students the correlation between the transition records and the stanine was only 0.01. For 177 students the correlation between squadron ranking and pilot stanine was 0.21. The difference between these correlations is significant at the 2-percent level.[31] On the assumption that more reliable and valid measures of proficiency will have higher correlations with other variables, one may conclude that the squadron rankings are better than the Transition training records.

Finally, two item analyses were made of the Transition training records. The first of these was on the basis of the upper and lower 27 percent of the Transition records rank-ordered by composite score and the second was on the basis of the highest and lowest quarters separated according to the ratings of the supervisory personnel. Discriminatory differences between the items were similar for both analyses and the grouping of the items clearly showed that there was so much "halo effect" that an item analysis could not be of much value. Furthermore, the grade of "poor" occurred so rarely that the rating was in effect reduced to the dichotomy of "good" and "fair." A total score was obtained by rescoring the Transition records on the basis of the twenty most discriminating items from the analysis against supervisors' rankings, but the correlation between this refined score and the pilot stanine was only 0.07 which is not reliably different from the results obtained by using the original score.

On the basis of these results it was decided that the Pilot Transition Training Record was not a useful measure of proficiency.

SUMMARY

In general, the correlations among subjective grades for different aspects of flying were high indicating the probable presence of a strong halo effect. These high correlations limited the possibility of using the various grades differentially, either in selecting students for different types of specialized training or as specific criteria for the separate validation for the different types of aptitude tests. The intercorrelations among scores derived from instructors' comments on student errors were considerably lower so that it seemed that there would be

[30] All correlations involving the supervisors' division of the class into quarters were corrected for broad categories for a rectangular distribution by table No. 33, p. 398 of Peters and Van Voorhis, *Statistical Procedures and their Mathematical Bases.*

[31] Computed by the formula for the difference between correlations from the same or correlated populations; E. F. Lindquist, *Statistical Analysis in Educational Research*, p. 218.

some possibility of getting more specific measures of different aspects of flying performance by developing a check-list of common errors. This was supported by the fact that factor analyses of scores derived from the comments isolated several independent factors. Some of these were predicted by the present battery of classification tests and others were not predicted, therefore suggesting areas in which new tests should be developed.

Within the same school, successive grades by instructors and check riders were not made independently; the rater almost always took previous grades into account before arriving at a final decision. Thus the correlations between such grades were spuriously high. When grades given at one school were compared with those at the next more advanced school, the evaluations were relatively independent. Under these conditions the correlations were low, in the range of 0.20 to 0.35. In fact, grades based upon the whole course of Primary flying instruction did not predict grades at the Basic level any better than did aptitude tests administered before the student entered preflight training. The unreliability of the subjective flying grades probably was one of the factors causing the correlations between grades at different levels of training to be so low. It is also possible that students who were slow to catch on at first did better in later stages of training; that a different level of training demanded different abilities from the students; or that some students changed their motivation or attitude. Whatever the reason for these low correlations, the fact that the grades at one level of training did not predict those at the next level is of considerable practical importance in that it suggests that a number of students eliminated in Primary might have received satisfactory grades if they had been allowed to proceed to Basic.

703322—47——9

CHAPTER SIX

Objective Measures of Flying Skill for the Primary Level of Pilot Training

Capt. Richard P. Youtz[33]

INTRODUCTION

Importance of Research at the Primary Level

Primary school was a crucial phase of pilot training. In Primary most students received their first flying instruction, became familiar with the fundamental maneuvers of flight, and exhibited their aptitude or lack of it. Of elimination for lack of flying proficiency in Primary, Basic, and Advanced schools, considerably more than half occurred at the Primary level. Definite objective standards were particularly desirable in Primary to select those students whose skill warranted further training and to eliminate early those students who would not be able to meet the high proficiency requirements of combat flying. For these reasons the Primary phase was chosen for intensive investigation.

The Primary School Curriculum

The maneuvers it was possible to measure in Primary schools were fairly well defined by the curriculum and customary practices of the schools. Because these measures of flying skill were designed for use with the regular Primary curriculum, it was decided to confine the measures devised to those maneuvers regularly taught at Primary school. If measures had been tried out on maneuvers which were not taught at Primary schools, the data would not have been applicable to any later student population which was routinely taught these maneuvers. In the first experimental investigation of measures of flying skill at a Primary school, one maneuver was measured which the

[33] Of the seven studies completed on objective measures at the Primary level for which the writer was responsible, the following people participated in the planning and execution of the field trips: Sgt. M. F. Connery in all seven; Lt. J. F. Kamman in Studies I, II, and III; Sgt. J. R. Rohrs in Study V; Sgt. J. J. Waeltermann in Study VI; and S/Sgt. W. W. Ismael in Study VII. Lt. R. E. Showalter supervised the test administration in Study VI and had charge of the field trips in Studies IV and V.

The statistical work on these studies was done by Lt. J. F. Kamman, Sgt. J. R. Rohrs, and Sgt. J. J. Waeltermann under the technical supervision of Lt. J. K. Hemphill and T/Sgt. W. O. Matheny. S/Sgt. C. P. Gershensen supervised the statistical analysis of the data in Study VII, the Primary section of which involved 1,400 cases. Valuable assistance and advice were received from Prof. Phillip J. Rulon, Acting Dean of the Graduate School of Education, Harvard University.

students had not previously practiced. It was found that large practice effects were present which radically changed the discriminating power of the measures from the first to the second administration. The decision to avoid maneuvers not regularly taught did not exclude the use of new and ingenious methods of measuring manuevers that students had already practiced, nor did it prevent the tester from measuring parts of maneuvers separate from the whole maneuver; it did mean that a careful examination of the curriculum and practices of Primary schools was necessary in making up measures of flying ability for administration at this level.

While the curriculum varied slightly from school to school, the fundamental elements remained the same. Table 6.1 was made up after examination of the curricula of three Primary schools and shows the maneuvers introduced during each successive week of training together with the number of hours completed by the end of each week. The name of the maneuver is followed by a footnote reference when personnel of the Pilot Project devised, tried out, and used or discarded one or more measures for that particular maneuver. It will be noticed that almost all of the maneuvers before solo were measured. This was

TABLE 6.1.—*New maneuvers introduced during each week of primary training*

Week No.	Flying hours at end of week	New maneuvers introduced during week
1	4	Straight and level,[1] use of trim tabs,[1] gentle [1] and medium [1] turns, effect of torque,[1] climbs,[1] climbing turns,[1] glides,[1] gliding turns,[1] taxiing, rectangular courses,[1] 8 turns.[1]
2	8	Coordination exercises,[1] stalls,[1] spins [1] and spin recovery,[1] forced landing,[1] take-offs,[1] landings,[1] wheel landings, crab [1] and slip [1] on final approach, crosswind landings,[1] steep turns,[1] traffic pattern,[1] recovery from bounce landings,[1] recovery from zoom landings,[1] drift-correction.[1]
3	14	Stalls and spins from unusual positions, control-timing exercise (supervised solo takes most of time during this week).
4	22	90° accuracy approach,[1] elementary 8's, rudder-exercise stalls,[1] chandelles.
5	30	180° side approach.
6	39	Loops,[1] pylon 8's, lazy-8's, 360° overhead approach.
7	48	Half-rolls, snap-rolls, slow-rolls, power-on approaches with maximum performance power glide.
8	57	Immelmann turns,[1] half-roll with roll-out recovery.
9 & 10	65	Review of all previous maneuvers.

[1] Maneuvers on which one or more objective measures have been devised, tried out, and used or discarded.

FIGURE 6.1.—*Percent of eliminations occurring at successive five-hour intervals during training in primary schools* [1]

Hour level	Percent of eliminees	Hour level	Percent of eliminees
1–5	0	36–40	7.5
6–10	10	41–45	5.5
11–15	31	46–50	5
16–20	11	51–55	2.5
21–25	7	56–60	3
26–30	7	61–65	2.5
31–35	7.5		

[1] Number of students eliminated=[illegible]; Classes 45 B, C, & D, AAF Central Flying Training Command.

true because later maneuvers tend to be made up of combinations of earlier, more fundamental maneuvers. Work was concentrated in the early part of Primary training because most of the eliminations occurred during that time.

Hours to Elimination in Primary School

The number of hours of flying training at which most eliminations occur is important in the development of a check-ride for use in Primary schools. If objective measures of flying skill are to be useful in standardizing the elimination of students, the measures used must be applicable at the stage of training at which the greatest number of eliminations occur. Only then can objective measures be used to set standards of elimination. In validating objective measures against the graduation-elimination criterion, it is of course also desirable to administer the measures before the bulk of the eliminations occur in order to avoid restriction of range. Figure 6.1 shows the percent of eliminations occurring at successive 5-hour intervals during training in Primary schools in the Central Flying Training Command in classes 45-B, 45-C, and 45-D, combined. There were 855 students eliminated and the figures shown are percents of these 855 eliminations.

From figure 6.1 it can be seen that 41 percent of the eliminations occurred before the fifteenth hour and 66 percent of the eliminations occurred by the thirtieth hour. It was desirable, therefore, to test the students as early as possible, but not so early that they had not had an opportunity to practice a representative sample of the basic maneuvers taught in Primary school. Table 6.1, showing the week and hour of introduction of new maneuvers, indicates that most of the fundamental maneuvers have been introduced before the fourteenth hour. An attempt was made to concentrate most of the work at a period which was late enough so that the students would have practiced enough maneuvers to afford a representative sample of the fundamental elements of flying skill, and early enough in order to include as many potential eliminees as possible. These considerations placed the ideal time for research at approximately the 8–12 hour level, when the students were just beginning to solo.

STEPS IN THE DEVELOPMENT OF OBJECTIVE MEASURES

In this new field of objective measurement of flying skill, a number of steps were necessary to develop relevant and meaningful measures. Throughout the development of each measure this Project benefited by the advice and cooperation of the members of the Primary Training Advisory Board. This organization was attached to the AAF Central Instructors School (Pilot) and had responsibility for the supervision of the Standardization Boards of Eastern, Western, and Central Flying Training Commands. Its members also made recommendations for the changes in the Primary school curriculum and were responsible for the contents of the Primary Flying Manuals used by students and instructors.[11]

In attempting to devise objective measures applicable to this new area of flying skill, the procedure was first to choose for analysis an important aspect of flying skill in conference with the Primary Training Advisory Board, then to prepare preliminary record forms for these measures and to try them out in flight. In practice this meant that pilots from the Primary Board and personnel from this Project alternately performed maneuvers and recorded measures of these maneuvers in an attempt to achieve the most effective form of measurement.

After these preliminary trials, the items were taken when possible to a nearby Primary school and a few students tested with the measures to make sure that the administrative instructions were clear and the measures really flyable by students at that level. The measures were then tried out on approximately 90 students who were divided into criterion groups for a first rough determination of validity and with a retest administration for determining reliability. After each measure was tried out, it was frequently necessary to make further changes as a result of the experience gained. Many measures were eliminated in the preliminary trial stage and in conference with the Primary Board because objective measurement of a particular aspect of a maneuver was not possible. This was due to the fact that necessary special recording equipment was not available in the primary trainer, or because the particular measure was too susceptible to the effects of differences between planes or changes in flying conditions. The stages in the development of a typical measure are illustrated below.

Example of Stages in the Development of a Typical Measure

In the initial analysis of the critical aspects of flying skill, it was decided in cooperation with the Primary Board that, in most maneuvers, maintaining a constant altitude is an important aspect of flying skill. All Primary trainers were equipped with a sensitive altimeter. Furthermore, maintaining a constant altitude was not greatly affected by turbulence, and plane and engine differences entered into this factor hardly at all. Of the many maneuvers in which altitude control can be measured, the 360° steep turn seemed likely to provide a distribution of abilities, since it is more difficult to maintain a constant altitude when the plane is banked steeply in a turn than in level flight.

In order to crystallize the discussion on essential flying skills, forms

[11] The members of the Primary Board who worked with this Project are Majors Chester Brown, Chairman, John E. Bowen, Harry H. Culler, and V. C. Denton, and Captains Earl D. Caton, and Philip B. Lockwood.

were made for each of the measures, showing the ability to be measured, the task to be performed, the measure to be recorded, the conditions under which the maneuver was to be performed, and a brief statement of procedures. The following form is for altitude control in a 360° steep turn.

Form 1

Ability: To maintain a given altitude in a steep 360° turn.

Task: To execute a steep 360° turn.

Measure: Amount of altitude gained or lost. Time in seconds to complete turn (as a measure of rate of turn and therefore of bank; a control condition to insure that the student banks as steeply as his power setting will allow).

Conditions: Cruising throttle (r. p. m. 2100 in the Fairchild Primary Trainer, PT-19) throughout turn. Altitude, 3,000 feet.

Procedure: The administrator takes the plane up to 3,000 feet and turns it over to the student, instructing him to execute a steep 360° turn leaving the throttle at cruising. He records the amount of altitude gained or lost and the time in seconds from the beginning of roll-in to the end of roll-out.

It will be noted that in the *Conditions* the throttle setting is "cruising," which would give an r. p. m. of 2100 in the cruising attitudes. Under *Measures*, altitude is measured but also the *Time in seconds to complete the turn*. Since the Primary trainer has no artificial horizon by which bank can be measured and has no rate-of-turn indicator, it was necessary to get some such means to force the student to bank the plane up sharply. It was not practicable to measure the angle of bank in terms of reference points on the plane, since the plane in use at that time was the PT-19 (Fairchild), a low-wing monoplane with no very good reference points for exactly determining the angle of bank. The student's task, then, was to complete the turn as rapidly as possible and to maintain altitude throughout the turn. This was difficult because if the student tried to turn too quickly, his bank would be so great that the amount of power available would not allow him to maintain altitude. Nevertheless, he had to bank the plane as sharply as possible in order to turn rapidly.

This is a good example of one problem met in developing objective measures of flying skill. It is frequently necessary to measure several aspects of a maneuver in order to be sure that various conditions of a maneuver are fulfilled without which the measurement of the critical aspect is meaningless. Without some measure of rate of turn or bank, the student would in the extreme case be able to use a gentle or medium bank and hold his altitude much more easily. This would reduce the distribution of scores on the measure and possibly make it worthless.

During preliminary flight testing it was found that cruising throttle did not allow a steep enough bank, so the condition of full-open throttle was added as shown in form 2.

Form 2

Level Turns (360°)

Time	Altitude: Starting	Altitude: Highest	Altitude: Lowest	Over-all evaluation [1]
At cruising rpm				
Left.................... sec.	...... ft.	...... ft.	...... ft.	4-3-2-1
Right.................. sec.	...... ft.	...... ft.	...... ft.	4-3-2-1
At full-open throttle				
Left.................... sec.	...... ft.	...... ft.	...... ft.	4-3-2-1
Right.................. sec.	...... ft.	...... ft.	...... ft.	4-3-2-1

[1] Performance graded on 4-point scale, 4 being high.

Analysis ....
of
Errors ....

In this second form of the measure, it will be noted that an over-all evaluation has been added as well as the full-open throttle condition. Space is also provided for a description of the errors not objectively measured. This was done so that if the subjective over-all evaluation did not agree with the objective measures used, new objective measures could be added or new conditions devised to secure a better total measure of the skills involved in the maneuver.

After this form of the measure had been tried out on several students it was found that two kinds of errors were made which would lower the subjective evaluation of the steep turn but which were not being measured objectively. One was the tendency of the student to keep his head in the cockpit staring at the altimeter with only occasional glances outside the plane to see where he was going and to look out for other planes. This was considered a serious error in a student because every student was told by his instructors to "keep his head on a swivel" and maintain a sharp look-out for other planes. The other unmeasured error was lack of coordination in the turn causing the plane to slip or skid. Since most of the planes did not have ball-bank indicators in them, and since calibrated marks would have to be made on the ball-bank indicator in order to measure objectively the entire slip and skid range, the possibility of obtaining ball-bank indicators for special installation and use was investigated. In order to measure the degree to which the student looked around in the turn an item was added, *Looks around 50% of the time,* Yes .... No .... This was not an objective measure but seemed to be a necessary condition of measurement of the steep turn.

One other change was found necessary. Since *Time to complete the turn* was being measured and might turn out to be valid and useful, either alone or in combination with altitude, it was necessary to control conditions for this time measurement. It was found that some students followed their instructor's teaching and rolled into the steep turn at a slow, steady rate, while others attempted to get the best time score possible by rolling in much more rapidly, sometimes gaining an advantage. Furthermore, it was not considered desirable to encourage students to force the plane into a steep bank too abruptly. Therefore, to avoid teaching the students this bad technique and also to keep irrelevant variability at a minimum, all students were given ample time to roll into the turn and to roll out. Instead of being given a 360° turn the students were given a turn and a half, or 540°. They were allowed 90° to roll into the turn and 90° to roll out of it, with only the center 360° measured. This is shown in the next form. This form was administered twice (for test-retest reliabilities) to 91 students at a Primary school using the PT-19 (Fairchild) training plane.

Form 3

540° Level Turn (Center 360° measured)

At Cruising R. P. M. (*Left*):
- Time ...... sec. Looks around 50%. Yes ... No ...
- Starting Altitude ...... ft.
- Lowest Altitude .... ft.
- Highest Altitude .... ft.

At Cruising R. P. M. (*Right*):
- Time ...... sec. Looks around 50%. Yes ... No ...
- Starting Altitude ...... ft.
- Lowest Altitude .... ft.
- Highest Altitude .... ft.

At Full Throttle (*Left*):
- Time ...... sec. Looks around 50%. Yes ... No ...
- Starting Altitude ...... ft.
- Lowest Altitude .... ft.
- Highest Altitude .... ft.

At Full Throttle (*Right*):
- Time ...... sec. Looks around 50%. Yes ... No ...
- Starting Altitude ...... ft.
- Lowest Altitude .... ft.
- Highest Altitude .... ft.

Comments..
....

Following this study the Training Command decided to shift from the Fairchild Primary training plane (PT-19) to the Stearman (PT-13 and PT-17). The Stearman is a biplane with a third more power than the Fairchild and somewhat different flying characteristics in stalls and spins. While the measures on the steep turn had worked out fairly well on the Fairchild Primary trainer, it was not known whether these same measures would also be effective on the Stearman. Accordingly, when new measures were ready for trial on the Stearman, several of the ones tried out on the Fairchild were included to give an indication of the relationship between the two planes.

Since Primary students are taught to fly "by the seat of their pants" as well as by reference to instruments, a comparison was made at the Stearman Primary school between performance on certain maneuvers with the instruments visible and performance on the same maneuvers with all the instruments (air-speed indicator, altimeter, and tachometer) in the student's cockpit covered. Thus, the only difference in the next form of the measure was that instruments were covered during one of the administrations. Since the tachometer was covered, it was necessary for the test administrator, who had the use of the instruments in his cockpit, to set the throttle for cruising and full-throttle conditions before the student performed the maneuver. In all other respects form 4 was exactly the same as form 3 and is not shown here.

By the time the next group of new measures was ready for administration, it had been determined that the full-throttle maneuver provided better measures than the cruising-throttle maneuver and the cruising-throttle had been dropped. Ball-bank indicators had been obtained and marked off in zones so that the degree of slip or skid in the turn could be measured objectively by the check-pilot. This form is shown below.

Form 5

540° Level Turn (Center 360° measured)

At Full Throttle (*Left*):

Time .... sec. Looks around 50%. Yes.... No...
Starting Altitude .... ft.
Lowest Altitude .... ft.
Highest Altitude .... ft.
Maximum deviation of ball-bank
Zone 1 .... Zone 2 .... Zone 3 .... Zone 4 ....

At Full Throttle (*Right*):

Time .... sec. Looks around 50%. Yes.... No...
Starting Altitude .... ft.
Lowest Altitude .... ft.
Highest Altitude .... ft.
Maximum deviation of ball-bank
Zone 1 .... Zone 2 .... Zone 3 .... Zone 4 ....
Comments ....
....

The 360° steep turn was revised again and used in the Primary section of a large-scale study of the effects of additional flying training, a description of which appears in chapter 10. Before this study the following revisions were made.

As a result of further discussion with experts it was decided that altitude control during the roll-in and roll-out was one of the more important factors of the turn. Therefore, in eliminating this factor in the 540° turn by measuring only the center 360°, an important aspect of the maneuver had been discarded. In the revision this aspect was emphasized by requiring the student to roll directly from a 360° steep right turn into a 360° steep left turn. Since experts were of the opinion that the poorer students might lose their orientation in the two turns, a record of whether or not the student came out within 45° of the original heading was included. This was a possible measure; it also helped to define the conditions more exactly.

After preliminary try-out, it was found possible to dispense with the time measurement and to substitute a report of whether the student maintained a 60° bank in both of the two steep turns, one of the steep turns, or neither of them. It was possible to do this with the Stearman Primary trainer because when the plane is in a 60° bank, the cabane strut, which is easily visible to both the student and the instructor, is parallel with the horizon.

Two measures were discarded. The item on *Looking around 50% of the time* was discarded as not being objective and relatively less important as a condition than had previously been thought. The study in which the students were compared under conditions of instruments covered and visible had shown that watching the altimeter made only a 10% difference in altitude control. The ball-bank deviation measure on this maneuver turned out to be not discriminative enough to be used as a measure and probably not important enough to be used as a limiting condition.

The altitude measure remained as the most important one and for the final form was described in terms of deviation from a starting altitude as shown in Form 6. One of the purposes of the preceding exploratory studies was to secure distributions of altitude variations which could be used to set definite scoring limits. After these distributions had been secured, it was possible to establish definite limits and allow the student to know exactly how he would be scored.

Form 6. The Final Form

Two 360° Steep Turns

(Roll Directly From Left Turn into Right Turn.)

A. Deviation From Starting Altitude:
Check Highest and Lowest.[1]

Lose feet[2] — Gain feet

←120	100	80	60	40	20	0	20	40	60	80	100	120→
0	1	2	3	4	5	6	7	8	9	10	11	12

[1] Scoring: Highest score is 9 points for range in zero interval; 1 point subtracted for each 20 feet of range so that check mark in gain 20 and lose 60 gives score of 5. No score below zero.

[2] Marking: 0 = +10 to −10; 20 = 11 to 30; 40 = 31 to 50, etc.

B. CONDITIONS TO BE FULFILLED: FAILURE IN EITHER OF THESE MEANS ZERO SCORE ON WHOLE MANEUVER.

1. REACHED A 60° BANK IN EACH TURN (CABANE STRUT PARALLEL WITH HORIZON)

Yes	No
1	0

2. FINAL ROLL-OUT WITHIN ±45° OF STARTING HEADING

Yes	No
1	0

No other changes in this maneuver seemed indicated and it was not further tried out in two later administrations of experimental check-rides at Primary schools. The measure has gone through a number of developmental changes, particularly in the *Conditions*, which have been revised and developed to provide a suitable background for the critical measure of altitude control. It will be noted that the throttle setting is no longer specified and that maintaining a steep bank and completing the turns were found to be the most important conditions to be fulfilled.

Other measures followed a similar evolution.—This succession of developmental changes was typical of the measures that showed promise and were retained for further work. Many measures were discarded when administrative or statistical inadequacies were revealed, but the better ones went through an evolution analogous to that of altitude control in the steep turn.

Inventory of Measures Devised for the Primary Level

An inventory of the measures devised is given in the table below. For the purpose of this inventory a separate measure is defined as any observation which was recorded and statistically analyzed separately. Thus, landings might include *point of landing* and *attitude on landing*. Readings on the same instrument are considered to be different when they are taken under different conditions or in different maneuvers. Thus *air speed range* in climbing turns and *air speed range* in gliding turns are different measures. Similarly, *altitude range* in steep turns and *altitude range* in the traffic pattern are two different measures. Minor variations such as *airspeed range* in climbing turns to the right and to the left, however, are *not* considered to be different measures.

TABLE 6.2.—*Number of measures devised and developed at Primary level of training*

Total number of measures devised	256
Measures discarded after preliminary try-out in the air	50
Measures discarded because of the need for special equipment, or not objectifiable at the Primary level	75
Measures retained for try-out on statistically significant numbers of students	131

FIVE ITEM-ANALYSIS STUDIES: PURPOSE AND PROCEDURE

Purpose

Five studies on objective measures at the Primary level of training were performed by the Pilot Project in order to obtain a first coarse screening of items on the basis of the reliability and validity of each individual measure and the experience gained during their administration. This section describes the procedure used in these five screening studies. The following two sections summarize the general results of these screening studies and present a description and evaluation of specific measures.

It was planned that selected measures from the five screening studies would be made into an organized check ride and put to a more severe test in a further study on a larger number of cases. This further study was to verify the reliabilities and predictive powers of these selected measures and also to determine their intercorrelations. It was also planned to obtain a total score so that the validity and test-retest reliability of the total objective check ride could be obtained. The end of the war and the consequent progressive disruption of the training program made the last planned study impossible, although all the arrangements had been made for it, including the preparation of the check ride score card in final form.

In addition to the five screening studies, identified as studies I, II, III, IV, and V, two other studies, VI and VII, were performed for other purposes and will be reported in later sections of this chapter.

Procedure

Methods of measuring reliability and validity.—In all of the five screening studies reliability was measured by the test-retest method with the interval between test and retest constant for any given experimental group at one day, two days, or three days. In all cases the test and retest were flown in different planes and, except in Study I, the test and retest were administered by different check riders.

Validity was measured in each study in one or more of three ways:

(1) the ability of the item to discriminate between two groups of students differing in number of hours of training;

(2) the prediction of graduation or elimination, and

(3) the ability of the measure to discriminate between two groups chosen to be different in proficiency as judged by their instructors. These different methods of measuring validity have been discussed in chapter 4.

In all of the five screening studies except study II, validity was measured by the ability of the measures to predict graduation and elimination in Primary school. In study II the number of eliminees was too small. In studies I, II, and III validity was further measured by finding the ability of each measure to discriminate between two groups of students, one with a greater number of hours of training and the other with a smaller number of hours of training. In studies IV and V a second measure of validity was the ability of the measure to discriminate between students judged by their instructors to be in upper or lower groups with respect to flying proficiency.

Summary of conditions.—Table 6.3 shows for each of the five screening studies a brief description of the groups tested. Column 1 shows the number of the study, the location of the Primary school where the study was performed, and the type of Primary training plane used. Column 2 indicates the size and nature of the groups used in obtaining graduation-elimination biserial correlations. Columns 3 and 4 show for studies I, II, and III the nature of the groups used to find the relationship between number of hours of flying training and score on the objective measures. For studies IV and V, columns 3 and 4 describe the groups used in calculating validity coefficients based on instructors' ratings or rankings. Column 5 shows the nature of the groups used in obtaining the test-retest reliability measures.

Sources of Variability

Constant errors in air speed indicators.—After the first few check rides in study I, it was found that the air speed indicators in the different planes had various constant errors. Following this discovery, each check rider was asked to report for each check ride the indicated air speed on his plane while the plane was in level flight at cruising r. p. m. (2050 r. p. m. at this school). This yielded a fixed figure which was the best available approximation for use in correcting absolute air speed measures during take-off and during spin recovery. It was found that the indicated air speed at cruising r. p. m. in level flight varied from 70 m. p. h. on one plane to 100 m. p. h. on another plane with most of the values clustering about 85 and 90 m. p. h., the correct values. Since Primary schools do not have instrument-calibrating shops, no other method of correction was possible. Since calibration measures on the air speed indicators were not available, it was necessary to assume that the error in the airspeed indicators was a constant one and that probably the range measurements were little affected, i. e., that 95 m. p. h. minus 75 m. p. h. was equal to 115 m. p. h. minus 95 m. p. h.

Change to Stearman plane.—Following study I, which was done with the Fairchild Primary Trainer (PT–19), it was learned that the Training Command planned to use only the Stearman Primary Trainer (PT–13 or PT–17) in Primary schools. The Fairchild training plane was a 175-horsepower low-wing monoplane, while the Stearman was a biplane with 225 horsepower and had possibly different flying characteristics. After experience had been gained in the Stearman, new measures on new maneuvers together with revisions of the maneuvers

TABLE 6.3.—*Summary of conditions involved in statistical analysis for each of the 5 item-analysis studies*

1	2	3	4	5
Study No., school, and plane	r_{bis}, graduation-elimination	r_{bis}, upper group versus lower group first day	r_{bis}, upper group versus lower group second day	Test-retest reliability (on different days in different planes)
I Coleman, Texas (Fairchild).	Average of first and second day for 15-hour group: 23 graduates, 12 flying-deficiency eliminees in primary and basic schools; (55-hour group too few eliminees).	15-hour group (N=38) versus 55-hour group (N=41).	15-hour group (N=38) versus 55-hour group (N=41).	z-method combination of reliabilities of 15-hour group and 55-hour group. Test and retest by same check rider.
II Tulare, Calif. (Stearman).	Number of flying-deficiency eliminees too small for calculation of meaningful coefficient.	15-hour group (N=43) versus 65-hour group (N=46).	15-hour group (N=43) versus 65-hour group (N=46).	z-method combination of reliabilities of 15-hour group and 65-hour group. Test and retest by different check riders.
III Tulare, Calif. (Stearman).	Average of first and second day for 10-hour group: 29 graduates, 13 flying-deficiency eliminees in primary and basic schools; (25-hour group too few flying deficiency eliminees).	10-hour group (N=45) versus 25-hour group (N=46).	10-hour group (N=45) versus 25-hour group (N=46).	z-method combination of reliabilities of 10-hour group and 25-hour group. Test and retest by different check riders.
IV Lancaster, Calif. (Stearman).	Average of first and second day for 11-hour students: 41 graduates, 10 flying deficiency eliminees in primary school.	*Grouped by instructor ratings.* Each instructor ranked his 4 students. The r_{bis} was done for average of first and second day for those ranked first versus those ranked last (N's=26 and 25).		51 11-hour students, test and retest administered by different check riders.
V Tuskegee, Ala. (Stearman).	Average of first and second day for 10-hour students: 30 graduates, 8 flying-deficiency eliminees in primary school.	Each instructor rated his students as in upper or lower half of all students he had taught. The r_{bis} was done for average of first and second day for those in upper half (N=15) versus those in lower half (N=30).		45 10-hour students, test and retest administered by different check riders.

and measures used in the first study were devised in collaboration with the Primary Training Advisory Board. Since none of the Primary schools in the Central Flying Training Command were using the Stearman Primary Trainer, it was necessary to go to the Western Flying Training Command to obtain students trained on the Stearman.

TABLE 6.4.—*Percents of validity and reliability coefficients at each level of significance*

Chances in 100 of obtaining positive coefficients as large as these, if there were no relationship between the variables.

	0.5 percent	2.5 percent	10 percent	More than 10 percent	Total number coefficients
r_{bis}, grad-elim [2]	10% [1](0.60–0.21)	5% (0.40–0.34)	23% (0.33–0.18)	62% (0.17 or less)	94
r_{pbis}, r_{bis} rankings [3]	11% (0.60–0.37)	26% (0.38–0.31)	30% (0.30–0.21)	33% (20 or less)	27
r_{pbis}, first day [4]	22% (0.62–0.28)	10% (0.27–0.20)	18% (0.19–0.14)	50% (0.13 or less)	138
r_{pbis}, second day [5]	11% (0.43–0.28)	11% (0.27–0.20)	14% (0.19–0.14)	64% (0.13 or less)	138
r_{tt}, test-retest [6]	25% (0.71–0.28)	11% (0.27–0.20)	13% (0.19–0.14)	51% (0.13 or less)	162
Total coefficients					559

[1] The figures in parentheses show the limits of the range of coefficients obtained at each significance level.
[2] The biserial correlations for the sum of scores on two daily administrations of an item versus graduation or flying deficiency elimination in Primary school.
[3] The biserial or point-biserial correlations for sum of scores on two daily administrations of an item versus upper and lower proficiency groups as rated by the students' instructors.
[4] The point-biserial correlations for each item score on the first day versus an upper group with more hours of flying training and a lower group with fewer hours of flying training.
[5] The point-biserial correlations for each item-score on the second day versus an upper group with more hours of flying training and a lower group with fewer hours of flying training.
[6] The product-moment reliability correlations for item-score on the first day of testing versus item-score on the second day of testing done one to three days later.

Propeller pitch variations from plane to plane.—Between studies II and III the PT–13 replaced the PT–17 planes in use at Tulare where the experimental work was being done. These two models of the Stearman had the same flying characteristics, but on the PT–13 two types of propellers were in use and these two types of propellers had different pitches. This meant that, at the same rpm, planes with different propellers had different actual air speeds. In study III it was necessary to use PT–13 planes with only one type of propeller—the MacCauley propeller. In spite of this control, the pitch of the propellers was found to vary somewhat when measured with a universal protractor. It is believed that this variation in pitch was not enough to affect materially the results of the experiment.

Necessity to shift work from a population of cadets to one of returnees.—After studies I, II, and III had been completed, it was announced that until further notice no aviation cadets would receive flying training. The only men then trained in Primary schools were bombardier and navigator officers returned from combat and former flying instructors from Primary schools. Most of the returnees had received various amounts of unofficial pilot experience in a variety of tactical equipment. It was difficult to evaluate the possible influence of this experience on Primary flying proficiency. With this kind of population, both the criterion of number of hours of training and the criterion of predicting graduation-elimination seemed less likely to provide useful results. Accordingly, it was planned to validate new objective measures against upper and lower proficiency groups of students at approximately the tenth hour of training with the proficiency determined by instructors' ratings. The procedure was followed in studies IV and V.

FIVE ITEM-ANALYSIS STUDIES: GENERAL RESULTS

In this section the results on measures tried out during the five screening studies at Primary schools are summarized in terms of the numbers of reliable and valid measures. In the next section the best individual measures on specific maneuvers are described and evaluated. The reliability and validity of each of the measures tested is presented in the appendix to this chapter.

Percents of Validity and Reliability Coefficients at Each Level of Significance

The results of the five screening studies are summarized in table 6.4, which shows the total number of each kind of validity and reliability coefficients calculated and the percents of these at each level of significance. As described in the earlier paragraphs on *Procedure*, five different groups of coefficients were calculated:

(1) The biserial correlations for the sum of scores on two daily administrations of an item versus graduation or flying deficiency elimination in Primary school;

(2) The biserial correlations or point-biserial correlations for the sum of scores on two daily administrations of an item versus upper and lower proficiency groups as rated by the students' instructors; biserial correlations were used when the proficiency categories were continuous, as in the instructors' rating of students in the upper half or lower half of all the students in the instructors' experience; point-biserial correlations were used when the proficiency categories were not continuous, i. e., when each instructor ranked his four students and the upper proficiency group was made up of students ranked first and the lower proficiency group was made up of students ranked fourth;

(3) The point-biserial correlation for each item-score on the first day's testing versus an upper group with more hours of flying training and a lower group with fewer hours of flying training; these groups were quite different in number of hours of training, e. g., 15 hours versus 55 hours, and 10 versus 25 hours;

(4) The point-biserial correlation for each item-score on the second day's testing versus an upper group with more hours of flying training

and a lower group with fewer hours of flying training, as described in the preceding paragraph (3);

(5) The product-moment test-retest reliability correlation for the item-score on the first day's testing versus the item-score on the second day's testing done 1 to 3 days later.

Out of a total of 559 coefficients summarized in table 6.4 one-half of 1 percent would be expected to reach the highest level of significance by chance. In these studies this level was reached by 10 percent of the biserial correlations indicating ability to predict graduation-elimination and by higher percentages up to 25 percent for the other measures of validity and reliability. While these figures are not large, enough of them are substantially above the chance level to indicate the discovery of some real relationships between objective measures and flying skill.

It will be noted that only 25 percent of the test-retest reliabilities are at the highest level of significance (36 percent at the top two levels). The problems of obtaining high test-retest reliabilities in the field of objective measures of flying skill will be discussed in a later section of this chapter.

Because the conditions in the different studies made it impossible to secure all of the types of coefficients on each measure, it is not meaningful in this table to compare the different rows reporting the different types of correlations.

Relationship Between the Different Indices of the Goodness of Each Measure

It will be recalled that the measures tried out in studies I and III were evaluated in three different ways by determining: (1) The ability of the measure to discriminate between students with different amounts of flying training, calculated separately for each of the 2 days of testing and yielded two point-biserial coefficients; (2) the ability of the sum of the two administrations of the measure to predict subsequent graduation-elimination; and (3) the test-retest reliability of the measure. In order to determine whether the measures which are best according to one index are also best according to another, the correlations among these indices were calculated. This was done for the 63 measures in studies I and III, in which all of the above-described types of correlations were available. The calculations were made separately for studies I and III and the results combined by Fisher's z-transformation technique. In order to correct for the fact that correlation coefficients were not distributed normally, the coefficient for each measure was transformed into a z-score before these correlations were calculated.

Similar results for same measure of validity on different days.—From table 6.5 it can be seen that the correlation between the point biserials on the first day showing relationship between item-scores and hours

ABLE 6.5.—*Intercorrelations of reliability coefficients and the validity coefficients for graduation-elimination and for hours of training* ($N=63$) [1]

	First day r_{pbis}	Second day r_{pbis}	Grad-Elim r_{bis}	r_{11}
1. First day, r_{pbis}, U. versus L.	....	0.41	0.00	−0.25
2. Second day, r_{pbis}, U. versus L.	0.41	....	−.07	−.01
3. r_{bis}, grad-elim	.00	−.07	....	−.03
4. r_{11}, reliability	−.25	−.01	−.03	....

[1] The figures shown are the z-transformation combination of separate calculations for studies I and III. Complete data are found in the appendix for chapter 6.

For a description of the different coefficients see table 6.4.

of training and the similar point biserials on the second day is 0.41. This means that the measures which give the best discrimination between the two groups with different amounts of training on the first day also tend to give the best discrimination on the second day. Since the point biserials on the first day are presumably measuring the same thing as the point biserials on the second day, this correlation gives an indication of the reliability with which differences in the discriminating powers of the different measures have been established. There is a definite relationship, but it is moderately low, probably because the point biserials which were correlated were originally computed on relatively small samples.

Different results for two types of validity.—It can be seen that under exactly the same conditions, there is no correlation between the biserials indicating the ability of the measures to predict graduation-elimination and the point biserials indicating their ability to discriminate between students differing in amounts of training. This means that the measures which were best at predicting which students would pass and which would fail were not necessarily best at discriminating between students with differing amounts of training. Under the conditions of these exploratory studies, the two criteria represented by these two types of indices seem to be independent and to test for different types of relevance. If this should be a general finding, repeated in other carefully controlled and more extensive studies, it would mean that the errors which characterize the student with low aptitude are different from those which characterize the inexperienced one. It would also indicate a necessity for developing two somewhat different batteries of measures: one to measure the results of training experiments and another to select students for elimination and to validate classification tests.

Low relationship between reliability and validity.—It will be seen that the correlations of the reliability coefficients with each of the other two indices is low, or even negative. This means that the measures which were most reliable were not necessarily the best at predicting graduation elimination or discriminating between the students with different amounts of training. This problem of the reliability of objective measures will be taken up in a later section of this chapter.

For the 26 measures tried out in study IV, the reliability coefficients correlate practically zero with the two coefficients of validity, *viz*, the biserial correlations with graduation elimination and the point-biserial correlations with the instructors' rating of students. However, the correlation between the graduation-elimination coefficients and the ranking-criterion coefficient is 0.46, which reflects the fact that the rankings and the recommendations for elimination were both done by the student's own instructor. The detailed data are presented in the appendix to this chapter.

FIVE ITEM-ANALYSIS STUDIES: DESCRIPTION AND EVALUATION OF SPECIFIC MEASURES

In devising measures of a particular maneuver the attempt was made to cover all of the relevant skills it was possible to measure objectively, particularly the ones most characteristic of the maneuver, and to choose items which measured the efficiency and safety of the student's performance. For many important aspects of the student's flying skill, it was not possible to devise objective measures. For instance, it is important for the student to look around constantly to be sure that no other plane is coming so close to him that there is danger of a collision. Instructors constantly warn students about this and the successful pilot keeps a constant lookout. As described in the section on the development of a sample measure, an attempt was made to get instructor's judgments of whether students looked around 50 percent of the time. This measure was discarded as too variable and not objective. Other possible measures were not tried out when it was discovered that they required the development of new measuring and recording instruments. A few such devices were made; for instance, a slip-skid recorder which was developed in collaboration with the Department of Psychology of the School of Aviation Medicine. It was decided, however, that the available research time could be spent most efficiently at first in exploring the objective measures that were possible using the flight and engine instruments available on the Primary trainer. Of the engine instruments, only the tachometer was useful since it was necessary in setting a certain specified r. p. m. as a constant condition for certain maneuvers. The only flight instruments available were the altimeter and the air speed indicator although in certain experiments ball-bank indicators were installed for the use of the check pilot.

In the following paragraphs the specific maneuvers and their measures and the experimental findings on them are briefly described. These measures are indicated which appeared worthy of further study and development and which would be desirable in a check ride designed to choose students with the greatest present and potential proficiency. A number of factors were taken into consideration in the selection of these best measures. The various estimates of validity and reliability were given first consideration. It was also necessary, however, to take into account the relevance of the measure to the general problems of flying skill and the importance of the measure in determining the pilot's safety and efficiency in performing the maneuver. Finally, when two items seemed to be measuring the same aspect of flying, only one was chosen for further trial. It was necessary to achieve a balance among these various factors in choosing the most likely measures. The score card for a check ride made up of these selected measures is presented in the last section of this chapter.

The complete data for the validity and reliability of each measure, including the means, standard deviations, and correlation coefficients are shown in the appendix for this chapter.

S-Turns

In the S-turns, the student's task was to make a ground track which formed equal semicircles on each side of a road that ran cross-wind. Approaching downwind the student crossed the road at right angles to it. After crossing the road the student made a semicircle, returning to the road at the end of this first loop, again at a right angle to the road. In the second loop, the student was on the upwind side of the road and again tried to make the ground track of the plane a semicircle, returning to the road at an angle of 90°. Loops 3 and 4 were exact repetitions of loops 1 and 2 so that at the end of the maneuver the student had made two complete S's. The test was difficult because in order to make a ground track of good semicircles, the student had to make proper drift correction by varying the angle of bank in relation to the wind. If the student made the most common error, he would undercorrect for drift and spend a long time on the loops on the downwind side of the road and a shorter time on the loops on the upwind side of the road, although with proper corrections, the downwind and upwind loops would take the same amount of time. Therefore, a good measure of drift correction was the sum of time on the downwind loops, 1 and 3, minus the sum of time on the upwind loops, 2 and 4. During all the loops, the student had to be careful to maintain a 500-foot altitude, and also to have the plane at 90° to the road and the wings level as he crossed the road.

In this maneuver the best measure was the one on *Drift correction*, the sum of time for the two downwind loops (1 plus 3) minus the sum of time for the two upwind (2 plus 4), as described in the preceding paragraph. The ability to make drift corrections in flying was fundamental to a number of maneuvers. This measure had a biserial correlation of 0.41 with subsequent graduation elimination ($N=42$). The maneuver was a difficult one and was regarded by pilot instructors as indicative of understanding of the principles of drift correction and requiring the ability to fly the plane automatically while paying attention to its course relative to the ground.

Altitude range during all four loops was also a good measure since it predicted graduation-elimination fairly well ($r_{bis}=.34$) and discriminated fairly well between students with 10 hours of training and those with 25.

The check-pilot's estimate of whether the *Plane's bank changed through level over the road* at the end of each loop of the S-turn was not discriminative.

Rectangular Course

In the rectangular course, the student's task was to keep the plane at a constant distance from the edge of a rectangular field, regardless of the direction of the wind. In going around the four sides of this rectangular field, the student's most difficult task was at the corners where he made his turns. For instance, if the student was going *downwind* along the side of the field and turned to go crosswind along the side of the field, he had to make the turn a little more than 90° so that his plane would be pointed slightly into the wind and would not drift away from the field. This crabbing into the wind had to be just the right amount. If he crabbed too much, he would gradually edge toward the field. If he did not crab enough or did not crab at all, the plane would drift away from the field. A similar problem existed when the student was going upwind along the side of a field and had to turn crosswind. His turn in this case was less than 90° if the ground track of the plane was to remain parallel with the field. As can easily be seen, the task becomes still more complex if the wind was quartering across the field and not blowing parallel to any of the sides. In addition to this, the student had to hold his altitude constant at 500 feet above the ground.

The *altitude range* on four legs of the rectangular course proved to be a discriminating measure. The *drift-correction* in the rectangular course was not a discriminating measure and apparently did not identify the good or poor student as well as the drift-correction measure in the S-turn.

Take-Off, Climb, and First Turn

The student's task in this maneuver was to line the plane up with the "T," open the throttle smoothly and continuously to full, make a straight take-off run, climb at a constant air speed of 70 m. p. h., level off at 300 feet above the ground, and make a coordinated, level 90° turn. During the take-off run and the climb, the student had to remain on a ground-track parallel with the "T." When there was a slight crosswind, he had to counteract it by holding proper rudder correction on the ground, or in the air by crabbing [34] slightly into the wind.

---

[34] When the pilot is flying on a ground track that is crosswind to any extent, the plane drifts away from the track unless it is headed into the wind enough to counteract the drift. This correction is called "crab," probably because the plane's movement in relation to the ground has a sideward component.

Of the eight measures tried out on this important maneuver, the following four were judged good enough to warrant further development:

Continuous or interrupted application of throttle during take-off was an important measure, since the indecisive student advanced the throttle in a series of short pushes, which lengthened the take-off roll and also made torque correction more difficult. The better student advanced the throttle slowly but continuously and without interruption, which shortened the take-off roll and made torque correction easier.[36]

The continuous or interrupted nature of throttle application was easy for the check rider to observe, since the throttle as well as the stick and rudder in the check-rider's cockpit were rigidly connected with the corresponding controls in the student's cockpit. During the take-off in the Primary trainer the check-rider's left hand was customarily on the throttle so that continuous or discontinuous application was easily observable.

Student lets plane fly off, pulls it off, or holds it on ground was found to be useful and also accurately judged by the pilot observer. It was first attempted to measure this part of the maneuver by having the check-pilot record the *Indicated air speed as the wheels left the ground.* This was neither a reliable nor valid measure even after the indicated air speed was corrected for variations in the air-speed indicator. This measure was poor, possibly because the correction was inadequate, but more probably because the check-pilot was understandably unwilling to turn his attention from the outside situation to the instrument panel while an inexperienced student was taking off.

Air speed range from 20-feet altitude to level-off for first turn was a good measure. The student who held his air speed fairly constant, even if it was a little too high or too low, was better than one who fluctuated, especially in a maneuver close to the ground.

Altitude range during first turn showed up fairly well statistically and was important because the student had to make a level turn while watching out for other planes and planning his exit from the traffic pattern.

The following measures did not show enough promise to warrant further development: *Ground control during take-off*, *Air speed reading as plane leaves ground*, *Plane drift during climb*, and *Slipping or skidding during first turn* as shown by a ball-bank indicator.

Accuracy Landing: Traffic Pattern and Power-Off Approach

In the usual Primary traffic pattern for a 90° power-off approach (illustrated in figure 3.1 in chapter 3), the student flew at an altitude

[36] In the Primary Trainer a number of factors, collectively labeled "torque," would, if uncontrolled, make the plane turn dangerously to the left. Torque correction consisted of enough pressure on the right rudder to hold the plane straight during the take-off roll.

of 500 feet above the ground and headed the plane toward the field in such a way that with a turn of 45° he was flying downwind outside the edge of the field (*downwind leg*). Flying downwind, when he reached a point approximately 500 feet beyond the downwind edge of the field, he made a 90° turn to the left and flew crosswind (***base leg***) part of the way across the end of the field. He then closed the throttle and made another 90° left turn and glided directly toward the field (*final approach*), gliding in and landing upwind.

In Primary schools the direction of the final approach and landing, and therefore of the other parts of the traffic pattern, were determined by the direction of the "T." The "T" was a ground marker set by the control tower so that it was on one of the eight points of the compass and pointed as nearly upwind as possible. Since the "T" had only eight settings, it was not always pointing directly upwind, the plane might have a tendency to drift sideways, which the student had to counteract on the downwind leg by crabbing and on the final approach by crabbing or slipping. If there was any wind at all, the student had to crab on the base leg in order to correct for drift and maintain the proper rectangular pattern.

The student's task in this maneuver was to enter the middle third of the downwind leg of the traffic rectangle at a 45° angle; maintain the correct 500 feet above the ground on the downwind leg and on the base leg until the throttle was cut; cut the throttle at the desired point which would allow landing in the correct area; glide with power off at the correct gliding speed and attitude; make a coordinated gliding turn onto the approach leg, starting the turn onto the approach leg at least 300 feet above the ground; make the final approach path parallel with the "T" and exactly lined up with the desired landing area, correcting for any drift produced by crosswind; and finally to break the glide and land. It was desirable to plan the pattern and point of closing the throttle so that the plane did not overshoot the field and so that no more throttle was needed to get into the field safely. If the plane was going to undershoot, however, the student had to see this himself and add power before it was necessary for the instructor to tell him.

The best measures of this particular maneuver were (1) ***Altitude range in the traffic pattern***, (2) ***Air-speed range in the final approach glide***, (3) ***Whether student changed final approach course or S'd to hit the correct landing spot***, and (4) ***Whether the student had correctly judged his total approach so that no added power was necessary on the final approach leg***. Altitude control was again one of the best measures. Air-speed range in the final approach glide on landing was one of the most important aspects of the landing. The student had to attend to a large number of duties and yet also had to keep his air speed high enough so that he did not stall and fall into the ground, and also had to keep his air speed low enough so that he could land successfully in the correct area. Students whose air speed fluctuated widely were not safe or accurate flyers.

Less valuable measures were (1) *Whether student entered the downwind leg in the middle third,* (2) *Whether the turn on to the final approach leg was started between 300 and 500 feet above the ground,* and (3) *Whether student maintained glide track parallel with "T" on final approach.*

Similar measures were tried out with power-on landings and with cross-wind landings. The results were the same as those described above.

Accuracy Landing: Place and Manner of Ground Contact

The student's task in an ordinary power-off landing was to put the plane down in the first third of the field. He had to cut the throttle at the proper point on the base leg so that, after turning and gliding on down without power, he would not land too close to the near end of the field and not beyond the middle of the field. If there was a slight crosswind and the student had to crab or put a wing down into the wind to compensate for this crosswind, he had to take off the crab just before the plane touched or the landing would be sideways to some extent. If the plane was moving sideways because of under- or over-correction for drift or for failure to remove crab at the last moment, the plane would have a strong tendency to groundloop. The student also tried to bring the plane down close to the ground so that when it is in a three-point, or stall, position it was no more than 3 feet above the ground. If the plane touched slightly wheels first, the angle of attack of the wings was changed and the plane appeared to bounce back into the air, after which the student had to make another landing. In the accuracy landing the student was trying to put the plane down in a specific area, in the case of these tests an area 200 feet wide and 600 feet long.

The two best measures of this maneuver were (1) *Landing attitude:* whether the student's landing was smooth three-point, tail first, wheels first, or three-point but dropped in 3 feet or more; and (2) *Whether plane bounced more or less than 3 feet on landing.* In contrast with these aspects of the landing, it was surprising to find that *place of landing* was not a discriminative measure, although the zone in which the plane lands was a matter of considerable interest to flying instructors and the object of considerable practice by students. However, it seems likely that the student's attempt to land in a particular zone was an important part of the maneuver, partly as a complicating condition which the student tried to fulfill, and partly because the place of landing was the principal goal of the maneuver. There was some evidence suggesting that all three of these measures of landing were somewhat more discriminating when the student's task was to land in a specified small area 200 feet wide and 600 feet long.

Two other measures were tried out in which the instructor judged whether the student *landed in a crab* or *landed drifting*. It was concluded that these two measures were too difficult to objectify. The reliability of these landing measures was particularly low, possibly because of the presence of gusts and irregular wind velocities near the ground. The reliabilities of landing measures are discussed in the section of this chapter concerned with the general reliability of objective measures.

Recovery Landing, After Bounce or Zoom

In these maneuvers landing attitude and bounce were measured when the student landed the plane after the instructor had either bounced the plane or zoomed the plane up after coming close to the ground. The student's task in each case was to take over the plane after the simulated bad landing and to do whatever was necessary to bring the plane down smoothly in a three-point landing. The validities and reliabilities of these measures indicate that they were not useful.

Simulated Forced Landing

In this maneuver the check rider closed the throttle when the plane was going downwind, 1,200 feet above the ground, and directly over a suitable field which had been pointed out to the student as the place where he should simulate a forced landing. Unlike the usual forced-landing practice situation, the student was completely familiar with where he was expected to land and with the aspects of the maneuver on which he was to be graded. The student's task was to glide down at the 70-m. p. h. gliding speed, establish a base leg on the downwind side of the field and to begin a final approach that would allow him to land, power off, upwind, in the first two-thirds of the field.

While the reliabilities of these single administrations were only occasionally above the chance level, graduation-elimination correlations based on the average of the test and retest scores, were significantly positive for *air speed range in glide*, *simulates base leg*, and *comes in upwind, downwind, crosswind*. The value of these measures might shrink upon trial with a large enough number of cases to be statistically stable. However, the obtained values were significant enough to warrant another trial. As with the other landing measures, *place of landing*, which was an item of obvious practical value in a forced landing, did not seem to have discriminative power as a measure of proficiency.

Four 180° Climbing Turns: Left, Right, Left, Right

In this maneuver the student's task was to climb the airplane at the proper air speed and r. p. m. and to correct for torque in such a manner that a climbing turn to the left would take the same amount of time as a climbing turn to the right. This was a real problem since without torque correction the airplane would make a climbing turn to the left much more rapidly than to the right.

The correction for torque was made with the rudder, in the case of climbing turns with right-rudder pressure. While holding right-rudder pressure the student also had to be careful to make the turn a coordinated one, that is, not to put on so much rudder that he would slip in the left turn or skid in the right turn.

Of the five measures tried out on this maneuver, none showed a useful combination of characteristics. The measure of proper *torque correction*, i. e., the difference in time for climbing turns to the right and left, showed some reliability but very little validity. This was characteristic of time measures. Students were apparently consistent in the amount of time taken to do particular maneuvers, but time was not relevant to their proficiency in any way. The *air speed range* in the four climbing turns gave variable results. Another measure, the student's tendency to *slip or skid* as shown by the maximum deflection of the ball in the ball-bank instrument was not discriminative.

Four 180° Gliding Turns: Left, Right, Left, Right

In a gliding turn the student's task was to set the plane in an attitude in which it would glide at the correct air speed and also to make the turns to the left and right holding the correct amount of left rudder [36] so that the rate of turn would be the same in both directions. If the student coordinated stick and rudder properly, the ball in the ball-bank indicator would remain centered throughout the turn. Of the four types of measures tried on this maneuver, one turned out well enough to warrant further investigation. This was the *air speed range* during the four gliding turns. In three successive experiments it showed ability to differentiate students with a greater, from those with a smaller, number of hours of training. The point-biserial correlations ranged from 0.17 to 0.45.

The other measures were not promising. The time measure, the *difference in duration for turns to the right and left*, again showed fair reliability but little validity. Coordination, as measured by *Deflection of the ball* in the ball-bank indicator, had little discriminating power.

360° Steep Turns at Cruising Throttle

In these steep turns at cruising throttle the student was allowed 90° to roll into the bank before entering the 360° turn and 90° to roll out after completing the 360° turn. Thus the student actually turned through 540°, although measurements were made on only the middle

[36] The left-rudder pressure was necessary in the Primary training plane because of built-in characteristics which made torque-corrections with either right or left rudder unnecessary at cruising speeds but made some left-rudder pressure necessary for correct flight at the lower air speed of a glide. This left-rudder pressure in gliding turns was sometimes known as "reverse-torque" correction.

360°. The student's task was to perform a 360° turn at cruising throttle as rapidly as possible and without changing altitude. The student's task was complicated by the fact that if he banked the plane up too steeply for the power setting, he was likely to lose altitude. On the other hand if he did not bank the plane as steeply as possible he would not get around the 360° as quickly. He also had to keep a sharp lookout for other planes.

The time measurements on this maneuver (*time in seconds to complete the 360° turns*) were among the most reliable of any measures that were obtained, ranging up to 0.71. However, the time measures did not correlate with the hours criterion or with graduation-elimination. *Looks around 50 percent of the time* was of doubtful objectivity and was put in at the insistence of flying instructors, as described in the introductory section of this chapter.

The best measure on this maneuver was *altitude range* which correlated very well with number of hours of training, although the correlation on the second test ride was not as good as on the first. This maneuver was probably not as good for testing purposes as the same maneuver at full throttle described next.

360° Steep Turns at Full Throttle

The student's task was the same in this maneuver as it was in the previous maneuver, although he had the greater amount of power allowed by full throttle. Again the *time* measurements had the highest reliability and the lowest correlation with hours and with graduation-elimination. *Looks around 50 percent of the time* was again of doubtful objectivity, had some reliability, some correlation with hours, and very little correlation with graduation-elimination.

Altitude range was again the measure with the best correlations with the hours criterion and some correlation with graduation-elimination. As described in an earlier section of this chapter, further administration of this maneuver will probably use a measure of degree of bank instead of requiring the check rider to measure the time that was taken for the 360° turn.

The measure of coordination in terms of deflection of the ball in the ball-bank indicator was again substantially valueless.

The instruments-covered versus instruments-visible measures in study II show comparable results, although it was unfortunately not possible to compare them on correlations with graduation-elimination.[37]

Maintaining Altitude While Reducing Air Speed

The student's task was to maintain altitude while reducing air speed from 90 m. p. h. to 70 m. p. h. and to hold air speed at 70 m. p. h.

[37] It was interesting to note that the students held a constant altitude almost as well with instruments covered as with instruments visible. Apparently flying was done by attention to the attitude of the plane and not by concentration on the altimeter.

until told to stop. Observations of altitude were made for a period of 1 minute after the beginning of the maneuver at a cruising air speed of 90 m. p. h. Two trials were given each day.

The altitude measures on the first day's trials correlated 0.34 with hours of training, but on the second day's trials the correlation dropped to 0.01. This maneuver was unfamiliar to the primary school students, at least at the 15-hour level, the level of the lower class. The upper class was evidently able to do it fairly well the first time, since they did not improve from the first trial of the first day through the second trial of the second day. The lower class, however, improved steadily from the first trial of the first day through the second trial of the second day and on the last trial of the second day did just as well as the upper class. Since these measures were being tried out for possible later continued use with classes at Primary schools, it did not seem desirable to develop this measure further. If one or two days of practice could remove the discriminating power of the measure, then continued coaching by flying instructors would probably do at least as much.

Forward Slip

In the forward slip the instructor closed the throttle and established a 70-m. p. h. glide along a long straight road or section line parallel to the wind. The student's task was to point the nose of the plane a little off to one side of the road and lower the opposite wing enough so that the plane would slip and remain over the road during a descent of 500 feet. He was instructed to slip enough so that the ball in the ball-bank indicator was deflected at least one-half ball-width throughout the maneuver. The student also had to raise or lower the nose of the plane sufficiently so that the indicated air speed would remain between 65 and 75 m. p. h.

Of the four measures on this maneuver, the best one was *highest air speed* during the slip, which had a biserial correlation with subsequent Primary graduation-elimination of 0.39, based on 57 students. Apparently the most important error made during the forward slip was to allow the air speed to become too high. *Slipped at least one-half ball-width* throughout the maneuver was also fairly diagnostic. Other measures that were less useful were *air-speed range during slip* and *plane over road throughout.*

Slipping Turn of 90°

In this maneuver the instructor closed the throttle and established a 70-m. p. h. glide. The student's task was to put the plane into a slip (as described under *forward slip*), make a turn of exactly 90°, maintain a 65–70 m. p. h. indicated air speed, and keep the plane in a slip throughout the turn without stalling it.

Of the five measures on this maneuver the two most promising were *remained in slip* throughout the turn and *recovered from turn at 90° ±5°.* The other, poorer measures, were *air-speed range, highest air speed* during turn, and *plane stalled or not.*

The measures in the *forward slip* and the *slipping turn of 90°* did not show enough promise to warrant the inclusion of either whole maneuver in the proposed check ride presented at the end of this chapter.

Estimation of Altitude, Air Speed, and R. P. M. with Instruments Covered

In many primary flying situations the most difficult aspect for the student was the perception of minimal cues which signaled the necessity for immediate adjustment on the controls. For instance, when the plane approached a stall in straight-and-level flight, a simple forward movement of the stick would eliminate this danger and the student's only problem was to recognize the visual, auditory, and tactual signs of the approaching stall. To the inexperienced student this was difficult, while to the experienced pilot it was easy. In many maneuvers the situation was similar: if the cues could be recognized, the control adjustments presented little problem. Apparently for these maneuvers the student's proficiency could best be measured in terms of his perceptual ability.

In study II it was possible to administer measures to students with all of the instruments in the student's cockpit covered. With instruments covered it would also be possible to check on whether or not students were performing maneuvers by concentrating too much attention on instruments during the check ride at the expense of keeping a sharp lookout outside the plane, i. e., whether objective measures were forcing students into undesirable mechanical flying. Three different maneuvers were used: straight-ahead climb, straight-ahead glide, and straight-and-level flight at an estimated altitude.

In the straight-ahead climb the student's task was to put the plane into the regular climb with the correct air speed and r. p. m. With instruments covered the student's cues for making the correct settings were the angle between the plane and the horizon, the sound of the engine and propeller, the "stiffness" or "mushiness" of the controls, and his feeling of sustentation. After the student had set the plane in what he believed to be the correct climb, he signaled the check pilot who recorded the air speed and r. p. m. The student was scored on his deviation from the correct values as read from uncovered instruments in the check pilot's cockpit.

In the straight-ahead glide the student's task was to set the plane in a straight normal glide at the correct air speed. The cues for this were the angle of plane and horizon, the sound of air hissing over the struts and fabric, and the feel of the controls. In this maneuver also, at the student's signal, the check rider recorded the air speed. The scoring was in terms of deviation from the correct value.

In the straight-and-level flight maneuver the student's task was to glide down from an altitude of at least 1,000 feet above the ground and level off at 500 feet above the ground at cruising air speed and r. p. m. He then had to fly at 500 feet above the ground, straight and level, for 30 seconds. This involved the student's judgment of absolute altitude without the aid of an altimeter and also his judgment of the air speed and r. p. m. at cruising. Since this test was carried out in level country, the altitude of the ground was the same as the altitude of the landing field and the check-pilot could tell within a small error how far the student was above the terrain. The specific measures were in terms of the deviation of the estimation from the correct altitude, air speed, and r. p. m.

Although the rationale for the above described measures seemed quite good, not one of them reached the lowest level of significance in differentiating between students with 15 and students with 55 hours of training. Reliabilities were low in most cases. Because of a small number of eliminations for flying deficiency, it was not feasible to calculate the biserial correlations with graduation-elimination. None of these measures was recommended for further trial.

Rudder-Exercise Stall

In the rudder-exercise stall the check rider closed the throttle, stalled the plane in a landing attitude, and held the stick all the way back throughout the maneuver. The student's task was to maintain the plane in as level an attitude as possible and to maintain the same heading by use of the rudders only while the plane lost 500 feet of altitude.

Of the three measures on this maneuver, two proved able to discriminate between students with more and fewer hours of training. One of these was the *maximum angle of the wings* with the horizon during the 500-foot descent. In the best form of this measure the check rider indicated whether the student allowed the wings of the plane to bank to an angle: less than 40°, between 40° and 60°, or more than 60°. The 40° and 60° angles were conveniently marked on the plane by the angle of the flying wires and the angle of the cabane strut respectively. The other good measure was the *maximum directional change* that the student allowed the plane to make. This was judged by the check rider to be: less than 22°, between 22° and 45°, or more than 45°. The third measure, *time to lose 500 feet*, which in these screening experiments was measured with a stop watch, did not discriminate. In this measure the longer time was the better because if the student allowed the plane to bank up steeply on one wing, it slipped off 500 feet very rapidly.

Power-On Stall

In the power-on stall recovery the student left the throttle at cruising power and pulled the nose of the plane up slowly to an angle greater than a normal climb. As the speed decreased, the plane was held in the same attitude by increasing back pressure on the stick and the wings were held level with the rudder, until the plane stalled. In Primary training, the plane was considered stalled when the ailerons were no longer effective. When the stall occurred, the student lowered the nose rapidly to an angle steeper than the normal gliding angle. With the forward stick motion he simultaneously advanced the throttle to full power, and as soon as flying speed was reached, the plane was returned to level flight. The student's principal tasks were to stall the plane thoroughly enough so that the ailerons were stalled, to bring the throttle forward with the stick in recovery, to hold the wings level with the rudder, and to return to cruising speed in level flight without stalling the plane again or diving excessively in recovery.

Of the four measures tried out on this maneuver, two seemed worth further development. The first was whether the student actually *stalled the plane* as indicated by the stalling out of the ailerons. The biserial correlation with graduation-elimination for this measure was 0.38. The other good measure was the *maximum air speed in recovery*, in which the lower the air speed the better the score because of the poorer student's tendency to dive excessively. This predicted graduation-elimination with a biserial coefficient of 0.43. The poorer measures were (1) *Time in seconds (measured with a stop watch) from forward stick until plane returned to level flight*, and (2) *Whether the throttle was full open when the stick was pushed forward during the recovery*. Although both of these were important aspects of stall recovery, neither showed useful validity or reliability.

Spin Recovery

In this maneuver the student's task was to put the plane into a normal spin and then to begin the recovery procedure by application of opposite rudder after a specified number of turns. After an additional one-quarter to one-half turn, the stick was moved briskly forward beyond the neutral position. This control position was held until the plane stopped rotating and went into a dive. The student then neutralized the rudder and pulled the nose of the plane up into level flight.

Of the three measures of this maneuver, two were chosen for further test. The first was a test of orientation, *whether the student applied opposite rudder in the spin within 22° of the specified number of turns.* The other was a measure of the *maximum speed during the recovery dive*. In this measure the best student was the one who had the lowest air speed in the dive. This method of scoring was decided on because the poorer students had a tendency to dive excessively and to take excessively long times to recover into normal flight. The measure not recommended for further trial was the *time in seconds from application of opposite rudder until the nose cuts the horizon*. Although this measure had fairly good predictive power for graduation-elimination, in practice it was necessary to measure this time with a stop watch, a procedure which only the most experienced check-pilots found possible, so that the measure appeared not to be suited to administration by the regular check-pilots or instructors in Primary school.

Student's Perception of Procedures in Spin Recovery

On the hypothesis that the student's ability to perceive and identify correctly the procedures during spin recovery was an important aspect of his proficiency, students were tested on their observation of a spin recovery performed by the instructor. Each check pilot was instructed to spin in a certain direction for a specified number of turns and then to perform a recovery either correctly (with the rudder first and then the stick), or incorrectly (with the stick first and then the rudder). Both student and check pilot were given score sheets. The instructor recorded what he did and the student recorded what he perceived the instructor to do. Scoring was on the basis of agreement between student and instructor on (1) *the direction of the spin*, (2) *the number of turns before application of opposite rudder*, (3) *whether stick or rudder was used first during the recovery*. Only the degree of agreement on the *number of turns* proved at all predictive of graduation-elimination. Since this presumably measured orientation in much the same way as the student's application of opposite rudder when he performed the spin, this perception measure was not added to the list recommended for further study.

Efficient Flight Planning

According to flying instructors and check pilots a student's ability to plan and execute a sequence of maneuvers was an important part of his general flying proficiency. Some students did not plan ahead while they were performing a maneuver and consequently, when one maneuver was completed, they were rarely ready to begin the next. Other students were able to plan so that, for example, they were near an appropriate road at the correct altitude when about to begin S-turns over a road, or could finish a sequence of maneuvers near the landing field.

In testing this ability, a sequence of 10 maneuvers was prepared which included a variety of problems. The student was instructed to perform these maneuvers in the regular manner, to observe the safety precautions directed by local flying regulations, and to plan and execute his flight so as to complete the check ride as quickly as possible. The student knew that each maneuver was graded by the check rider so that no maneuver could be slighted. The check rider recorded the duration of the flight from take-off to landing.

Although the test-retest reliability of the time measure was low, the average of the two tests correlated 0.48 with later graduation-elimination. This coefficient was based on only 30 graduates and 8 eliminees, so that the figure can be regarded only as tentative, although a positive coefficient of this size would be expected only 2.5 times in 100 if there were no real relationship between the variables.

FACTORS INFLUENCING THE RELIABILITY OF MEASURES AT PRIMARY SCHOOLS

In the five screening studies it was found that most of the reliabilities were low, and, as shown in table 6.4, there was no correlation between the reliability and validity of the measures. Some light is thrown on the probable cause of the low reliabilities by an analysis of data available on measures of the landing maneuver used in the Primary section of a large-scale study of the effects of additional training on flying skill (reported in chapter 10). Three aspects of this landing maneuver were measured: the zone of landing, the attitude of the plane (3-point or wheels-first) on landing, and whether the student dropped in or bounced more than 3 feet. Furthermore, the student was given a zero score on all three aspects if his landing was so bad that the instructor had to assist him in order to prevent an accident. Experts are of the definite opinion that landing in the proper place in the proper attitude without dropping the plane in or bouncing it involves important aspects of flying skill, namely, the ability to judge space and plan a course through it, to control the attitude and air speed of the plane, and to feel when it is about to stall.

A study of observer reliability was carried out on 152 students in class 45–B who had had 15 weeks of training at Curtis Field, Brady, Tex. During the last week of this Primary training the students were given two successive landings, each of which was scored by an observer on the ground as well as by the check rider in the plane. The correlations between the scores of the ground and aerial observers were determined separately for the first and second landings and combined by the z-transformation, yielding a coefficient based on a total of 304 pairs. The results are presented in table 6.6.

Test-retest reliabilities on the landing measures were obtained on a total of 170 primary school students from class 45–B who had completed 15 weeks of Primary training. Seventy of these were from Garner Field, Uvalde, Tex., and 100 from the Wilson-Benfils Flying School, Chickasha, Okla. These students were given a pair of successive landings on each of two days. The landings on the second day were with a different check pilot and in a different plane. The reliability of test and retest in immediate succession was determined by correlating the two landings on the first day with each other and also correlating the two landings on the second day with each other. These two sets of correlations were then combined by the z-transformation to yield a coefficient based on a total of 340 pairs. Similarly, the test-retest reliability on different days was determined by correlating the first landings on each of the two days and also the second landings of each of the two days. Then these coefficients were combined by the z-transformation, yielding a correlation based on 340 pairs. These results are also shown in table 6.6.[38]

TABLE 6.6.—*Reliability of landing measures in primary training*

	N	Zone of landing	Landing attitude	Dropped or bounced
Observer reliability [1] [2]	304	0.83	0.79	0.88
Test-retest on the same day [3]	340	.12	.32	.32
Test-retest on different days [4]	340	.02	.04	.00

[1] Since scores from 2 different ground observers and 46 check pilots were involved in these correlations, they represent absolute as well as relative agreement. Furthermore, there was almost perfect agreement between the means and standard deviations of the scores of the ground and air observers.

[2] These relatively high observer reliabilities agree very well with those secured for most measures of the Advanced Twin-Engine Scale reported in chapter 7. The observer reliability of the measure of zone of landing in the twin-engine scale was 0.80 for 47 cases. If corrected for broad categories, these reliabilities would approach unity.

[3] A calculation of the correlation between the two landings for the rest of the Primary schools involved in the large-scale study on the effects of added training (a total of 1,400 cases) confirms these figures. The reliabilities when test and retest follow each other in the same check ride are 0.13 for zone of landing, 0.28 for landing attitude, and 0.26 for bounce or dropped in.

[4] These very low reliabilities when the test and retest are administered in different planes on different days agree with those secured for landing measures in the other studies described in this chapter.

Observer Reliabilities Are High

The observer reliabilities are relatively high, ranging from 0.79 to 0.88. Since this degree of absolute agreement was achieved by the way in which the measurements were defined rather than by specially training the observers in comparing their estimates with each other, it seems likely that the standards used are stable with a large degree of generality. Thus, these correlations between observers indicate the objectivity of the scoring as well as its high observer reliability. Since the scoring directions were available to the students as well as to the check pilots, the entire process of measurement may be considered to be relatively objective.

Reliabilities from Test and Retest on Same Flight Are Low But Positive

The test-retest reliabilities of two immediately successive trials on the same check ride are constantly lower, ranging from 0.12 to 0.32. The agreement between the same observer's scoring of two immediately successive landings is much poorer than that of the two observers scoring the same landing. Since the observer reliability is relatively high, the lower test-retest reliability must be produced by conditions which introduce variability into the actual performance of the test rather than by inaccuracy in scoring the performance. These conditions are probably minute-to-minute fluctuations in wind and turbulence and variability in the student's performance.

[38] The two studies and the analysis involved in this table were designed and supervised by Maj. N. E. Miller.

Reliabilities From Test and Retest on Different Days Vary About Zero

The correlations between test and retest on different days are still lower. They are practically zero. These rides are on different days and are by different check pilots in different planes. The high observer reliabilities indicate that the change of check pilots does not produce a great effect upon the scoring of performance, although it might influence the student's confidence slightly and hence his manner of performance. The instruments of different planes are likely to be more or less out of calibration and various planes have somewhat different flying characteristics. Other factors such as fluctuations in the wind, turbulence, density of the air, and other weather differences would also be expected to vary more from day to day than they would within the period of a single check ride. Variability of the wind would be expected to have an especially important effect on landing. Finally, it seems likely that factors within the individual student may produce some variability in his performance and that such factors may vary more from day to day than within a single check ride.

Reliabilities Within a Day and Between Days Compared for a High-Altitude Maneuver

Correlation between maneuvers on the same check ride.—In studies I and III a left steep turn and a right steep turn were scored separately on each of the two testing days. It was thus possible to correlate the *altitude-range* scores on the left steep turn and the right steep turn on the same day. This was done separately for the first day and the second day and for the upper and lower groups in studies I and III. These correlations were then combined by the z-transformation and yielded a correlation of 0.38, based on 332 pairs of turns.

Correlation between maneuvers on different days.—For the between-days correlation, the altitude range in the left steep turn on the first day was correlated with the altitude range on the right turn on the second day, and the right turn on the first day was correlated with the left turn on the second day. A combination by the z-transformation of these correlations yields a coefficient of 0.17, based on 332 pairs of turns.

Within a day and between days compared.—The correlation between turns on the same day was 0.38 and between turns on different days was 0.17. Under the conditions of this experiment a difference of this size would be expected by chance less than 1 time in 100. This difference was probably the result of the different planes used on the two days and changes in turbulence, as well as day-to-day variability in the students' performance.

It seems unlikely that the difference was the result of variability among the check riders. Since the Primary trainer has room for only two people, the student and the check rider, it was not possible on this high-altitude maneuver to obtain the correlations between the scoring of two different observers. However, in a study reported in chapter 7, two different observers read the same set of instruments during altitude-control maneuvers in the B–25 (Mitchell). The observer reliability on these altitude-control measures ranged from 0.57 to 1.00. Since this was very probably also true for observations of altimeters in Primary trainers, observer error was not likely to be a large factor in the above difference in reliabilities.

Assumptions Involved in Conventional Test-Retest Reliability May Not Apply to Flying Measures

In measuring flying ability, a low test-retest reliability might easily result from the inexperienced student's observation of his own errors on the first test and consequent correction of these errors on the retest. According to instructors' reports, in a given maneuver including two measurable tasks the student pilot was often unable to perform both tasks at the same time. On the first test he might fail in one of the measured tasks and *note his error*. On the retest he corrects this error but fails on the other measured task. If a number of students did this on a given measure, the test-retest correlation for individual measures would be considerably reduced and might approach zero. However, for these measures the sum of scores on test and retest might correlate well with graduation-elimination, since students who performed well on both test and retest had better scores than those who failed on one or both.

The fact that correlations between maneuvers performed in immediate succession on the same check ride were higher than correlations between maneuvers on different days seems incompatible with the ideas of the preceding paragraph. This correction during the retest of errors made on the original test, while it has been reported by instructors, might not occur frequently enough to affect the correlation.

Implications for Further Work

Both landings and steep turns were very important parts of the task of all pilots. According to experts, the variables measured in these scales were very relevant aspects of these maneuvers. This study showed that the difficulty of measuring these aspects was more the result of factors introducing variability in performance from day to day than of errors in measuring any given sample of performance. Landings were probably especially susceptible to variable conditions, such as gustiness of wind, which changed the performance from day to day. Steep turns do not seem likely to be more susceptible than the average maneuver and it is believed that similar results would be secured with many other types of measures.

The fact that the correlation between measures administered on different days, in different planes, and by different check riders was considerably lower than that between tests given in immediate succession in the same plane by the same check rider has important practical implications. Two different procedures should be used for two different situations:

(1) In measuring the results of training experiments it is important to control the effects of conditions of testing by giving check rides to matched pairs of experimental and control subjects as nearly in succession as possible in the same plane by the same check pilot; and

(2) If a number of rides are being given to increase the reliability of the measures on the same student, these should be given on different days in different planes by different check pilots in order to cancel out as much as possible the effects of the variable conditions.

It should be noted that reliabilities determined with test and retest in immediate succession, by the same check rider or in the same plane, probably will considerably overestimate the results that can be expected in any practical application of the measures tested.

One of the most practicable ways of coping with the factors producing day-to-day variability probably is to repeat the measures on a number of different days. These studies have shown that, in general, single administrations do not produce sufficiently reliable results.

RELIABILITY OF OBJECTIVE MEASURES WITH REPEATED ADMINISTRATIONS

Purpose

Since most of the reliabilities and many of the validities of objective measures in the screening studies were extremely low, it was decided to give repeated daily administrations of a check ride in an effort to obtain higher reliabilities and validities. Because it was desired not to interfere with training, the tests were administered by each student's own instructor and the check ride was necessarily made up of measures on maneuvers that were performed on most of the flights during the early part of training. The only maneuvers certain to be performed on every flight during the early part of training were take-off, traffic pattern, and landing. On these maneuvers it was possible to try out some 18 measures, most of them objective. Some apparently less objective measures were also included with the thought that methods of making them more objective might be suggested in the course of the experiment.

Procedure

Study VI[30] was conducted at Uvalde and Brady Primary Schools. In it the items were administered every day on which there was a dual ride from the sixth through the thirtieth hours of training. The tests were administered each day to 73 students by the students' own instructors. An average of 12 administrations per student was ob-

[30] Study VI was done at a time when flying students without overseas service or previous flying experience were still being trained in Primary schools. The studies are grouped logically rather than chronologically. Thus, the item-analysis studies are numbered I through V and the special studies on reliability and intercorrelation are numbered VI and VII.

tained. Since the same instructor gave all of the 12 administrations of the test, the odd-even reliability from these data might be spuriously high. In order to obtain an estimate of this possible spurious rise in reliability, arrangements were made in planning the experiment to have each student tested at approximately the sixteenth hour by another instructor. This administration by the other instructor was called the "switch-ride."

In order to obtain an estimate of the reliability of the measures when they were administered by different instructors, the switch-ride results were correlated with the scores given by the regular instructor on the previous day. They were also correlated with those given by the regular instructor on the following day. Thus there were two correlations between the scores given by the "switch" and the regular instructors. These two correlations combined by the z-transformation technique gave the best possible estimate of the test-retest reliability of a single administration of these measures when test and retest were given by different instructors. These figures are shown in column 1 of table 6.7.

TABLE 6.7.—*Odd-even reliabilities for 12 administrations by same check-rider compared with reliabilities when different check riders were used (N=73)*

	1[1] $r_{s \cdot r+r}$	2[2] $r'_{s \cdot r+r}$	3[3] r'_{oe}	4[4] $r'(1.3)(2.4)$	5[5] r_{bis} grad-elim
VI-A1. Continuous application of throttle in take-off.	0.45	0.91	0.80	0.71	0.56
VI-A2. Deviation of plane from straight path during take-off roll	−.03	(²)	.90	.86	.48
VI-A3. A/S reading when plane leaves ground	.34	.86	.86	.86	.06
VI-A4. Plane pulled off ground, held on ground, or flown off correctly	.19	.74	.84	.76	.73
VI-A5. A/S range in climb to first turn	.19	.74	.84	.75	.66
VI-A6. Plane drift during climb	−.12	(²)	.87	.81	.48
VI-A7. Altitude deviation from 300 feet during first turn	.24	.70	.92	.86	.77
VI-B1. Traffic pattern: Enters downwind leg in middle third	−.10	(²)	.83	.71	.43
VI-B2. Traffic pattern: Maximum deviation from altitude of 500 feet in pattern from entry till throttle is cut	.04	.33	.92	.78	.90
VI-B3. Final approach: A/S range from beginning of last turn until student begins to break glide	.17	.71	.92	.91	.86
VI-B4. Traffic pattern: Turn onto approach leg started 300 feet or above	−.07	(²)	.64	.46	.67
VI-B5. Final approach: After throttle closed more power not necessary, or added as required	.20	.76	.66	.81	.63
VI-B6. Final approach: Maintains glide track parallel with T	.08	.81	.80	.63	.30
VI-B7. Ground contact: Landed in crab	.13	.64	.73	.66	.20
VI-B8. Ground contact: Landed drifting	.09	.54	.82	.76	.44
VI-B9. Landing attitude: Smooth 3-point, tail first, wheels first, 3-point but dropped in 3 feet or more	.07	.47	.78	.60	.33
VI-B10. Plane bounced	−.01	(²)	.85	.72	.00
VI-B11. Place of contact: Goal, first ⅓ of field	−.12	(²)	.67	.64	−.23

[1] $r_{s \cdot r+r}$.—This correlation is a z-transformation combination of: (1) The correlation between results from the "switch-ride" administered by a different instructor at approximately the 16th hour of training and the results of the *preceding* day's test ride by the student's regular instructor; and (2) The correlation between the same "switch-ride" and the *following* day's ride by the student's regular instructor.

[2] $r'_{s \cdot r+r}$.—The figure in column 1 was then corrected by the Spearman-Brown formula to estimate the reliability of 12 administrations. Negative reliabilities were considered meaningless and were not corrected.

[3] r'_{oe}.—This correlation is between the average of 6 odd and the average of 6 even administrations of each measure, corrected by the Spearman-Brown formula to estimate the reliability of all 12 administrations.

[4] $r'(1.3)(2.4)$.—This correlation is the odd-even correlation with hours held constant for the average of the odd scores and the average of the even scores by the semipartial correlation formula described in the text. The correlations necessary for entering this formula are shown in the appendix.

[5] r_{bis} grad-elim.—The correlation is between grad-elim in Primary school and the sum of four test administrations for each measure between the eighth and eleventh hours of Primary training. There were 53 graduated and 20 eliminated for flying deficiency.

Results: Agreement of Measures by Same Instructor Compared With That of Measures by Different Instructors

When the reliabilities involving the switch ride were corrected by the Spearman-Brown formula to estimate the reliability of 12 rides, they could be compared with the corrected odd-even reliability of the 12 rides given by the regular instructor. The two sets of values are shown in columns 2 and 3 of table 6.7.

As would be expected when the test and retest were given by different instructors (column 2), this more rigorous procedure produced a general drop in the reliabilities. In comparing the correlations between the different instructor and the regular one shown in column 2 of table 6.7 with the odd-even correlations for the 12 administrations given by the regular instructor in column 3, it was found that in all but two instances the values in column 2 were smaller than or equal to those in column 3. It may be inferred that the measures which showed the biggest drop in reliability are the ones on which the two instructors had the most different standards and that these were the least objective items. With only 73 cases in this study and the fact that the switch-ride correlations were based on only two trials each, some of these differences may be produced by chance. Of the six measures with the negative correlations, VI–A2, VI–A6, VI–B1, VI–B4, VI–B10, and VI–B11, the first two were on observations which were particularly hard to make objectively. In VI–A2 the check pilot had to judge whether during the take-off roll the plane deviated from the straight path:

> Not at all.
> One wing length (approximately 15 feet).
> One plane width (approximately 30 feet).
> Over one plane width.

In VI–A6 the check pilot had to make the same judgment during the straight-ahead climb to 300 feet after take-off. Since at Primary schools most of the take-offs were from grass, without a runway for a reference either on the ground or in the take-off climb, the judgment was difficult and allowed minimal possibilities for common standards among pilots. The other four measures with negative correlations seemed more likely to be objective. It is possible, of course, that some or all of the six negative coefficients were the result of chance factors.

Results: Effect of Holding Hours Constant in the Odd-Even Correlation

The odd-even correlations for the 12 test administrations by the regular instructor are shown in column 3 of table 6.7. Because it was not possible in the test situation to control the hour levels on the successive administrations it was necessary to correct this odd-even reliability for the different average hours of training of the various students. The students varied in the number of hours of flying preceding any given test administration for two reasons:

(1) Because of variations in the weather, attendance, etc., some students had more flying periods than others and thus consistently had more hours at each administration of the test;

(2) Some students were ready to solo with fewer hours of training, accumulated hours of solo between the dual check rides and thus had more hours of flying experience at each later test administration.

The fact that different students had somewhat different numbers of hours of flying instruction increased the range of skill represented in the population and thus raised somewhat the estimates of reliability. It should be noted that measures which were so unreliable that they did not measure anything would not be affected by this increase in range; it would not make an unreliable measure seem reliable but only make a reliable one seem more reliable.

It seemed desirable to get some estimate of the effect upon the reliabilities of the fact that hours of training were not held completely constant. In order to obtain the odd-even correlation, with hours held constant for both odd and even average scores, it was necessary to partial out hours in both the odd scores and even scores. This was accomplished by a semi-partial correlation.*

While the odd-even correlation *without* holding hours constant gave an estimate of reliability somewhat higher than it would if hours could have been held constant experimentally, the semi-partial technique gave too low an estimate. This occurs because the relationship between hours and skill had two aspects:

(1) The student who had more hours had more chance to improve their skill through practice. This was the relationship that had to be removed before a fair estimate of reliability could be obtained.

(2) The students with more aptitude soloed sooner and thus, because of their greater aptitude, had more hours on the average on later test administration. To the extent that the greater number of hours was indicative of greater aptitude, the semi-partial correlation, in which hours were held constant, gave too low an estimate of odd-even reliability.

Although the semi-partial correlation overcorrected to some extent, the reduction when hours were held constant was not large, as can be seen from column 3 and 4 in table 6.7. This means that although it was impossible to hold the number of hours of flying training completely constant for all students in this study, it did not have any very great effect on the results.

---

* $r_{(1.3)(2.4)} = \frac{r_{12} - r_{14}r_{24} - r_{13}r_{23} + r_{13}r_{34}r_{24}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{24}^2}}$

In which: Variable 1=odd score, variable 2=even score, variable 3=odd hours, variable 4=even hours.

J. P. Guilford, *Psychometric Methods*, McGraw-Hill Book Co. Inc., 1936. pages 403 and 404, formula number (188).

Results: Prediction of Graduation Elimination

In the five screening studies it was possible to determine the predictive power of the measures tried out by correlating the sum of the test and retest administrations with subsequent graduation or elimination from Primary school. It seemed probable that the sum of these two administrations would be more reliable than the single administration on either the test or retest. This seemed likely not only because of the increased length but also because the sum of the two administrations on different days with different planes and check riders would tend to average out these sources of day-to-day variability.

In the present experiment in which students were tested every day on which there was a dual ride from the sixth through the thirtieth hour of training, the four test administrations during the eighth through the eleventh hours were chosen to be summed and correlated with subsequent Primary graduation-elimination. These four were chosen for the following reasons: (1) Before the eighth hour the students had so little flying experience that objective tests would not measure their skills adequately; (2) After the eleventh hour the number of eliminations increased rapidly with a consequent restriction of proficiency range in the population; and (3) Students recommended for elimination in Primary schools received an average of approximately four independent check rides before elimination, so that if an objective check ride were accepted for use in Primary schools, four administrations would be administratively feasible.

Column 5 in table 6.7 shows the biserial correlation with Primary school graduation for the sum of four administrations of each measure between the eighth and eleventh hour of flying training. Of those 18 measures, 12, or 67 percent, were higher than 0.41, compared with 10 percent of the graduation elimination biserial correlations for the 94 measures in the five screening studies that were reported in table 6.4. Since these biserial coefficients are based on 53 graduates and 20 eliminees, for the 14 highest of these 18 coefficients there was only 1 chance in 200 of obtaining positive coefficients as large as these if there had been no relationship between the variables.

These coefficients indicated the predictive power of these measures when the measures were administered four times by the student's own instructor. The coefficients indicated a useful degree of predictive power when administered in this way in the early part of Primary training. However, it should be kept in mind that the values obtained are high possibly because the student's instructor usually made the initial recommendation for elimination before the student received independent elimination check rides by three members of the supervisory staff.

There remained the problem of whether similar predictive power could be obtained if objective measures were administered four times with a different check rider on each test, a suitable procedure if the objective measures were to be used as a regular part of the flying tests given to students proposed for elimination during Primary training. Arrangements were made for such a trial of a check ride made up of objective measures, but the end of the war and the consequent changes in training made this impossible. Such a check ride and the objective measures for use in it are described in the last section of this chapter.

INTERCORRELATIONS AMONG PRIMARY MEASURES IN THE LARGE-SCALE INVESTIGATION OF EFFECTS OF ADDED FLYING TRAINING

Purpose and Procedure

It was desirable to obtain intercorrelations of the measures tried in the item-analysis studies but the small number of students tested in any given study and the low reliabilities of the measures made it inappropriate. The first opportunity to obtain intercorrelations appeared when 1,400 students were tested in the Primary section of a large-scale study of the effects of additional training on flying skill. Other details of this study are reported in chapter 10, but since the intercorrelations are relevant to the present discussion, they will be presented here. In the sequence of Primary studies this was counted as study VII.[11] The 1,400 students tested on 13 measures had received 15 weeks of Primary training. Each student was tested by an instructor (other than his own) who had been carefully briefed on the methods of scoring.

Results

The intercorrelations of the 13 measures are shown in table 6.8. None of the intercorrelations between measures on different maneuvers was over 0.12, with the exception of the correlation of 0.21 between the *Loop* and the *Immelmann*, which are relatively similar maneuvers. These low intercorrelations were probably for the most part the result of the low reliabilities of the single administration of each measure. Unfortunately it was not possible to correct most of the intercorrelations for attenuation because the only reliability values available (except on landings) were for test and retest on different days. These day-to-day values were lower than the test-retest correlations within a day, as seen in table 6.6, and would *overcorrect* for attenuation if applied to intercorrelations for measures given on the same day.

Measures within a single maneuver have higher intercorrelations.—Some of the intercorrelations among the measures within a given maneuver were considerably higher. These higher correlations within a maneuver might have been produced, however, by the fact that

[11] Study VII was done at a time when flying students without overseas service or previous flying experience were still being trained in Primary schools. As stated in footnote 8, the primary studies are grouped logically rather than chronologically in this chapter.

Table 6.8.—*Intercorrelation among the measures in the primary check ride administered during the large-scale study on effects of added training* [1]

N=1410 S. E. of Correlation of 0.00=0.026

Study, maneuver and measure code		J	K	L	N	O	P	Q	R	S	T	U	STA	M	SD
VII-A1. I—*Immelmann*		0.21	0.06	0.06	0.09	0.09	0.02	0.09	0.10	0.04	0.05	0.12	0.07	2.27	0.91
VII-B1. J—*Loop*	0.21	....	.07	.07	.02	.04	.07	.07	.07	.00	.08	.09	.02	2.08	.97
Two 360° turns (left and right)															
VII-C1. K—Altitude deviation	.06	.07	....	.00	.01	.01	.06	.05	.04	.04	.03	.03	.09	1.63	.68
VII-C2. L—Maintained 60° bank or more	.06	.07	.00	....	.05	.03	.03	.07	.07	.01	.03	.02	.05	2.70	.60
VII-C3. M—Final roll out [2]	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Simulated Forced Landing															
VII-D1. N—Simulates a base leg	.09	.02	.01	.05	....	.15	.03	.06	.06	.02	.02	.03	.05	2.71	.70
VII-D2. O—Area where plane would have landed	.09	.04	.01	.03	.15	....	.06	.09	.08	.07	.08	.08	.02	2.35	.94
First Accuracy Landing [3]															
VII-E1. P—Area in which plane first touches	.02	.07	.08	.03	.03	.06	....	.34	.29	.13	.08	.11	.03	1.66	.76
VII-E2. Q—Landing attitude	.09	.07	.05	.07	.06	.09	.34	....	.63	.10	.28	.17	.37	2.45	.74
VII-E3. R—Bounced or dropped in	.10	.07	.04	.07	.06	.08	.29	.63	....	.06	.19	.26	.07	2.38	.80
Second Accuracy Landing [3]															
VII-F1. S—Area in which plane first touches	.04	.00	.04	.01	.02	.07	.13	.10	.06	....	.28	.27	.01	1.82	.78
VII-F2. T—Landing attitude	.08	.08	.03	.03	.02	.08	.08	.28	.19	.28	....	.58	.04	2.54	.68
VII-F3. U—Bounced or dropped in	.12	.09	.03	.02	.03	.08	.11	.17	.26	.27	.58	....	.02	2.41	.78
Pilot Stanine	.07	.02	.09	.05	[illegible]	.02	.03	.37	.07	.01	.04	.02	....	6.90	1.53

[1] For these correlations the IBM tabulation was done by S/Sgt. Gershenson of the Pilot Project and the computations were done by the statistical unit of the Psychological Section, Office of the Surgeon, Hq. AAFTC.

[2] Omitted because it lacked a sufficient distribution.

[3] If a bad landing necessitated a go-around, the student was given the lowest score, 1, for P, Q, and R, or for S, T, and U, as the case may have been.

there were simple physical causal relationships between different aspects of the airplane's behavior. This seemed more likely than the occurrence of common elements in the human abilities involved. The clearest example of a purely physical causal relationship was in the correlation between *Attitude on landing* and *Bounced or dropped in*. If the Stearman plane's landing attitude was tail high or slightly wheels first, it hit the ground while it still had flying speed and then as the shock of hitting the ground threw the tail down, the plane rose into the air again, as a result of the sudden increase in angle of attack of the wings. This is clearly shown in the correlation of 0.63 between *Attitude* and *Bounced or dropped in* for the first landing and 0.58 for the second landing. This is further substantiated by the fact that the correlations for *Attitude* on the one landing and *Bounced or dropped in* on the other are only 0.17 and 0.19. The correlations between these measures on the same landing and on different landings are shown in table 6.9.

It will be noted in table 6.9 that the correlations between *Area plane touches* and *Attitude* and also the correlations between *Area plane touches* and *Bounced or dropped in* were higher within the same landing than for different landings. The relationship was not as strong in these two aspects of landing but still remained higher within the same landing. If a student was going to overshoot the area, he may have tried to put the plane down too soon and made a landing in a wheels-first attitude, which in turn produced a bounce. If the student saw that he was not quite going to reach the desired area, he may have tried to hold the plane off the ground too long, stalled it too high, and dropped in more than 3 feet.

Reliabilities of test and retest in immediate succession for the three measures in landing were shown to be 0.13 for *Area plane touches*, 0.28 for *Attitude*, and 0.26 for *Bounced or dropped in*. In a further study it was planned to give administrations of each measure on four or more successive days. Under these conditions, test-retest reliabilities and intercorrelations should become more meaningful.

Reliability and Validity of a Total Check-Ride Score

One method of obtaining greater reliability in a homogeneous test is by increasing the length. Although a check ride, made up of a number of different measures, is not homogeneous, the intercorrelations between the various measures are for the most part positive, so that the total score for a number of measures on the same check ride should be somewhat more reliable than the average reliabilities of the individual measures. Of the 24 measures in study I, the 16 measures were chosen which best differentiated between students with 55 hours of training and those with 15 hours of training. The scores on these 16 measures were equally weighted and the test-retest reliability of the total score was then calculated separately for the 55-hour group and the 15-hour group. The biserial correlation with graduation-elimination was also calculated for the 15-hour group. These results are shown in table 6.10.

It should be noted that the biserial correlation with graduation-elimination is based on the sum of the scores for the two administrations on successive days. This validity coefficient, then, should be compared with the corrected reliability coefficient in the second column rather than the simple test-retest reliability shown in the first column. It should be noted that the measures were not selected on either of the bases compared but on their ability to differentiate between students with more and fewer hours of training.

TABLE 6.9.—*Correlations between the same measures during a single landing and on immediately successive landing*

	Same landing	Different landings
Area plane touches (first versus second landings)		PS[1] 0.13
Landing attitude (first versus second landings)		QT .28
Bounced or dropped in (first and second landings)		RU .26
Area versus attitude	PQ[1] 0.34	PT .08
Area versus attitude	ST .28	SQ .10
Area versus bounce	PR .29	PU .11
Area versus bounce	SU .27	SR .06
Attitude versus bounce	QR .63	QU .17
Attitude versus bounce	TU .58	TR .19

[1] The letters in this table correspond to the identifying letters on the measures in table 6.8. For example, PS is the test-retest reliability for area plane touches; PQ is the correlation between area plane touches and landing attitude for the first landing, and ST is the correlation between the same variables for the second landing; PT is the correlation between area on the first landing and attitude on the second and SQ between area on the second and attitude on the first, etc.

TABLE 6.10.—*Reliability and validity of total score on 16 measures in study I*

	N	r_{tt} test-retest	c_{tt}[1] double length	r_{bis} grad-elim v. sum of 2 administrations
Upper class (55 hours)	41	0.50	0.67	(2)
Lower class (15 hours)	35	.39	.56	0.32

[1] Spearman-Brown estimate of the reliability of the sum of the scores on test and retest together.
[2] Not calculable because no students in this group were eliminated in primary school after the fiftieth hour of flying training.

RECOMMENDED FURTHER INVESTIGATION OF SELECTED MEASURES

Plan of Experiment

It was planned to investigate further the measures which seemed most promising in all of the seven studies described up to this point. Since most of these measures were used in a single test and retest and given to a relatively small number of students, their results were not highly reliable statistically. They were the first coarse screening of the measures. The next study was to be a finer and more definitive screening. In order to make each measure more reliable, it would be administered on four successive check rides at each of two levels of training. These check rides would be administered to a larger number of students in order to secure statistically more stable results.

The following types of data were to be secured on each measure: (1) ability to predict graduation-elimination, (2) ability to discriminate between groups of students with different amounts of flying training, (3) reliability of odd versus even administrations, and (4) intercorrelations with other measures. The first two types of data would indicate the power of each measure for the two different possible uses; the third would help to indicate whether any lack of power is due to unreliability. The intercorrelations would show whether or not certain measures overlap so much that some of them should be omitted to avoid needless duplication and would also throw some light on the nature of flying skills. Finally, a comparison of the first two types of data, secured more reliably, would indicate the extent to which it might be necessary to develop two somewhat different scales of flying skill, one to measure those aspects which are the result of aptitude, and another to measure those aspects which are the result of learning.

It was planned to gather these data by administering a check ride of selected measures to 100–150 students each day for 4 days from the eighth through the eleventh hour of training. Each test administration would be by a different check rider in a different plane and results from the 4 days were to be averaged to obtain the most reliable measure of the students' ability at this level of training. The same procedure with the same check ride was to be followed for 4 successive days at a higher level of training, between the thirtieth and fortieth hours of primary flying. It was planned to test the entire class at the one remaining Primary school, using from 20–30 specially trained check riders under the supervision of members of the Primary Training Advisory Board.

These students were to be tested at the 8–11 hour level and again at the 30–40 hour level. In this way it would be possible to compare the same group at two different levels of its training. Any students who would have been eliminated between the first and second testing periods would be kept in school for the purpose of this comparison. Proper authority was secured to postpone these eliminations which were not immediately essential to safety until after the second period of testing. The correlation with graduation elimination would be obtained on the same group used to determine the ability of the measures to discriminate hours of training so that a direct comparison could be made between these two different criteria.

This study became impossible when the end of the war brought a sharp reduction in primary training. The only students being trained in Primary schools were bombardier and navigator officers returned from combat who had had varying amounts of incidental pilot training in course of flying their combat missions. Such training as continued was also severely disrupted by the release or transfer of instructors, supervisory personnel, and many of the students. It was concluded that further studies should not be conducted until conditions became more settled and possibly not until cadet training was again instituted.

Bases for Selecting the Measures

With the information gained from the first coarse screening of the various measures described in the preceding sections of this chapter, it was now possible to select better measures for a more rigorous trial on a larger number of students. The principal bases for the selection of these measures were as follows:

(1) *Suitability for use at the 8–11 hour level.*—From an examination of the results of the seven Primary school studies, from the distribution of elimination at the various hour-levels and from the pattern of introduction of new maneuvers in Primary schools, it was decided that the 8–11 hour level was the best time to administer an over-all check ride. By the eighth hour the fundamental maneuvers had been introduced and practiced so that measures of these maneuvers could be used. In order to be sure that the measures would be tested on their ability to discriminate fundamental skills rather than on the more transient factor of the opportunity of the students to be exposed to the maneuver involved, all of the measures used were to deal with maneuvers with which the student was already familiar, though not yet skilled, after 8 hours of flying training.

(2) *Comprehensiveness.*—It was desirable to have a set of maneuvers and measures which tested all of the fundamental skills of flying. This was true not only because a comprehensive set of measures was more likely to be discriminative but also because a check ride which emphasized only a few of the important measures would leave the way open for coaching by instructors who were preparing students to pass such a check ride. The ideal check ride measures *all* of the important aspects and *only* the important aspects. An instructor who coached his students for such a check ride would then be utilizing his instructional time most efficiently.

(3) *Reliability.*—If a measure had validity enough to predict graduation-elimination or to discriminate between groups with different amounts of training, it must have had enough reliability to measure something. If it was highly reliable, it may or may not have had validity. Therefore, when suitable criteria were available, validity was a more important basis for choosing measures than reliability. Relatively little weight was given to reliability in choosing the measures for this study. When tests of validity are available, separate measures of reliability are chiefly useful in indicating whether low validities are the result of the irrelevance of the thing measured or whether they are the result of the unreliability of the measure. In the latter case, the low validities may be raised by improving the relia-

bility through steps such as controlling variable conditions or administering more trials.

(4) *Validity.*—The two kinds of validity, ability to predict graduation-elimination and ability to discriminate between students with different amounts of flying training, were both considered in selecting measures. As described in the preceding section, it was planned to investigate further the relationship between these apparently different tests of the measures. If there is substantially a zero correlation between ability to predict graduation-elimination and ability to discriminate hour levels, it will be desirable to construct separate check rides for the two purposes. For those reasons it was important to include in this study measures with both kinds of validity.

Measures Selected for Further Trial

According to plan the following maneuvers and measures were chosen and the form prepared for use in the next trial check ride.

1. Take-Off, Climb, and First Turn

A. Application of Throttle During Take-Off Roll:
- (2) ...... Smoothly opens throttle to full position.
- (1) ...... Not opened full, uneven, or rammed.

B. At Take-Off, Student:
- (2) ...... Lets plane fly off.
- (1) ...... Pulls plane off the ground.
- (1) ...... Holds plane on ground.

C. A/S Range From 20 Ft. to Level-Off:
Check highest and lowest A/S.

← 55	60	65	70	75	80	85	90 → m. p. h.
9	8	7	6	5	4	3	2

D. Altitude Range During First Turn:
Check highest and lowest.

Less-Feet — 300 ft. — More-Feet

← 100 80 60 40 20 20 40 60 80 100 →

0 1 2 3 4 5 6 7 8 9 10

2. Two 360° Steep Turns

[Roll directly from left turn into right turn]

A. Deviation From Starting Altitude:
Check highest and lowest.[1]

Lose-Feet[2] — Gain-Feet

←120	100	80	60	40	20	0	20	40	60	80	100	120 →
0	1	2	3	4	5	6	7	8	9	10	11	12

[1] Scoring: Highest score is 9 points for range in zero interval; 1 point subtracted for each 20 feet of range so that check in gain 20 and lose 60 gives score of 5. No score below zero. Similar scoring for all range items.

[2] Marking: 0=+10 to −10; 20=11 to 30; 40=31 to 50, etc.

703322—47——12

B. CONDITIONS TO BE FULFILLED: FAILURE IN EITHER OF THESE MEANS ZERO SCORE ON WHOLE MANEUVER.

1. REACHED A 60° BANK IN EACH TURN (CABANE STRUT PARALLEL WITH HORIZON).

Yes	No
1	0

2. FINAL ROLL-OUT WITHIN ±45° OF STARTING HEADING.

Yes	No
1	0

---

3. POWER-ON STALL RECOVERY

A. STALLED AILERONS.

Yes	No
2	1

B. MAXIMUM AIR SPEED IN RECOVERY:

← 65	70	75	80	85	90	95	100	105 →
9	8	7	6	5	4	3	2	1

---

4. RUDDER-EXERCISE STALL

A. MAXIMUM ANGLE OF WINGS DURING 500 FEET:

(3) ...... less than 40° (40°=angle of flying wires and horizon).
(2) ...... 40° to 60° (60°=angle of cabane strut and horizon).
(1) ...... more than 60°

B. MAXIMUM DIRECTIONAL CHANGE DURING 500 FEET:

(3) ...... less than 22°.
(2) ...... 22° to 45°.
(1) ...... more than 45°.

---

5. SPIN RECOVERY

SPIN TO LEFT

A. DID STUDENT APPLY RUDDER AT 1½ TURNS WITHIN—

(2) ...... less than 22° either way.
(1) ...... 22° or more either way.

B. MAXIMUM AIR SPEED IN DIVE:

← 100	105	110	115	120	125	130	135	140	145 →
9	8	7	6	5	4	3	2	1	0

---

6. FOUR 180° GLIDING TURNS

[With medium bank]

A. FIRST TURN (LEFT)

A/S, CHECK HIGHEST AND LOWEST

← 55	60	65	70	75	80	85	90	95	100 →
0	1	2	3	4	5	6	7	8	9

B. Second Turn (Right)

A/S, Check Highest and Lowest

← 55	60	65	70	75	80	85	90	95	100 →
0	1	2	3	4	5	6	7	8	9

C. Third Turn (Left)

A/S, Check Highest and Lowest

← 55	60	65	70	75	80	85	90	95	100 →
0	1	2	3	4	5	6	7	8	9

D. Fourth Turn (Right)

A/S, Check Highest and Lowest

← 55	60	65	70	75	80	85	90	95	100 →
0	1	2	3	4	5	6	7	8	9

---

7. Simulated Forced Landing

[In field approx. ¼ mile square]

A. From 1,200 Ft. Directly Above Specified Field, Going Downwind

A/S, check highest and lowest.

← 55	60	65	70	75	80	85	90	95	100 →
0	1	2	3	4	5	6	7	8	9

B. Student Simulates Base Leg

Yes	No
2	1

C. Student Comes In

Upwind	Crosswind	Downwind
3	2	1

D. Student Would Have Landed in Field

In first 2/3	Outside first 2/3
2	1

---

8. Two Complete S-Turns

[Over crosswind road]

A. Altitude Deviation in all Four Loops Together

Check greatest deviations above and below starting altitude:

Lose-Feet								Gain-Feet		
← 100	80	60	40	20	0	20	40	60	80	100 →
0	1	2	3	4	5	6	7	8	9	10

B. Duration of Each Downwind and Upwind Loop in Two Complete S-Turns.

Position of clock second hand each time plane crosses road:

	Downwind 1st loop	Upwind 2d loop
Beginning		
End		
Time in seconds		

	3d loop		4th loop		Score
End					
Time in seconds					
Total time in seconds		–		=	

C. Student Attempts Semicircular Ground-Pattern in Each Loop:

Yes	No
2	1

Time of Day .... Wind Velocity ....

---

9. Traffic Pattern and Final Approach

A. From Entry Into Traffic Until Throttle Is Cut, Altitude Range:

Check highest and lowest

← Less-Feet						More-Feet →				
100	80	60	40	20	500	20	40	60	80	100
0	1	2	3	4	5	6	7	8	9	10

B. Air speed in Final Glide

Check highest and lowest

← 55	60	65	70	75	80	85	90	95 →
9	8	7	6	5	4	3	2	1

C. Changed Final Approach Course or S'd To Hit Correct Line

Yes	No
1	2

D. After Throttle Is Cut:

(3) ...... No power added.
(2) ...... Student added power by himself.
(1) ...... Instructor told student to add power, or instructor added power.

---

10. Place and Manner of Ground Contact

A. Place of Landing:

Middle Zone	Outer Zone	Not in Zones
3	2	1

B. ATTITUDE ON LANDING:

Smooth 3-point	Tail First	Wheels First	3-Point and Dropped in 3 ft. or more
4	3	2	1

C. BOUNCE ON LANDING:

Not at all	Less than 3 ft.	3 ft. or more
3	2	1

D. INSTRUCTOR ASSISTS

No	Yes
2	1

SUMMARY

Primary school was a crucial phase of pilot training. In Primary most students received their first flying instruction, became familiar with the fundamental maneuvers of flight, and exhibited their aptitude or lack of it. Of the total number of eliminations for lack of flying proficiency in Primary, Basic, and Advanced schools, considerably more than half occurred at the Primary level. Definite objective standards were particularly desirable in Primary to select those students whose skill warranted further training and to eliminate at an early stage those students who would not be able to meet the high requirements of combat flying. For these reasons the Primary phase was chosen for intensive investigation.

The first task in developing objective measures of flying skill was to select the critical aspects of the various training maneuvers and to devise relevant and comprehensive objective measures for them. For 31 maneuvers, or parts of maneuvers, 256 measures were devised and, after consultation and preliminary trial, the 131 best and most practicable of these were tried out in Primary schools on groups from 45 to 90 students who had not had previous flying training. For each of the measures, information was obtained on (1) reliability of test and retest on different days, (2) validity in differentiating between students with more and fewer hours of training, and where possible, (3) validity in terms of the ability of measures to predict subsequent graduation or elimination from Primary school.

Positive correlations significant at the 2.5 percent level or above were obtained for approximately 35 percent of the reliabilities. This level of significance was also reached by 25 percent of the correlations indicating ability to discriminate between groups with different amounts of training and 15 percent of those indicating ability to predict subsequent graduation or elimination.

The measures which on the first day of testing were the best at discriminating between students with different amounts of training were also best on the second day of testing. There seemed to be no relationship, however, between the ability of measures to discriminate between students with different amounts of training and their ability to predict graduation-elimination. The seeming independence of these two types of validity, under the conditions of this experiment, suggested that it may be necessary to construct one type of scale for measuring the results of training experiments and another somewhat different one for discriminating between students with different amounts of aptitude.

A special analysis was made of the reliability of certain representative measures. One of these was *range of altitude deviation* in a high altitude maneuver, a steep turn. Three others dealt with an important low altitude maneuver, an accuracy landing. These were *zone of landing*, *landing attitude*, and *bounced or dropped in*. On all of these measures it was found that when test and retest were given on different days in different planes by different check riders, the reliability was much lower than when test and retest were given in immediate succession in the same plane and by the same check rider. From this result two conclusions were drawn: (1) in any check ride used to measure the results of training experiments, it is important to control the effects of conditions of testing by measuring matched pairs of experimental and control subjects as nearly in succession as possible in the same plane and by the same check rider; (2) if a reliable measure of the individual student is desired, check rides must be repeated on a sufficient number of days in different planes by different check riders in order to average out the effect of variations in conditions of testing and of possible fluctuations in the individual's ability to perform.

As a further step in this analysis, the three objective measures on the accuracy landing were scored by two independent observers, one in the plane and one on the ground. The correlations between the scores by the two independent observers were found to be relatively high, ranging from 0.79 to 0.88. From this it was concluded that the low test-retest reliabilities were produced not so much by errors in measurement as by variations in the performance measured. These variations probably resulted from changes in turbulence, temperature, and other characteristics of the air, differences in planes and check riders, and fluctuations in the student's ability to perform.

One way of counteracting this variability was to give repeated administrations of the tests so that variable errors were more likely to average out. In a special experiment the student's own instructor graded him on objective measures each day that a dual ride was taken from the sixth to the thirtieth hour of training. It was found that useful graduation-elimination validities were obtained from the sum of four administrations of objective measures between the eighth and eleventh hours of training. Repeated administration also raised the reliabilities to useful levels when these administrations were made by the student's own instructors. The problem remained of applying a more rigorous test and finding whether adequately valid and reliable measures could be obtained if four successive administrations were made by different check riders on different days in different planes.

A final experiment was prepared in which 27 selected objective measures were to be given to 150–200 Primary students in 4 consecutive daily administrations between the eighth and eleventh hours of training, with each administration by a different check rider. For each measure a determination was to be made of the odd-even reliability of the four trials and also of the ability of the total score on these trials to predict graduation or elimination from Primary school. Another four administrations between the thirtieth and fortieth hours of training were planned in order to determine the ability of the selected objective measures to discriminate between students with more and fewer hours of training. It was planned to compare the effectiveness of each measure in discriminating between students with different amounts of training with its ability to predict pass-fail in order to determine whether these two criteria are similar or different. Finally, intercorrelations between the measures were to be calculated for the purpose of eliminating those which duplicated each other, and also to throw light on the structure of flying skill. It was impossible to perform this final experiment because of the end of the war and the consequent changes in training activities.

The proposed final check ride, containing the most reliable and valid objective measures chosen from the seven Primary studies, is presented in this chapter. The Pilot Project recommended this experiment as the next step in developing objective measures of flying skill at the Primary level.

CHAPTER SEVEN

Measures of Two-Engine Flying Skill (Contact)

Capt. Stanford C. Ericksen

INTRODUCTION

Use of Two-Engine Planes in the Training Command

The use of two-engine aircraft in the Training Command began in 1941 when small, selected groups of students were given instruction in the Lockheed Hudson and the B-18 prior to graduation. By 1944 the majority of students were being graduated from two-engine schools where they received 10 weeks training in the TB-25 Mitchell bomber. As late as 1941 it would have appeared extremely radical to propose that a plane with the power, speed, weight, and tactical characteristics of the B-25 could be used successfully as the standard Advanced two-engine trainer. This transition was the result of a number of contributing factors: (1) The tremendous need for multi-engine pilots to fly the larger combat and transport aircraft, (2) better methods of pilot training, (3) improved pilot selection, and (4) the fact that the use of tactical type aircraft at the advanced level materially improves pilot proficiency and shortens the precombat training required for two-engine pilots.

Advantages of Multiengine Aircraft for Research

Multiengine aircraft offer at least two distinct advantages for doing research in the area of pilot evaluation: (1) The instructor sits beside the student instead of in a separate cockpit and can observe in greater detail the quality of the student's performance, (2) there is space for a second observer to fly and make independent observations of pilot performance.

Because of these advantages, a series of exploratory studies were made in this area. On the basis of the experience gained in these studies, the decision was made to concentrate further work at the Advanced level of training on developing and testing the objective measures of instrument flying skill which are described in the next two chapters. The four preliminary studies on the measurement of two-engine contact flying skill are reported in this chapter.

STUDY 1: RELIABILITY AND VALIDITY OF A SUBJECTIVE CHECK RIDE [43]

The Pilot Project, working cooperatively with the two-engine group at the Central Instructors School, constructed a more detailed and analytical rating scale than the one normally used for grading students. This check ride covered the more important maneuvers included in the two-engine Instructors School curriculum. This subjective check ride provided the initial opportunity for the Pilot Project to apply reliability and validity item analyses to the problem of measuring flying skill.

Procedure

After a preliminary try-out of items, a rating scale of 94 "rated work sample" items was constructed. Sixty-nine student-instructors in the two-engine group were checked at two different times by separate raters, with an intervening interval of from 2 to 4 days. The raters were the 20 most experienced check pilots in the two-engine group and had never flown with the students before. Each officer made an original rating of approximately four students and also made the recheck of about four students previously graded by one of the other rating officers. The ratings were made after approximately 15 hours of flying instruction at the Central Instructors School.

The Scale

There were 14 different maneuvers, each one composed of 2 to 9 items. These are listed in table 7.1. The normal landing maneuver is a good example of the measures composing the scale. The check pilots were told to encircle the appropriate word or number, with "1" low and "5" high. The importance of "spreading" the ratings was emphasized.

Normal Landing:

Turn onto approach	1	2	3	4	5
Glide	1	2	3	4	5
Gear check on approach	No				Yes
Roundout and landing	1	2	3	4	5
Landing roll	1	2	3	4	5
Smoothness on controls	1	2	3	4	5
Trimming	1	2	3	4	5
After-landing check correct	No				Yes
Had to go around because of overshooting	Yes				No

Results

The product-moment test-retest reliabilities for the individual items were found to range from 0.00 to 0.26. For items marked

[43] Lt. Col. Charles M. Wharton, Commanding Officer, Two-Engine Group, Central Instructors School, Randolph Field, encouraged this research and made available the facilities of the Two-Engine Group for Studies 2 and 3 described in this chapter. Capt. A. T. Gordon, Research Officer, Two-Engine Group, assisted in the construction and administration of the check rides used in Studies 2 and 3. S/Sgt. C. P. Gershenson and S/Sgt. Walter W. Ismael made the statistical analyses for both projects.

either "yes" or "no," the tetrachoric correlations indicate reliabilities as high as 0.58.

The scores on the separate measures were added together to obtain a total score on each of the 14 maneuvers. The reliability coefficients of these maneuver scores ranged from 0.00 to 0.40, with the exception of one maneuver, forced landing, which was −0.29. The reliability of the total score was 0.25.

Using instructors' final grades as the criterion, validity coefficients were obtained for the results of the two rides separately. These values ranged from −0.06 to 0.41. The validity for the total check-ride score (sum of all maneuver scores) for the two rides was 0.39 for the first ride and 0.29 for the second.

In table 7.1 are the reliability and validity coefficients and significance levels for each maneuver score and the total check-ride score.

TABLE 7.1.—*Reliability and validity coefficients for each maneuver in the rated work sample check ride*

[N=69]

Maneuver	Number of items	Test-retest reliability	Validity [1]	
			1st Ride	*2d Ride*
Starting	5	*0.21	0.11	0.11
Taxiing	6	*.20	.09	.07
Take-off	8	.11	.13	−.06
Climb and level-off	6	.14	**.28	*.23
Power-off stalls	8	**.29	**.25	*.20
Slow flying	6	*.23	*.18	.13
Single-engine drill	8	*.15	.05	***.41
Single-engine stall	5	.03	.10	**.24
Forced landing	8	−.29	.02	.09
Pattern	7	.02	***.32	*.20
Normal landing	9	***.40	**.26	***.40
Take-off (second)	8	**.26	.13	.05
Single-engine landing	8	*.21	***.32	*.15
Parking and stopping	2	.01	−.02	.09
Total check-ride	94	**.25	***.39	**.29

[1] Correlation with final grade.

The one, two, or three asterisks indicate the statistical significance of these coefficients. If there were no relationship, positive coefficients as large as these would be expected by chance:

(No asterisk) = more than 10 times in 100.
* = 10 times in 100 or less.
** = 2.5 times in 100 or less.
*** = 0.5 times in 100 or less.

Conclusion

This form of subjective grading system in which the detailed specific aspects of pilot performance are scored separately holds some promise for use within any given training station. The merit of this type of scale is limited to its use only under conditions where the work of the check pilots can be continually standardized and their subjective ratings compared at frequent intervals. This type of grading system is similar to the one developed and used extensively by the Royal Canadian Air Force.

The reliabilities and validities obtained fall in about the same range as those of the objective measures tested under similar experimental conditions at the Primary level, described in chapter 6. The reliability of any one administration of a check ride of this type is so low that the results of a considerable number of independent observations must be combined to discriminate any but the most extreme cases.

STUDY 2: COMPARISON OF OBJECTIVE AND SUBJECTIVE MEASURES TAKEN DURING THE SAME CHECK RIDE

A check-ride booklet was constructed to test most of the basic maneuvers given in the Central Instructors School two-engine curriculum. This new check ride contained several subjective measures covering those important aspects of maneuvers which were not readily adapted to objective measurement. As a partial control for the large day-to-day variability in conditions and performance, this experiment was designed to give two check rides to each student within the same half-day. The primary purpose of this study, then, was to determine the test-retest reliability of objective and subjective measures when each student was tested twice within the same half-day by two different check pilots.

Procedure

On 15 and 16 August 1944, 103 student-instructors in the two-engine group were given the check ride by two different staff instructors. The students flew in their regular training planes, 49 in AT-10's (a two-place plywood and plastic two-engine trainer made by Beechcraft) and 54 in the UC-78 (Cessna Bobcat, a standard advanced two-engine trainer and utility cargo plane with fabric wings and fuselage). Two auxiliary fields were used exclusively for this experiment thus avoiding the heavy traffic normally found at Randolph Field. Unfortunately the wind was fairly strong and gusty on both days of testing with turbulence from moderate to severe. All students and the check pilots were given a copy of the scale and were thoroughly briefed before making the flights.

The Scale

Of the 38 separate measures, 20 are objective and 18 subjective work samples. The slow flying maneuver is reproduced to illustrate the form of instructions and the measures. Items 1, 2, and 3 are strictly objective; 4 and 5 are subjective work-sample measures.

Slow Flying

"At 4,000 Feet, With a Heading of 90°, Begin Slow Flying at 45 M. P. H. (UC-78) or 50 M. P. H. (AT-10). Maintain This Altitude and Heading."

To the rater.—Make observations during a 3-minute period, beginning with moment student reduces throttles. At the end of this time, cut the right mixture control.

1. Altitude deviations: Lowest .... ft.
   Highest .... ft.
2. Directional control: Starting heading ....°
   Heading of maximum deviation ......°
3. Air-speed control: Lowest .... m. p. h.
   Highest .... m. p. h.
4. Use of throttles to check approaching stall:
   .... excellent .... poor
   .... good .... never approached a stall
   .... fair
5. Attitude of plane during maneuver:
   .... nose too high .... fluctuated between too high and not high enough
   .... three-point
   .... nose not high enough

A listing of all the maneuvers and the measures used is given in table 7.2.

TABLE 7.2—*Test-retest reliability of objective and subjective measures of two-engine flying skill*

Item Description: N=103	Correlations	
	Objective measures	Subjective measures
TAXIING:		
1. Use of throttles	....	0.16C
2. Use of brakes	....	.24C
NORMAL TAKE-OFF:		
3. Use of throttles on roll	....	.13C
4. Level-off between 10 and 25 ft	....	.14tet
5. During level-off (attitude)	....	.41C
6. Reduction of power	....	.28C
7. Climbing air speed	0.44tet	....
8. Flight path	....	.60tet
9. Cleared area	.25tet	....
SLOW FLYING:		
10. Altitude	.33tet	....
11. Direction	.26tet	....
12. A/S	.10tet	....
13. Use of throttles	....	.41tet
14. Attitude	....	.36C
SINGLE-ENGINE PROCEDURE FROM SLOW-FLYING:		
15. Directional control speed	....	.11C
16. A/S	.37tet	....
17. Heading	.45tet	....
18. Altitude	.32tet	....
19. Procedure check-list	....	.64tet
STEEP TURN:		
20. Attitude	.20tet	....
21. Bank constancy	.20C	....
22. Bank maximum deviation	.31tet	....
23. Position of ball	.36tet	....
PATTERN:		
24. A/S } at turn to downward leg	.04tet	....
25. Altitude } at turn to downward leg	.05tet	....
26. Altitude during base leg	.25tet	....
27. Horn (or light) check	....	−.15tet
28. Gear checks	....	.18C
LANDING (POWER OFF):		
29. Altitude at approach turn	.21tet	....
30. A/S at turn	.38tet	....
31. A/S on glide	−.07tet	....
32. Gear check	.35tet	....
33. Roundout	....	.30C
34. Landing attitude	....	.22C
35. Point of landing	.40C	....
36. Directional control on roll	....	.24C
OVER-ALL RATINGS:		
37. Self-confidence and aggressiveness	....	.35C
38. General pilot ability	....	.42C

tet=tetrachoric correlation; C=coefficient of contingency.

Results

When all the scores were distributed, it was found that the range was unexpectedly small and that the objective deviations were usually recorded in even-numbered units (nearest 5° heading, 5 m. p. h., etc.). Each measure of heading, air speed and altitude was, therefore, scored in a simple dichotomy with the "split" as near the 50 percent point as possible. The results from students using the AT-10 and UC-78 were combined since the distributions and means were approximately equal and the 50-50 split normalized the distributions of the heading, air speed and altitude measures. Tetrachoric correlations and coefficients of contingency [43] were used to determine an estimate of test-retest reliability. Table 7.2 presents these coefficients for all items.

Inspection of table 7.2 shows that the objective and the subjective items had approximately the same reliabilities. The comparison is complicated by the fact that the performance measured by each type of item may not be equal in difficulty, ease of measurement, and consistency from day to day. Not enough is known of the inherent variability of the specific skills measured by the objective and the subjective items to make a detailed analysis of these results. However, this check ride provided enough measures of both types to show that the objective measures are as reliable as the subjective measures when obtained under the experimental conditions of this study. Had the study involved students from a number of different schools, however, one would expect the reliability of the subjective measures to be further lowered by differences in the standards of the different schools; the reliability of the objective measures should not be reduced since they are made in terms of measurements which are relatively independent of the observer.

Validity

Since these check rides were given near the end of training, failing students had already been eliminated. So many of the passing students were given the same grade, C, that it was impossible to compute reliable correlations with final grades.

Summary Comments

This study demonstrated several principles to guide further research on objective measurement of flying skill.

1. The form of the items: The form of the items in objective check rides should allow the observer to maintain a running account of the successive increases in instrument deviations. For example, if the student loses 100 feet while making a steep turn, returns to his starting altitude and then follows with a second deviation of 150 feet, the construction of the item should be such as to permit the check pilot to record each deviation as it occurs rather than waiting until the end

[43] J. P. Guilford, *Psychometric Methods*, p. 357, McGraw-Hill Book Co. Inc. New York, N. Y. 1936.

of the maneuver to record what he remembers to be the largest single deviation. In the present booklet the check pilot had to "write-in" the maximum instrument deviations. As a result, most readings were made to the nearest even number (5°, 5 m. p. h., 50 ft., etc.). This results in some loss of discriminating power of the measures.

2. The population to be tested: As far as possible the objective measures should be tested with a normal sample of cadets in training. If the population has been specially selected on the basis of superior aptitude or special training, the total range of scores will be restricted making it more difficult to obtain adequate estimates of reliability and validity. Two conditions were present in this investigation which increased the homogeneity of the students: (*a*) They were a select group of above-average pilots since most student officers sent to Central Instructors School were selected from the upper half of their Advanced school class and the poorer members of this selected group had been eliminated during the first phases of training at Central Instructors School. (*b*) The students had extensive prior training on the maneuvers being tested since Central Instructors School training is, for the most part, an extension of Advanced school training. All of those pilots had a good deal of previous practice performing the maneuvers in this check ride.

3. Standard testing conditions: About the only way the interfering effects of weather can be controlled is to schedule the testing at periods when it is expected that turbulence, wind, visibility, air density, temperature, etc., will remain somewhat constant and at near average levels. Objective instrument readings cannot adjust for error factors produced by gusty wind, turbulence, low visibility, or such nonweather conditions as plane differences and traffic. Since the check pilot can make necessary compensations when making a subjective evaluation of the students' performance, it is apparent that these nonpilot factors, operating in a random manner, tend to reduce the reliability of the objective measures more than that of the subjective ratings.

Conclusion

The results of this experiment do not show any consistent or marked differences in the reliability of the objective and subjective measures. The subjective items referred to performance considered highly important in each maneuver and for which the check pilots at the Instructors School had been frequently "standardized" as to the correct manner of teaching and grading. Even though the subjective measures might give as good a prediction of final pilot flying ability, two basic limitations of that approach favor the more objective measures: (1) Objective measures are more easily standardized from school to school and over successive periods of time; (2) it is easier to interpret the meaning of objective measures. The relatively low reliabilities reported in studies 1 and 2 are probably due to variable pilot performance from day to day as well as the change in testing conditions (weather, planes, etc.).

The Revised Check Ride

On the basis of the information obtained in the preceding two studies, a more comprehensive objective scale was constructed to include additional important maneuvers such as formation flying and stalls.

The first draft of this scale was constructed in cooperation with Standardization Board members at the Central Instructors School and at Pampa Army Air Field. Eighty-eight items covering 12 maneuvers were originally constructed and then criticized and modified by these expert pilots. After flight testing the measures with 5 students and receiving further criticisms from experts, a final form was drawn up of 75 measures covering 11 maneuvers. Tables A7.1 and A7.2 in the appendix give an inventory of the maneuvers, the variables measured, and one complete sample maneuver. It will be noticed that a good deal of emphasis is given to formation flying since this is a phase of contact flying which is highly important in tactical operations.

Due to the higher priority of research in the instrument area, no statistical data were obtained on this scale before the end of the war. It is believed, however, that this revised scale may serve as a helpful starting point for future research work in the contact phase of multi-engine flying skill.

Additional objective measures of two-engine contact flying skill will be found in the check-ride booklet which is reproduced in the appendix to chapter 10.

STUDY 3: INTERCORRELATIONS BETWEEN 19 OBJECTIVE MEASURES OF TWO-ENGINE FLYING SKILL

Problem

Intercorrelations were secured in an attempt to obtain information on the various factors involved in flying skill and the ways items should be combined to measure them.

Procedure

The intercorrelations of 19 objective measures, plus the pilot stanine, were obtained. These 19 measures made up a check ride designed to sample differences between large groups of students who had received different amounts of training at Advanced two-engine schools during the temporary training freezes. A copy of the check-ride booklet containing these measures is included in the appendix to chapter 10 which summarizes the differences found between the groups. The present section will deal with only the intercorrelations of these measures.

A total of 725 students in class 44-J at five two-engine schools was used in this study. They were just completing their 10 weeks of Advanced training in the TB-25 when given the objective check ride.

Results

Intercorrelations are presented in appendix tables A7.3 and A7.4. Inspection of these tables shows consistently low intercorrelations. With 725 cases, any coefficient with a value of ±.10 or more is significant at the 1 percent level of confidence. There are only 25 (13 percent) intercorrelations significant at this level of confidence. Some of these intercorrelations, such as the ones between air speed and altitude within the same maneuver, are higher because the measures are tied together aerodynamically; i. e., when a plane loses altitude it usually picks up air speed. The pilot stanine has practically no correlation with any of the items.

Two reasons can account for these results: (1) specificity of the skills being measured by the individual items, and (2) the low reliability of the items. From the low test-retest reliabilities secured in other studies it seems likely that this second factor was of greater importance in producing the low intercorrelations.

Effect of Heterogeneity of Items on Reliability of Total Scores

To the extent that these low intercorrelations are characteristic of objective measures and indicate heterogeneity of the skills measured, the amount of reliability gained from combing scores in different maneuvers will be less than that which would be predicted by the Spearman-Brown formula. In such cases the following more general formula should be used for predicting the reliability of the combination of standard scores on different items:

$$r_{(a+b)(a'+b')}=\frac{r_{aa'}+r_{ab'}+r_{a'b}+r_{bb'}}{2\sqrt{(1+r_{ab})(1+r_{a'b'})}}$$

in which: $r_{(a+b)(a'+b')}$ = the test-retest reliability of the combined score of items *a* and *b*.

$r_{aa'}$ = test-retest reliability of item *a*.

$r_{bb'}$ = test-retest reliability of item *b*.

$r_{ab'}$, $r_{a'b}$, etc. = the intercorrelations of scores *a*, *b*, *a'*, *b'*, etc.[44]

From an examination of this formula, it is apparent that when the intercorrelations of the items being combined equal their reliabilities ($r_{aa}=r_{ab}$, etc.), this formula reduces to the Spearman-Brown and, of course, predicts exactly the same gain. As the intercorrelations approach zero there is less gain from combining the items. When they are zero the reliability of the combined score is the average of that of the items. When the intercorrelations are negative, combination of items results in a loss of reliability.

[44] Jackson, Robert W. B., and Ferguson, G. A. *Studies on the Reliability of Tests.* Toronto, Department of Educational Research, University of Toronto. p. 56.

703322—47——13

The low intercorrelations reported in this section lead one to suspect that the factors measured may be heterogeneous, so that the gain from combining items will be less than that predicted by the Spearman-Brown formula. The average test retest reliability of the 14 maneuvers in study 1 in this chapter was 0.14 (computed by Fisher's z-transformation); the reliability of the sum of these 14 measures was only 0.25. From the Spearman-Brown formula one would expect the reliability of the total score based on these 14 measures to be far greater than 0.25. This seems to indicate that the factors being combined actually were heterogeneous so that the assumptions of the special case covered by the Spearman-Brown formula do not hold. If such specificity should be characteristic of all measures of flying skill, it would increase the difficulty of securing a reliable total score.

STUDY 4: AGREEMENT BETWEEN OBSERVERS AS A TEST OF OBJECTIVITY [a]

Problem

The low reliabilities obtained so far with objective measures of flying skill may have been produced by either: (1) Variable day-to-day performance or, (2) errors in measuring that performance. Evidence already reported in chapter 6 on Objective Measures of Flying Skill for the Primary Level of Pilot Training, indicates that the first of these was the greater source of unreliability for the measures tested. This experiment was designed to supply further evidence on the sources of unreliability by investigating the observer reliability of objective measures at the Advanced two-engine level of training. The observer reliability was determined by correlating the scores assigned by two independent observers who measured the same performance at the same time.

Procedure

The study was done between 13–20 November 1944 with class 44–1, flying the TB–25 at Brooks Field, Tex. The measures were the same 19 referred to in the preceding section as being designed to sample differences between large groups of students with different amounts of training received during a temporary training freeze. A copy of the check-ride booklet is given in the appendix to chapter 10. Nine enlisted men from the Pilot Project flew as second observers to score the maneuvers simultaneously with, but independently of, the check pilot. These nine men all received prior training and practice in scoring these measures at Randolph Field. The correlation between the independently derived scores of the two observers was calculated for each item of each maneuver and for the total score for each maneuver. Due to adverse weather conditions, the number of cases

[a] Lt. John K. Hemphill and T/Sgt. William O. Metheny were chiefly responsible for this experiment.

available on which all maneuvers were administered to the same student was so small that it was not worth while to calculate the total score for the completed check ride.

Results

In general, the correlations expressing the agreement between the two independent observers were satisfactorily high. Since the observations of a number of randomly paired observers are involved in these correlations, they indicate absolute as well as relative agreement. The number of cases, the means and sigmas for each set of observers, and the correlation coefficients expressing the agreement between the observers are given separately for each measure in table 7.3.

Two series of correlations are given: (1) The uncorrected Pearsonian coefficients for items scored in three categories and phi coefficients expressing the correlation in a fourfold point surface for those scored in only two categories, and (2) the Pearsonian coefficients corrected for the effects of grouping the data into the three broad categories used in scoring this scale. The uncorrected correlations are the best estimate of the reliability of a scale scored in these broad categories. The corrected coefficients are the best estimate of the reliability which would be obtained by scoring the scale in finer categories, i. e., by feet of altitude, miles per hour, etc. The corrected correlations must be used in comparing the observer reliabilities obtained in this study with those of other studies in which the measures were scored in finer deviation intervals. One of these other studies is the comparison of three methods of scoring basic instrument measures, which is reported in the next chapter.

Observations of the same instruments occurred in different maneuvers, e. g., altimeter readings in single-engine procedure, straight and level flight, etc. In some cases the reliabilities of these similar observations were different in different maneuvers. An examination of the conditions under which these observations were made suggests certain factors which may be related to the objectivity, or scoring reliability, of observations.

Air speed deviation was scored on three maneuvers: Single-engine procedure (item K), single-engine landing (item O), and instrument straight and level flight (item U). The reliabilties of this same measure in the three different maneuvers were 0.70, 0.25, and 0.69, respectively. During the single-engine landing maneuver there was a smaller range of differences in air speed so that it had to be measured in smaller units (5 instead of 10 m. p. h.) which were more difficult to observe. Furthermore, since part of this maneuver was done while approaching the ground, there was greater stress on the check pilot as a safety observer. These two conditions probably account for the lower reliability (0.25) of the air speed reading in this maneuver.

TABLE 7.3.—*Agreement between observers on the Advanced 2-engine scale*

	Check pilot		EM observed		N	r_{11}	$_c r_{11}$[1]
	Mean	*SD*	Mean	*SD*			
SINGLE-ENGINE PROCEDURE:							
J. Heading: (Maximum deviation)	2.25	0.70	2.11	0.75	65	0.62	0.84
K. Air speed: (Minimum A/S)	1.75	.63	1.68	.58	65	.70	.95
L. Altitude: (Maximum lost)	2.14	.86	2.17	.80	64	.89	1.00
M. Procedure check list	2.32	.95	2.48	.88	65	ϕ.61	ϕ.61
Total score	8.53	1.92	8.44	1.93	64	.79	.84
SINGLE-ENGINE LANDING:							
N. Turn onto approach lined up with runway	2.46	.89	2.12	.99	50	ϕ.53	ϕ.53
O. A/S on approach: (Maximum deviations from 140)	1.72	.66	1.54	.67	50	.25	.34
P. Was gear checked after lowering on approach?	2.14	.99	2.21	.98	58	ϕ.65	ϕ.65
Q. Where was power cut completely? (Check nearest answer)	2.18	.98	1.98	.98	57	.59	.80
R. Where was landing first made? (Goal is zone 2)	1.66	.72	1.70	.77	47	.81	1.00
S. Landed in skid	2.54	.84	2.49	.87	39	ϕ.51	ϕ.51
Total	12.97	3.45	12.42	3.46	33	.76	.79
STRAIGHT AND LEVEL FLIGHT: (INST.)							
T. Heading deviations	2.10	.79	2.02	.87	41	.60	.81
U. A/S variations	2.41	.70	2.37	.85	41	.69	.93
V. Altitude variations	2.27	.63	2.20	.80	41	.72	.97
Total	5.78	1.67	6.59	2.08	41	.81	.88
INSTRUMENT LETDOWN AND LOW APPROACH:							
W. Position—Plane's track in relation to ground markers	1.58	.76	1.67	.85	24	.95	1.00
X. Time—Difference between prescribed time and student's elapsed time past low cone	2.05	.88	1.95	.93	22	.95	1.00
Y. Altitude—Number of feet difference between prescribed altitude and student's indicated altitude	1.88	.88	1.83	.85	24	.42	.57
Total	5.58	2.00	5.47	2.21	19	.90	.98
POWER-OFF APPROACH AND LANDING:							
Z. Altitude at roll onto approach varied above or below the altitude (.. ft.) by	2.92	.34	2.88	.44	48	.48	.65
AA. Was power added on the apprach?	2.21	.98	2.17	.99	48	ϕ.98	....
BB. Where was landing first made	2.07	.79	2.14	.79	43	.80	1.00
Total	7.24	1.11	7.26	1.22	42	.80	.90

ϕ Phi coefficients.
[1] Corrected for broad categories, see: Peters, Charles C. and Van Voorhis, Walter R., *Statistical Pr--dures and Their Mathematical Bases*, pp. 393-399.

Altitude deviations were scored on four maneuvers: Single-engine procedure (L); straight and level flight (V); instrument let-down and low approach (Y); and power-off approach and landing (Z). The reliability coefficients were 0.89, 0.72, 0.42, and 0.48, respectively. Here again the coefficients vary with the conditions under which the observations were made. In the instrument let-down and low approach (Y) the units of measurement were only in 20-foot intervals and the check pilot was busy scoring other aspects of the maneuver and acting as safety observer in the heavy traffic conditions usually present when flying low over the field. In the power-off approach and landing (Z) only six percent of the students deviated more than 100 feet. This poor distribution probably accounts for the lower reliability.

Heading deviations were scored on two maneuvers: Single-engine procedure (J), and instrument straight and level flight (T). The reliabilities of the scoring of heading deviation on the two maneuvers were 0.62 and 0.60, respectively. The correlations are moderately low compared with those obtained on other measures. These lower observer reliabilities were probably produced by the fact that, since the students held their heading quite well during these maneuvers, it was necessary to use a small scoring interval (5°), which represents a relatively slight movement of the compass and hence is not easy to observe.

Zone of landing was judged on two maneuvers: Single-engine landing (R) and power-off approach and landing (BB). The reliability coefficients were 0.81 and 0.80, respectively. Observer agreement is high for both maneuvers indicating little difficulty in determining the 500-foot zone in which the plane landed.

Five measures of a somewhat different type were included in the scale. These measures were scored by encircling either 1 or 3 on the score card; the middle category (2) not being used. Phi coefficients were accordingly used. The reliability coefficients ranged from 0.51 to 0.96 for these five measures.

These measures included observations of sequences of procedure and also judgments such as whether or not the plane was lined up with the runway or whether or not it landed in a skid. Most of these were of a type which requires more training and experience for accurate observation than does the reading of instruments. The enlisted men used as second observers had a minimum of time in which to acquaint themselves with the sequences of procedure and other factors involved in these observations. This may account for the somewhat lower observer reliability of this type of measure. One of these measures, "Was power added on the approach?" was very easy to observe, and the high agreement between the experienced check rider and the less experienced enlisted man is reflected in a scoring reliability of 0.96.

Summary

The reliability of scoring of several objective measures of flying skill was investigated. These observer reliabilities are in general high, but readings from the same type instrument, e. g., altimeter, sometimes has a different reliability when used in different maneuvers. An analysis of these differences in the light of the experience gained in this study suggests three considerations which are important in constructing objective measures of flying skill:

1. The conditions under which the measure is to be scored must be well defined. For example, the beginning and end of the period during which the students are measured should be clearly specified.

2. The range of variation in performance of the student should be such that small errors in the actual reading of the measuring instrument become insignificant in comparison to the large differences between students.

3. The amount of division of attention demanded of the observer

should be kept as low as possible, especially during periods of stress when the check pilot must concentrate on the safety of the airplane.

Where the measures used in the Objective Scale of Flying Skill (Advanced two-engine) met the conditions specified in 1, 2, and 3, above, scoring reliability was high enough to demonstrate the possibility of creating measures of flying skill which are satisfactorily objective.

An inference from these findings is that the low test retest reliabilities reported elsewhere are probably due to the changing test conditions and variable pilot performance rather than to errors made in scoring and recording this performance.

SUMMARY

The importance and use of two- and four-engine aircraft continued to grow during the course of the war until by the end approximately two-thirds of all new pilots were graduating from Advanced two-engine schools. Preliminary experiments were conducted in this phase of pilot training for the purpose of finding the most efficient area for concentrating research. After these studies it was decided to concentrate work on developing and testing the objective measures of instrument flying skill which are described in the next two chapters. Four studies on the measurement of two-engine flying skill (contact) are reported in this chapter.

Study 1.—In cooperation with the Two-Engine Group at the Central Instructors School, Randolph Field, a subjective (rated work sample) phase check was administered to 69 student instructors. Using different check pilots for test and retest on different days, the reliabilities of the individual measures ranged from 0.00 to 0.26. Correlations of individual measures with the instructor's final grade were in the same low range.

Study 2.—In this experiment both objective and subjective measures were included in a check-ride booklet which was administered twice to the same student within the same half-day by two different check pilots. The median test-retest reliability coefficient for all the measures was 0.27 with the objective and the subjective items showing equal reliabilities. On the basis of this information, a more comprehensive objective contact check ride was constructed including objective measures of formation flying skill. The end of the war prevented the administration and statistical analysis of this revised scale.

Study 3.—The intercorrelations between 19 objective measures, plus the pilot stanine, were found to be extremely low. This was probably due to the unreliability of the separate measures.

Study 4.—The performance of students on each of 19 objective measures was scored simultaneously by two independent observers.

Comparing these with the seven less accurately scored measures suggested principles to be followed in reducing the observer error in objective measures.

The data in this chapter, like the results in the previous one, dealing with objective measurement at the Primary level, indicate that the low test-retest reliabilities of objective measures are more a function of changing test conditions and variable pilot performance than of errors in observing and scoring this performance.

CHAPTER EIGHT

Objective Measures of Multi-engine Instrument Flying Skill

Capt. Stanford C. Ericksen

INTRODUCTION

Reasons for Concentrating on this Area

Instrument flying refers to the control of the plane by cues from a group of instruments in the cockpit when the normal cues from outside the plane are obscured by conditions such as clouds or darkness. Skill in instrument flying was exceedingly important during the war since the tremendous mobility of the airplane as weapon or transport was useless whenever it had to be grounded because of bad weather. The engineering developments which produced planes capable of flying long distances through difficult weather, created a demand for pilots with a high level of instrument flying skill.

As has already been pointed out, the Pilot Project concentrated a considerable proportion of its work on developing objective measures of instrument flying skill. The importance of instrument flying was one reason for this decision; another was the fact that this type of flying seemed to be adapted to objective measurement because it was conducted solely with reference to instruments and radio aids.

The main work on developing objective measures of flying skill was started on an Advanced two-engine plane, the TB-25. This was selected for a number of reasons. One of these reasons was because any improvement in measuring proficiency at the Advanced level would be more useful in selecting pilots for special combat assignments than would an improvement at an earlier level. From two-thirds to three-fourths of the Advanced students were trained in two-engine planes, and instruments were especially important in flying heavy, long-range aircraft of this type. Furthermore, in these two-engine planes it was possible for the aviation psychologist to ride as an additional observer along with the student and his instructor.

After measures for two-engine planes at the Advanced level had been developed and evaluated, the knowledge and experience gained was used in planning work on single-engine planes at the Basic level. This is described in the next chapter.

Main Features of Instrument Training

The complexities of instrument flying have been described in chapter 3. It was necessary for the pilot to keep responding to a half-dozen different instruments. A deviation in any one of a number of different variables, for example, altitude, heading, or air speed, meant that he was not flying correctly. Furthermore, it was important for him to cross-check and interpret the readings from several different instruments.

The difficulty of instrument flying made it necessary to give the students extensive specialized training. Students were not introduced to instrument training until after learning the fundamental elements of flying by reference to the ground and horizon, i. e., contact flying. Instrument training started at the basic level. The student was placed under a cloth hood which eliminated outside cues and taught to rely solely on his instruments in flying the fundamental maneuvers such as straight and level, climbs, glides, and turns, and in recovering from unusual positions. Approximately 20 hours were devoted to this training at the Basic level.

At the two-engine Advanced level, the student received 20 more hours of instrument flying. The first few hours were spent in relearning on the more powerful two-engine plane the fundamental maneuvers practiced at Basic. The transition was usually fairly rapid and few students were eliminated because of failure to pass the "Basic Instrument Check" covering these maneuvers.

The remaining, larger portion of the time was spent in learning to use radio aids in problems such as navigation, flying on the beam, and making the instrument let-down and low approach which allowed the pilot to come down through an obscuring layer of clouds and break out into the clear immediately above the field in a position to land. This part of the training caused the poorer students considerable difficulty, since, in addition to holding precise limits of altitude, heading, and air speed, they had to interpret and respond correctly to a series of radio signals which were often obscured by static and difficult to distinguish.

Studies to be Described

The first objective measures of instrument flying skill constructed by the Pilot Project were part of a large-scale study of the effects of additional training on flying skill. The next study compared the effectiveness of three different methods of scoring instrument deviations. In this study the observer reliability, test-retest reliability, and correlation with instructor ratings were determined for representative measures.

On the basis of the experience gained in this work, a list was made of the procedures to be followed in developing objective measures of flying skill, and a 34-item check ride was developed for the TB-25. The measures in this ride were administered to students on each dual ride in order to determine (*a*) the reliability of each measure as shown by the correlation between the scores on odd and even days of administration, (*b*) the ability of selected measures to discriminate between different amounts of training, and (*c*) the correlation of measures with instructors' ratings.

In order to get preliminary information on additional measures and to secure answers to certain questions which interested the administrative staff at the school where the TB-25 studies were being conducted a 70-item check ride was constructed and administered to all students on the last day of their regular instrument training and again just before they graduated four weeks later. An experimental group of students who had been receiving the shorter scale on every dual ride was compared with a control group who had not used this scale. The scores of students at the end of the formal period of instrument training were compared with those four weeks later.

On the basis of this work, a refined scale for the TB-25 was constructed. The temporary disruption of training produced by the end of the war prevented experimental evaluation of this scale.

Finally some work was started on developing and giving preliminary try-outs to measures of four-engine instrument flying skill and an exploration was made of the possibilities of securing records by attaching a work-adder device to the flight-control cables.

RESULTS OF THE INSTRUMENT MEASURES USED IN A LARGE-SCALE STUDY OF THE EFFECTS OF ADDITIONAL TRAINING

The Pilot Project designed a number of measures of instrument flying skill to determine the effect of an additional 5 weeks of training given to all student pilots during temporary training freezes. The complete study is presented in chapter 10. Items to measure the accuracy of the instrument let-down and low approach (to landing) were developed since most instrument training was directed toward the development of skill on this final maneuver.

At the Advanced two-engine schools data were obtained on students with three different amounts of training:

- Group I: 10 weeks in lighter two-engine aircraft, AT-10 and UC-78, followed by 5 weeks in the TB-25.
- Group II: 10 weeks, all in the TB-25.
- Group III: 15 weeks, all in the TB-25.

The extra 5 weeks of training was found to improve the skill of the students in group III who had all 15 weeks in the TB-25. On the other hand, the students in group I who also had 15 weeks of Advanced training, but in two different type aircraft, were inferior to group III in the accuracy of the instrument let-down and low approach. In fact, they were also inferior to the students with only 10 weeks of Advanced training, all in the TB-25. Pilots have frequently assumed a high degree of positive transfer of instrument flying skill from one type aircraft to another. In general, this is probably true, but these results indicate the importance of being familiar with the particular type plane being flown as a prerequisite for precision instrument flying.

Measures were also developed at the Basic and Advanced single-engine levels to determine the way in which the extra five weeks of training effected both contact and instrument flying skill. The results, presented in chapter 10, show that in general more reliable differences were secured on the instrument than on the contact measures. This result confirmed the decision to concentrate further work on the instrument measures.

COMPARISON OF THREE DIFFERENT METHODS OF SCORING INSTRUMENT MEASURES [46]

One of the first problems systematically studied in developing objective measures of instrument flying skill was the question of how best to score and record instrument deviations in order to maximize the reliability and validity of these measures.

Statement of the Specific Problem

The problem was to determine how the following three different methods of scoring instrument deviations compared with respect to: observer reliability, test-retest reliability, and validity:

1. *Time sample method.*—Instrument readings of a student's performance were taken at specific, equally-spaced periods during the course of a maneuver. For example, deviations from the starting altitude were taken every 20 seconds. The student's score was the average of the deviations on all the readings of the appropriate instrument (altimeter, compass, or air speed indicator).
2. *Limits method.*—This was recording the student's maximum single deviation from the specified starting point on the instrument being observed, e. g., greatest number of feet lost (or gained) in altitude at any time during the maneuver.
3. *Range method.*—In this case the score represented the student's total range of deviation between the two extremes, e. g., difference between greatest deviation to the left and to the right of the initial heading, or difference between lowest and highest altitude reached during the whole maneuver.

In all these methods the observer scored the maneuver by making a mark at the appropriate place on a facsimile diagram of the instrument. When using the time sample method, a separate scoring sheet was used for each successive reading. The range and the limits scores could both be obtained from the same score sheet, as illustrated in figure 8.1. The check pilots were instructed to make a new mark

[46] Sgt. John R. Rohrs and Sgt. Maurice F. Connery are chiefly responsible for this study.

each time they observed a deviation in either direction larger than the greatest one previously recorded in that direction. In the example in figure 8.1, the student's range score for altitude was 340 feet, his limits score 240 feet.

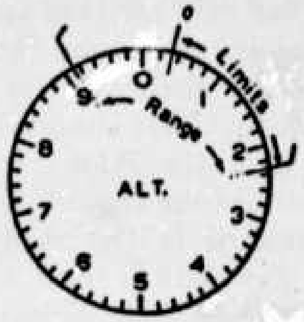


FIGURE 8.1—*Sample score sheet used for the range and limit methods of measuring altitude in the steep turn. (Reduced to ½ size used.)*

While the range and limits readings could both be obtained from observations of the same performance, the maneuver had to be repeated to derive the time sample score. The sequence of repetition was counterbalanced throughout the study.

The Maneuvers on Which the Students Were Tested

Straight and level course with changing air speeds.—The trial started at the student's signal when he was flying straight and level at 6,000 feet on a heading of 360° and at an air speed of 200 m. p. h. The total maneuver was made up of four separate parts but throughout each of these parts, the task remained the same: to maintain the original heading and altitude. The four parts with a common task were:

a. During the first 2½-minute period the student reduced air speed from 200 m. p. h. to 160 m. p. h.

b. He held this air speed for ½ minute.

c. During the next 2½ minutes air speed was increased back to 200 m. p. h.

d. For the last ½ minute, the student held this air speed. Separate scores were obtained for each of these four legs.

Two steep turns.—Starting at 6,000 feet, at cruising throttle, the student made a 45° bank turn to the left, holding this bank and continuing the turn for one minute. The maneuver was then repeated to the right. The student was graded on altitude control and holding the 45° bank.

Procedure

The experiment was conducted at the Advanced two-engine school at Brooks Field, Tex., in December 1944. The instructor and supervisory personnel selected the 30 best and the 30 least proficient instrument students from class 44-J currently receiving the extra 5 weeks of TB-25 training resulting from the second training freeze. Because the students had only received an average of approximately 5 hours instrument training at Brooks Field, the separation of these two criterion groups was probably not as good as that which could be obtained under more favorable circumstances.

Ten enlisted men from the Pilot Project received preliminary training and then served as the second observers for the ten Brooks Field check pilots assigned to the study. Each check pilot and second-observer team flew with a total of six students—three from the most and three from the least proficient group of students. Equal numbers of good and poor students were flown during each half day. No check pilot flew with his own students, nor did he know from which of the two extreme groups they came. All students and check pilots were carefully briefed prior to the flights.

Results

Observer reliability.—The first analysis was to determine the observer reliability of the three methods by comparing the scores recorded by the check pilot with those recorded by the second observer. The only maneuver on which observer reliability data were available was the straight and level course with changing air speed. During the other maneuver, the steep turn, the check pilot functioned as a safety observer and only the second observer recorded the deviations.

The data on observer reliability are given in table 8.1. When interpreting these data it should be recognized that the time sample score is the mean of 22 observations while the range and limits score are based on the maximum deviation occurring at any time during the maneuver.

For both the heading and altitude measures the time sample score gave the highest observer reliability. This was expected since this method provided rigid control as to just when and what deviations were recorded. The observer did not have to try to keep track of all of the instruments all of the time. This narrowing down of the task of the observer increased the accuracy of observation. It was also desirable to determine its effect on the test-retest reliability and the validity of the observations.

Test-retest reliability.—An estimate of test-retest reliability (or perhaps it should be called internal consistency) was obtained by correlating the right versus the left steep turn and the first two legs versus second two legs of the straight and level course with changing air speed. These data are given in table 8.2 and are already adjusted for double length by the Spearman-Brown correction.

Since these coefficients were based on measurements taken from the same trial, variables such as turbulence, student day-to-day variability and plane differences, were held at a minimum. In this sense these are maximum estimates of test-retest reliability.

TABLE 8.1.—*Correlation between the scores recorded by two observers during the straight-and-level course with changing air speed*

N=60

Method of scoring	Heading deviation	Altitude deviation
Time sample: Average of deviations observed at 22 specified times	0.90	0.87
Range: Difference between extreme deviations in each direction	.86	.69
Limit: Greatest single deviation in either direction	.84	.79

Means and standard deviations are given in appendix table A8.1.

TABLE 8.2.—*The test-retest reliability of the three different methods of scoring based on tests administered in immediate succession to 60 students*

N=60

Method of scoring	Steep turn altitude	Straight and level	
		Heading	Altitude
Time sample	0.37	0.82	0.67
Range	.61	.80	.56
Limits	.68	.75	.61

Means and standard deviations are given in appendix table A8.2.

It will be noted that the time sample method did not produce consistently higher test-retest reliabilities than the range or limits methods. Though limiting the observations to certain specified times produced a higher observer reliability than continuous observation, the fact that the behavior between observations was neglected apparently offset this advantage so that the test-retest reliability was no higher.

Validity.—Which of these three methods of scoring best differentiates the two criterion groups? Table 8.3 gives the validity coefficients for the three measures together with an indication of their level of significance.

TABLE 8.3.—*Point-biserial correlations between the measures and the criterion*

N=60

	Steep turn altitude	Straight and level	
		Heading	Altitude
Time sample	**0.26	***0.30	0.16
Range	***.33	***.31	**.28
Limits	**.29	.15	*.20

Probability of obtaining a positive correlation of this size by chance alone:
No asterisk = more than 10 times in 100.
* = 10 times in 100 or less.
** = 2.5 times in 100 or less.
*** = 0.5 times in 100 or less.

Means and standard deviations are given in appendix table A8.3.

Examination of this table shows that the range method of scoring correlates somewhat higher with the criterion than do the other methods.

The next question was to determine the significance of the differences between the validities obtained from the use of the three methods. Table 8.4 gives the probabilities that differences as large as those observed would be expected to occur by chance alone. Since the range and limits scores were highly correlated, in computing the standard error of the differences between these correlations, the intercorrelation of the scores upon which they are based was taken into account. The correction formula used is given on page 218 of *Statistical Analysis in Educational Research* by E. F. Lindquist, 1940, Houghton Mifflin Co., N. Y.

TABLE 8.4.—*Probability of securing by chance differences between the three methods of scoring as large as those observed (see table 8.3)*

	Steep turn altitude	Straight and level	
		Heading	Altitude
Range versus limits	0.16	0.01	0.07
Range versus time sample	.25	.44	.26
Limits versus time sample	.44	.13	.45

From the results in tables 8.3 and 8.4 it can be concluded that the range method is slightly better than the limits method. While the time sample method has higher observer reliability, it does not have higher validity and, in addition, is more difficult for the check pilot to administer in flight when he is also busy acting as safety observer, instructor, and copilot.

Comparison of the Total of Four Part-Scores Versus a Single Maximum Deviation Score

The range and limits scores analyzed up to this point were based on the largest deviation occurring during any part of the entire maneuver. As was described previously, the straight-and-level course was made up of four legs: Reducing air speed, holding this reduced air speed, increasing air speed, and maintaining the final air speed. Throughout the maneuver, the student's task was to maintain the original heading and altitude. On each of the four legs separate range and limits scores were obtained. A total score was then computed which was the sum of the maximum deviations occurring during each of the four legs. This total score approaches the time sample technique but with the decided advantage that the observer maintained constant watch and obtained a record of the student's largest errors rather than simply recording what deviations existed at specified time intervals.

The same statistical analyses were applied to the total scores as to the single scores described in the preceding pages. The results of these two types of scores are presented in table A8.4 in the appendix. It can be concluded that the two types of scores are approximately equal. The differences are small, inconsistent, and could easily have occurred by chance. Under the conditions of this experiment, a measure of a student's largest single deviation during a maneuver seems to be as accurate an index of his instrument flying ability as a series of four readings taken from the same maneuver and added together.

An incidental finding was that the test-retest reliability of the altitude measures were less reliable than the heading measures in the straight-and-level course with changing air speeds. This can perhaps be understood in terms of the opposite effect produced by decreasing power during the first half of the maneuver while adding power during the second half.

Summary

1. The validity coefficients obtained for the separate items in this study are high enough to be promising for single items to be included as one of a series of objective measures of instrument flying skill. This validity was established despite the fact that the criterion groups were selected by instructors who had on the average only 5 hours of dual time in which to observe their students' instrument-flying ability.

2. The range method gave slightly higher validity coefficients than the limits method although there was no difference in either observer or test-retest reliability.

3. The time sample validities were no greater than those of the range method.

4. The time sample method had higher observer reliability than the other two methods. Its test-retest reliabilities, however, were about the same as those for the other two methods. This high observer reliability did not, in this case, give the time sample method a higher test-retest reliability or validity than the range method.

5. There were no consistent differences in observer reliability, test-retest reliability, or validity between two types of scoring, one based on the largest single deviation during the entire maneuver and the other a total of the largest deviation during four parts of the maneuver, straight-and-level course with changing air speeds.

PROCEDURE TO BE FOLLOWED IN DEVELOPING A COMPREHENSIVE CHECK RIDE TO MEASURE INSTRUMENT-FLYING SKILL

Before constructing instrument items for a comprehensive check ride of instrument-flying skill, a general outline of procedures was prepared, based on the accumulated experience of the Pilot Project working in all phases of objective-scale construction and experimentation. The optimum sequence in the development of objective measures can be summarized as follows:

1. Analysis of the fundamental skills essential to practical success in operational instrument flying and analysis of the curriculum used to teach these skills.

2. Construction of tentative items with the aid of technical experts.

3. Testing the items with a few students in order to determine practicability and to refine administrative details, scoring, and sequence of maneuvers.

4. Administration to groups large enough to allow statistical item analysis. The reliability of each measure is obtained by two or more administrations of the test. The validity of each measure is obtained by any one or a combination of three methods: Testing of two extreme groups selected on the basis of known flying skill; testing two groups with widely different numbers of hours of training; testing of a large enough group to permit correlations between instructors' subjective ratings of flying ability and objective scores. Intercorrelations may also be obtained at this stage to eliminate duplication of items measuring the same skills. Normative data are secured in order to help in setting class intervals for scoring weights.

5. On the basis of the statistical data provided by the preceding analysis, the best measures are selected and refined. Then these are tried out again on a large number of students as a second check on the results already obtained. This final testing will profit from the experience gained about how best to administer objective check rides within the normal student-training schedule; briefing and supervision of instructors and students; vital supplementary information such as weather reports, training curriculum, and flying schedules. The main purpose of this step is to obtain conclusive data on the validities, reliabilities, and intercorrelations of the measures so that they can be weighted and combined into total scores to produce the most efficient results.

RESULTS OF AN OBJECTIVE INSTRUMENT CHECK RIDE GIVEN ON EACH DUAL RIDE [47]

The goal of the next stage of research was the development of an objective check ride to measure instrument flying skill in the TB-25. While the Psychological Research Project (Pilot) was primarily interested in determining the feasibility of objective measures of instrument flying, a secondary aim was the construction of a standardized check ride which could be used by supervisory and instructor personnel in making the prescribed periodic checks of instrument flying proficiency.

[47] Capt. Wm. V. Hagin and S/Sgt. W. W. Ismael of the Pilot Project, and Capt. B. N. Henderson and Lt. Jack O. Dean of the Instrument Training Squadron, Pampa Army Air Field, assisted in the construction and administration of the instrument measures described in this section. Sgt. Ezra Glaser was responsible for supervising the statistical analysis.

A comprehensive study was made at the Advanced two-engine school at Pampa, Tex., to determine the reliability, validity, and practical value of objective measures of instrument flying skill. This, of course, required the preliminary work of test development summarized as steps 1, 2, and 3 in the preceding section. The final statistical analysis of the data obtained at Pampa Army Field represents the completion of step 4. The end of the war occurred before work could be started on the final phase, step 5, refining the measures for application in the routine training situation.

Problem

The purpose of the first phase of the study was to determine the reliability and validity of the objective measures of instrument flying skill when administered on each dual ride. It was highly advantageous to repeat the objective measurement on as many rides as possible since other studies had shown that the low reliabilities of objective measures were caused by factors producing day-to-day variability in performance. By measuring performance on each dual ride for a number of different days, these factors could be averaged out and the reliability increased.

The problem of securing a satisfactory criterion for validating the objective measures was complicated by the fact that each student usually had two or three instrument instructors, no one of whom could be expected to give an accurate subjective report of a student's instrument flying skill. Nevertheless, the instructors' rankings of their students' instrument flying ability were correlated with the average objective scores. A second validity analysis compared the objective scores made during early versus later hours of instrument training.

Procedure

The check-ride booklet.—A 34-item check was designed to cover the fundamental instrument maneuvers which were practiced at least once during each instructional flight. A copy of this booklet is presented in appendix table A8.5. Figure 8.2 gives a sample set of measures showing how they are marked by the check pilot.

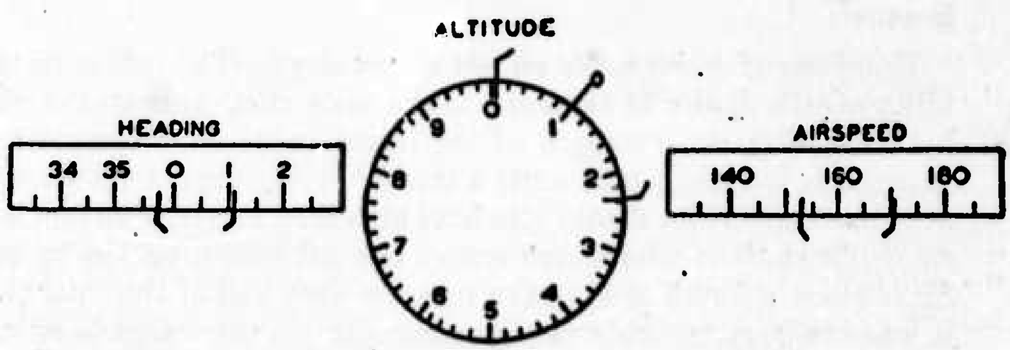


FIGURE 8.2.—Sample measures with deviation markings by the check-pilot.

To score these measures, the check pilot simply marked the highest and lowest deviations on the altimeter, air speed, and heading dials. Several marks in the same direction of deviation could be recorded as the student continued to increase the amount of altitude, air speed, or heading variation during the maneuver. The check pilot indicated the exact starting altitude on the dial by a line with a small zero on the end (?).

Testing procedures.—The experiment was designed to fit into the training conditions with minimal disruption of the routine flying schedule of students and instructors.

At least 20 hours of Advanced flying training were spent in instrument flying, divided between 13 hours dual (with instructor) and 7 hours so-called solo (without instructor but with a second student, who was not under the hood, watching to avoid collisions with other planes). Instrument training was given on alternate days during the first 5 weeks at the Advanced school. When not receiving instrument training, the students continued transition flying with their contact squadrons.

Ordinarily, at least six dual flights were given during the 5 weeks of instrument training. The instructors gave the objective check during the first part of each dual ride—before instructing the student on that day's lesson. Some of the radio measures could not be graded during the first rides because the training on radio procedures had not yet begun.

The subjects used in this study were the 80 students in flights G and H in class 45-D.

Both the students and the instructors were briefed thoroughly on the nature of the objective scale. The instructors were indoctrinated as to how the measures were to be interpreted, administered, and graded. Representatives from the Pilot Project were present at all times to assist in the administrative details, maintain a log of weather conditions and flight schedules, and to facilitate the maintenance of standard testing conditions in all aspects of the study.

Results

Reliability of scores made on odd v. even days.—The odd-even reliability of each of the 34 measures in the daily check ride was obtained by correlating the averages of the scores made on alternate rides. If a test is reliable (consistent) a student who earned high scores on even numbered rides should also have high scores on his odd numbered rides; the student who scores low on one set should be low on both. Since either the odd or the even rides are only half of the total given, it was necessary to apply the Spearman-Brown correction to estimate the reliability of both sets of rides taken together. Table 8.5 on the following pages presents these corrected coefficients. Appendix table A8.6 gives the more detailed summary statistics.

TABLE 8.5.—*Reliability of each objective measure in the TB-25 daily check ride*

[Based on Average Scores of Three Odd v. Three Even Days Corrected for Double Length [1]]

Measures	Test-retest reliability	N
A. *Instrument take-off:*		
1. Instructor help keep on runway	0.06	68
2. Attitude in initial climb	.26	70
3. Air speed range in climb to 4,000 feet	*.29	72
B. *Spiral climb:*		
4. Air speed in climb 4,000-5,000 feet	−.21	74
5. Air speed in climb 5,000-6,000 feet	−.02	75
6. Air speed in climb 6,000-7,000 feet	.17	76
C. *Level out:*		
7. Heading during level out	**.39	76
8. Altitude during level out	***.48	76
D. *180° turn:*		
9. Time error in 180° turn	.17	77
10. Altitude range in 180° turn	***.64	77
11. Heading at roll-out of 180° turn	−.10	77
E. *Partial panel turn:*		
12. Time error in partial panel turn	.18	77
13. Altitude range in partial panel turn	***.56	75
14. Heading at roll-out in partial panel turn	.18	77
F. *Power let-down:*		
15. Heading range during air speed change	*.31	71
16. Altitude range during air speed change	.25	76
17. Did student complete GUMP check	.26	76
18. Heading range at level-out	*.36	71
19. Altitude range at level-out	***.48	76
20. Air speed range at level-out	.23	76
G. *Single-engine operation:*		
21. Heading range during single-engine operation	.21	75
22. Altitude range during single-engine operation	*.33	75
23. Lowest air speed reached during single-engine operation	*.31	72
H. *Radio procedure:*		
24. Tune in and identify proper station	−.06	67
25. Check directional gyro	***.56	60
26. Heading range while tuning radio	−.06	61
27. Altitude range while tuning radio	*.35	65
28. Air speed range while tuning radio	.17	60
I. *Quadrant identification:*		
29. Recognize fade or build	**.50	64
J. *Beam crossing:*		
30. When started turn after crossing beam	−.26	66
K. *Station recognition:*		
31. Error in seconds when crossed cone	*.37	57
L. *Let-down and low approach:*		
32. Position when signal over field in let-down	−.29	27
33. Time error of let-down from cone to field	.11	26
M. *Voice procedure:*		
34. Use proper voice procedure	.12	16

Probability of obtaining a positive correlation of this size by chance alone:

No asterisk = more than 10 times in 100.
* = 10 times in 100 or less.
** = 2.5 times in 100 or less.
*** = 0.5 times in 100 or less.

Means, standard deviations, uncorrected odd-even reliabilities and significance levels are given in the appendix table A8.6.

[1] For the radio measures (24-34) only 2, 3, or 4 separate scores could be obtained, since these procedures were introduced later in training.

The results in table 8.5 show that, as a group, the altitude measures have the best reliabilities, since 6 of the 10 measures with positive correlations significant at the 2.5-percent level are altitude items. Table 8.6 gives the distribution, showing the range of corrected odd-even reliabilities for the 34-item objective scale.

TABLE 8.6.—*Distribution of corrected odd day-even day reliabilities*

[Median = 0.25]

Corrected reliability	Frequency	Corrected reliability	Frequency
Below −0.20	3	0.31 to 0.40	7
−0.11 to −0.20	0	0.41 to 0.50	3
±0.00 to −0.10	4	0.51 to 0.60	2
±0.00 to +0.10	1	Over 0.60	1
+0.11 to +0.20	7		
+0.21 to +0.30	6	Total	34

In the original design of the experiment it was planned to have each student graded by the same instructor on all but one ride, on which he would be graded by a different instructor, in order to determine the effect of instructors on day-to-day reliability. However, it was administratively necessary for the instrument-training supervisory personnel to switch instructors from student to student at frequent and uncontrolled intervals throughout the 5 weeks of instrument training. As a result, most students had either two or three different instructors on their 6 dual rides when the objective checks were given. The odd-even reliabilities reported in table 8.5 include any effect resulting from this changing of instructors which was fairly well randomized throughout the instrument training.

In order to determine whether having different amounts of flying time for the odd and even days had any effect on the reliability coefficients, semipartial correlation coefficients partialing out total TB–25 time, were calculated for three measures, Nos. 9, 13, and 15, having low, high, and intermediate reliabilities. The reliabilities could have been distorted if some students had consistently more flying time (thus more possibility for learning) on one set of rides than on the other. Since the semipartial correlations were found to be almost identical with the uncorrected reliabilities, no further control for the "flying time" factor was made. The results are given in appendix table A8.7.

The range of reliabilities in table 8.5 is about the same as that obtained in other studies in which the students were given tests on two different days. On the basis of these previous results one might have expected the reliabilities of measures given on 6 days to be somewhat higher. The distribution of low reliabilities presented in table 8.6 emphasizes the importance of the many variable factors operating in pilot testing procedures: different planes from day to day; variable weather and turbulence; different kinds and amounts of interpolated flying; different instructors; different motivation and attitudes, and differential rates of learning. The homogeneity of the group made the influence of these factors greater than they might have been with a group of students representing a wider range of flying abilities. The students tested had all been screened with the same battery of aptitude tests and had received similar training in the Primary and Basic schools, where the weaker students were eliminated.

Comparison of scores made at two different levels of training.—One type of validity of the objective instrument measures can be estimated by comparing scores made at two different levels of training. Presumably students should make better instrument scores as training progresses and the valid measures should show these differences. A comparison was therefore made between objective scores obtained on the second versus the sixth rides.

Twenty-two measures were selected for further analysis on the basis of their learning curves, and on the basis of whether or not they covered an important area. The results are presented in table 8.7. This table shows the level of significance of the mean difference between the two rides.

Of the 22 measures analyzed, 13 are significant at the 1 percent level, and 17 at the 5 percent level, leaving only 5 objective measures with differences which might reasonably be expected by chance alone. Four of these 5 nondiscriminating measures are radio items in which the comparison was between the fourth and sixth rides rather than between the second and sixth as with the nonradio measures. (Radio procedures were not introduced until later in training.)

These results can be interpreted as demonstrating that the objective measures of instrument flying skill reflect improvement in proficiency and can be accepted as valid within the limits of the study. It may be concluded that these objective measures are sufficiently sensitive to measure any important differences between groups which have received different amounts or types of training or different opportunities to forget.

Correlation between objective scores and instructors' subjective rankings.—At the end of training, each instructor ranked his students on the basis of over-all instrument flying ability. These rankings were then used as a criterion against which to validate the objective scale scores. The objective score used for each student was the average of

TABLE 8.7.—*Differences between objective instrument check-ride scores on early and late rides* [1]

Measures	N	Mean difference [2]	S. D. mean difference	P [3]
2. Attitude in initial climb	32	1.000	0.370	0.0030
3. Air speed range in climb to 4,000 feet	32	.938	.265	.0002
4. Air speed in climb 4,000-5,000 feet	31	.912	.197	.0000
5. Air speed in climb 5,000-6,000 feet	39	.718	.176	.0001
6. Air speed in climb 6,000-7,000 feet	42	1.094	.171	.0004
7. Heading during level out	49	.714	.162	.0000
8. Altitude during level out	49	.429	.168	.0050
9. Time error in 180° turn	50	.200	.118	.0160
10. Altitude range in 180° turn	48	.313	.233	.0000
13. Altitude range in partial panel turn	48	1.104	.212	.0000
16. Altitude range during A/S change	45	1.010	.240	.0000
17. Did student complete GUMP check	47	.596	.148	.0000
19. Altitude range at level out	47	.340	.173	.0210
21. Heading range during S-E operation	46	.870	.259	.0004
22. Altitude range during S-E operation	45	.980	.217	.0000
23. Lowest A/S reached during S-E operation	41	.195	.218	.1900
25. Check directional gyro	35	.400	.180	.0130
26. Heading range while tuning radio	32	.154	.213	.2300
27. Altitude range while tuning radio	36	.667	.279	.0030
28. Air speed range while tuning radio	30	.433	.197	.0240
29. Recognize fade or build	31	−.065	.144	....
31. Error in seconds when crossed cone	27	222	.404	.2900

[1] Second and sixth daily checks for basic measures (numbers 2-23); fourth and sixth for radio measures (numbers 25-31).

[2] The mean difference was computed from the algebraic sum of the differences (score more training minus score less training) of coded scores, and the S. D. was computed from this series of differences. A positive number means that the students were better after more training.

[3] Probability of obtaining by chance a difference in favor of more training as large as the one observed.

all his daily rides. Only those measures with reliability coefficients of 0.29 or better were used for this analysis. The results are presented in table 8.8.

TABLE 8.8.—*Correlation between average daily objective instrument scores and instructor's subjective rank order*

Measure	N	r	P[1]
3. Air speed range in climb to 4,000 feet	71	0.18	0.07
7. Heading during level out	71	.15	.11
8. Altitude during level out	71	.04	.37
9. Time error in 180° turn	71	.24	.02
10. Altitude range in 180° turn	71	.23	.03
13. Altitude range in partial panel turn	71	.18	.07
15. Heading range during A/S change in power let-down	71	.09	.23
18. Heading range at level-out in power let-down	71	.03	.40
19. Altitude range at level-out in power let-down	71	.14	.12
22. Altitude range during single-engine operation	71	.04	.37
23. Lowest air speed reached during single-engine operation	70	.06	.30
Radio Procedures			
26. Check directional gyro	69	.20	.05
27. Altitude range while tuning radio	71	.04	.38
29. Recognize fade or build	70	.09	.24
31. Error in seconds when crossed cone	68	.16	.09

[1] Probability of obtaining positive coefficients of this size by chance alone.

Only 3 of the 15 measures reach the 5-percent level of significance for positive correlations. The failure of the other measures can probably be accounted for by two conditions:

(1) There were a great many changes between instructors and students during the period of instrument training. Several instructors reported that they had flown with their originally assigned students on only one or two rides. Therefore, not much weight can be placed on the accuracy of these subjective ratings.

(2) Being an average of all rides, the objective scores reflect both early and late performance. The instructors' ratings, on the other hand, were based on final instrument-flying performance or final ability estimated on the basis of a few rides in the earlier stages of training.

Comparison of the range and limits methods of scoring.—Earlier it was pointed out that the range score includes the maximum deviations in both directions while the limits score is based on the greatest single deviation from the initial instrument reading. The problem was to determine which of the two methods showed the greatest differences between student performance on the second versus the sixth rides. The data were taken from the 34-item daily objective instrument check ride which permitted both types of scoring. Ten items representing measures of heading, altitude, and air speed were used in this analysis. It was assumed that the method which reflected the greatest improvement was the more valid measure of instrument flying skill. The results are presented in table 8.9.

Examination of this table shows that larger differences were secured by the range method on six measures and by the limits method on four. What little difference there is in the over-all picture seems to be in favor of the range method again, but the differences are so inconsistent that they could easily have been produced by chance.

RESULTS OF A COMPREHENSIVE OBJECTIVE CHECK RIDE TO MEASURE FINAL INSTRUMENT FLYING SKILL

The personnel supervising instrument training were particularly anxious for the objective measures to be used to yield evidence relevant to the following problems:

a. Does the use of a standardized daily check improve the final level of instrument flying proficiency?

b. How well is instrument flying skill retained after formal instruction has ended?

c. Is it feasible to construct a comprehensive standardized check ride for use in evaluating the instrument flying proficiency of students prior to graduation?

To help answer these problems the Pilot Project constructed a 70-item final instrument check ride. The maneuvers selected for testing were those prescribed by AAF memorandum 50-3, which defines the standardized instrument flight check used throughout the Air Forces to check the instrument flying proficiency of rated pilots at yearly intervals, and which was also used to evaluate student pilots just prior to graduation.

TABLE 8.9.—*Differences between Second and Sixth dual rides as measured by the range and the limits methods*

Measure [1]	Limits method				Range method			
	N	Mean difference [2]	S. D. mean difference	*P* [3]	*N*	Mean difference [2]	S. D. mean difference	*P* [3]
3. Air speed range in climb to 4,000 feet	32	0.47	0.22	0.018	32	0.94	0.27	0.000
5. Air speed in climb 5,000–6,000 feet	42	.52	.17	.001	39	.72	.18	.000
7. Heading during level-out	47	.64	.33	.025	49	.71	.16	.000
10. Altitude range in 180° turn	48	.40	.18	.016	48	.31	.23	.000
13. Altitude range in partial panel turn	47	1.19	.21	.000	48	1.10	.21	.000
16. Altitude range during air speed change	47	.96	.25	.000	45	1.04	.24	.000
19. Altitude range at level-out	48	.17	.16	.147	47	.34	.17	.024
21. Heading range during S-E operation	46	.78	.28	.003	46	.87	.26	.000
26. Heading range while tuning radio	32	.19	.19	.156	32	.16	.21	.230
28. Air speed range while tuning radio	30	.67	.24	.002	30	.43	.20	.024

[1] Second and sixth daily checks for basic items (numbers 3-21); fourth and sixth for radio items (numbers 26 and 28).

[2] The mean difference was computed from the algebraic sum of the differences (score more training minus score less training) of coded scores and the S. D. was computed from the series of differences. A positive number means that the students were better after more training.

[3] Probability of obtaining by chance a difference of this size in favor of the scores based on more training.

The measures were scored in the same manner as in the 34-item daily check ride. In order that the student would know exactly how to make his best score, the first page of this longer check ride had a scoring box showing the score points for each unit of instrument deviation. The score points for the procedural measures were indicated in parentheses with each item. Twenty of the measures were identical with those included in the short daily check-ride booklet.

Effect of Daily Objective Check Ride on Proficiency at the End of Instrument Training

The personal reports of instructors using the daily check ride were favorable. After the completion of the Pampa experiments, many instructors continued to use the booklet as a guide for their teaching and grading. The objective instrument readings and procedures were particularly helpful during the recapitulation of the training flights after the instructor and his students returned to the ready room. The instrument training personnel suggested that this type of grading system be modified for use as the daily grade slip. In the light of this favorable comment, the administrative personnel wanted to secure any evidence possible on the effect of using the 34-item objective check ride during the first part of each period of dual instruction.

Since only one of the two flights in the instrument training squadron used the 34-item objective check, it was possible to determine the effect of the use of this check on student proficiency at the end of instrument training when both flights were given the 70-item objective final check. Fourteen sample measures were selected for this analysis, nine measures were common to both the daily and the long final check, while five occurred only in the objective final check. The results are presented in table 8.10.

In 11 out of the 14 measures, the experimental flight (using the daily objective tests) was superior to the control flight. The differences on the three measures which favored the control flight were all small. On 5 of the 14 measures, the difference in favor of the experimental flight was significant at the 1-percent level or above. These five items represent measures of air speed, altitude, heading, and procedures, but no clear-cut explanation, in terms of the pilot skills being measured, is apparent to indicate why these particular measures should show significant differences while others do not.

One of the main goals of instrument flying is precise control of heading, air speed, etc., as indicated by deviations on the instruments. This is exactly what the objective items measured. It seems likely, therefore, that the superiority of the experimental group represents a significant improvement produced by the more rigorous conditions of daily grading. The effects of this procedure on other aspects of the performance remain to be determined.

Level of Instrument Flying Skill 4 Weeks After the End of Formal Instrument Training

Approximately 4 weeks after formal instrument training, the long 70-item check was again administered to the same students the last week before graduation. At this time the students were given a final check on their instrument flying skill by repeating the AAF 50-3 check ride which they had to pass again in order to graduate. The objective check-ride booklet was administered by an instructor who had never flown with the student before. Ten measures representing the more common types of items (altitude, heading, air speed, procedure) were selected for this analysis. The cessation of hostilities prevented more complete analyses using all of the 70 measures in the objective final check-ride booklet. The results are presented in table 8.11.

TABLE 8.10.—*Differences between scores on the final instrument check ride of students using the daily objective check and a control group*

Measure [1]	N		Mean difference	S. D. difference	P [3]
	Control group	Experimental group			
3. Air speed range in climb to 4,000 feet	56	52	0.91	0.187	0.001
7. Heading during level-out	64	67	[2] −.16	.214	....
13. Altitude range in partial panel turn	62	68	.52	.217	.008
15. Heading range during air speed change	58	56	.63	.221	.008
17. Did student complete GUMP check	56	65	.02	.187	.460
22. Altitude range during S-E operation	60	65	.32	.249	.100
28. Air speed range while tuning radio	64	66	.19	.063	.002
31. Error in seconds when crossed cone	59	60	[2] −.18	.293	....
32. Position when signal over field	61	65	.17	.173	.160
A. Altitude in steep turn	66	68	.28	.219	.100
B. Beam following	61	67	.09	.053	.050
C. Air speed in-bound to low cone	57	68	.81	.182	.003
D. Position at low cone	54	64	[2] −.08	.246	....
E. Altitude at field	49	57	.30	.239	.108

[1] Measures 3 through 32 were common to the daily and the final check rides; measures A through E were in the long check only.
[2] The control group, which did not use daily objective check, secured a higher score than the experimental group which used this check-ride.
[3] Probability of obtaining by chance a difference of this size in favor of the experimental group.

TABLE 8.11.—*Comparison of objective scores on the 70-item check rides separated by an interval of 4 weeks*

Measure	N	Mean difference	S. D. difference	P [1]
A. Air speed in initial climb	91	0.15	0.100	0.108
B. Altitude in steep turn	127	−.09	.138	....
C. Altitude in S-E operation	115	−.00	.170	....
D. Heading in level-out	129	.23	.133	.018
E. Altitude in partial panel turn (left)	123	.05	.139	.345
F. Position at low cone	116	.60	.184	.003
G. Procedure in let-down	113	.02	.140	.418
H. Heading in let-down	99	.10	.188	.295
I. Air speed in-bound to low cone	122	−.02	.132	....
J. Position over field	111	.04	.133	.394

[1] Probability of obtaining by chance a difference of this size in favor of the tests given at the end of training.

In general, the differences between the two administrations were small and inconsistent. Seven out of the ten differences were in favor of the scores obtained at the end of training; two of these were significant at approximately the 1 percent level. This would seem to indicate that general effects of additional contact training and the specific instrument practice, which the students gave themselves during the solo periods immediately preceding the final check, counteracted any tendency to forget during the 4 weeks following the end of formal instrument instruction.

SUMMARY OF OBSERVATIONS MADE DURING ADMINISTRATION OF THE OBJECTIVE INSTRUMENT CHECK RIDES

In addition to the statistical results of the Pampa studies, the following information was gained about the construction and administration of objective measures of flying skill:

1. Insofar as possible, the measures should be constructed and developed in close conjunction with the personnel at the station having the responsibility of final administration. This helps to indoctrinate these pilots and to motivate them to accept the controls required for experimental administration, e. g., switching instructors for the final checks.

2. Whenever possible the items should be adapted to the specific conditions of the school giving the checks, e. g., traffic pattern altitudes, air speed, etc. This makes it easier to obtain maximum clarity of instructions and helps to minimize the novelty of the check ride.

3. It is sometimes necessary to compromise strict objectivity to include subjective measures of vital aspects of some maneuvers which are not adapted to objective grading, e. g., volume control during radio range procedures.

4. Research personnel must be continually present to supervise every detail of the experimental administration, e. g., maintaining a detailed log of unavoidable day-to-day variations of procedure such as the effect of adverse weather conditions, and abnormalities of the schedule produced by chemical warfare lectures, military parades, and other interfering factors.

5. Some arrangement should be made whereby the local supervisory and instructor personnel who are doing extra work will be shown as soon as possible some results of a type which are interesting to them. For example, plotting of learning curves on six measures during the daily checks at Pampa was effective in maintaining the interest and cooperation of instructors.

Possible Uses of Such Scales

Pilots responded favorably to these scales; their comments and suggestions were encouraging. They recognized two of the more practical applications of this type of measuring instruments: (1) A more objective and standardized check ride which could be used in place of the one specified in AAF Regulation 50-3 whenever it is necessary to select pilots, individually or in groups, on the basis of instrument flying skill; (2) an objective daily grade slip. This may ultimately be the most valuable application of these objective measures since it is only by repeated tests of the same students that the large day-to-day variations can be balanced out.

These objective measures have potential value also as a means of measuring the results of training experiments and stating the effects of such factors as: different kinds and amounts of instruction in special ground training devices, e. g., Link trainer; transfer of training from one type plane to another; different methods of scheduling (distributing) instrument training. Objective scores also permit a more precise and clear description of the curves of learning and forgetting. Finally, such scales may be used as a criterion for validating parts of the selection and classification battery.

A RECOMMENDED FINAL INSTRUMENT CHECK-RIDE FOR THE TB-25

The many changes in the Training Command following the end of the war made it impossible to continue the statistical analysis of the two objective instrument scales for the TB-25. For this reason, Step 5 of the outline of objective test development, page 190, was never completed. However, a revised form of the final check-ride booklet was made and criticized by expert instrument pilots at three different two-engine schools.

These specialists were asked to include any and all aspects of instrument flying skill which are considered important in a final check ride. As a result, wherever an important aspect of a manouver could not be measured accurately in objective terms, subjective measures were included. However, the total number of strictly subjective measures was rather small, 27, compared to the 92 objective measures in the total scale. The most common subjective item is an evaluation of technique which is included in most of the maneuvers.

A copy of the scale is given in appendix table A8.8. The primary purpose of this revised scale is to serve as a sample of the type of standardized final instrument check ride which might be used for checking graduating students as well as rated pilots being reexamined for the AAF 50-3 "white" instrument card. The maneuvers included in this check were the ones prescribed during the late summer of 1945.

DEVELOPMENT OF INSTRUMENT SCALES AT THE FOUR-ENGINE LEVEL

Four-engine aircraft, particularly the TB-24 and the TB-29, are frequently referred to as "instrument ships" since the pilot sits forward of the wings and does not have easy reference to the position of the wing on the horizon as an aid in controlling the attitude of the plane. Therefore, instruments such as the flight indicator and the rate of climb indicator are habitually used to control the attitude of the aircraft. Since four-engine planes frequently make long-range flights, the pilot must also be prepared to fly by instruments at night and through poor weather conditions. Many experienced pilots hold the opinion that instrument flying skill is the most important single pilot requirement for superior performance when flying four-engine aircraft.

For these reasons, the Pilot Project constructed an instrument scale for use at the four-engine level, but the end of the war prevented conducting an experiment to obtain statistical data on reliability and validity.

The First TB-24 Scale Developed at Fort Worth Army Air Field

This scale combined both contact and instrument measures. This was advisable in order to utilize the flying time of the planes and the students since it was not possible to schedule flights for testing instrument items alone. The first draft contained 118 separate measures; after flight testing with a few students this was reduced to 87 measures.

Fifteen TB-24 Transition student officers were tested with this scale; eight were rated by instructors as being excellent student pilots and seven were judged mediocre or weak TB-24 pilots. No statistical analyses were made of the results of this testing. Scores were distributed, however, for the two groups and the data inspected for any apparent differences or trends. These were taken into account in the later revisions of the four-engine instrument scale. The many hours which personnel from the Pilot Project spent observing and scoring the scale helped them to see ways in which objective measures could be improved.

Four-Engine Instrument Scales from Other Sources

In preparation for the final revision of the TB-24 instrument scale, a survey was made of similar work being done elsewhere at the four-engine Transition level. Two examples of tests of this sort were found which had a reasonable degree of objectivity. These were the "Instrument Flight Check Form" for the TB-24 developed by Lt. Col. L. H. Ware, AAF School of Applied Tactics, Orlando, Fla. The second scale is the "Preliminary Form of Objective Scale of Instrument Flying Skill in the TB-17," devised by the psychological group in the Research Section, Medical Division, Hq., 3rd Air Force. The first of these two scales consisted of 15 maneuvers: 10 basic instrument and 5 radio with a total of 76 measures. A sample item, instrument take-off, is presented to illustrate the form of the first of these two scales.

INSTRUMENT TAKE-OFF

☐	2—Gyro within 2° runway heading
☐	1—Wings level
☐	1—Climb 150 IAS, minimum
☐	1—Gear up, flaps up at proper time
☐	5—Total for maneuver

Scoring is accomplished by checking the appropriate box when performance was satisfactory. This score is indicated beside the boxes. A minimum grade of 60 out of a possible grade of 100 is required for passing this scale. This scoring was based on the judgment of experienced pilots; no reliability or validity data were collected.

The Third Air Force instrument scale includes 42 measures—24 being objective and 18 rated work samples. These items involve the following seven maneuvers:

1. Instrument take-off.
2. Spiral climb.
3. Level flight.
4. 90° and 180° turns.
5. Steep bank.
6. Stalls.
7. Glides.

This scale was administered to 48 experienced four-engine pilots (including 25 instrument instructors and check pilots) and 20 four-engine student officers. The biserial correlation of the total score against these two groups differing in experience was 0.20. No reliability data were reported in this preliminary study.

Items for the Final TB-24 Instrument Scale[48]

Measures previously developed by the Pilot Project and other groups were tailored specifically for a final TB-24 instrument check ride. On the first page of the booklet, the student was given instructions as to how he could make his best score by defining the passing and failing limits for altitude, heading, and air-speed deviations.

Appendix tables A8.9 and A8.10 present an inventory of the maneuvers included in the scale and two sample complete maneuvers.

A B-29 Scale

The importance of instrument skill in flying the TB-29 suggested that research be initiated in the instrument phase, for even under contact conditions the B-29 pilot flies with frequent reference to his instruments.

At the time the war ended only a few measures specifically designed for the TB-29 had been constructed. However, it was expected that most of the TB-24 measures could be easily modified to fit this larger plane.

[48] Lt. H. H. Hagy assisted in the preparation of this scale.

OTHER RESEARCH ON INSTRUMENT FLYING SKILL

Exploration of Specially Designed Apparatus

The Psychological Research Project (Pilot) purposefully limited its work in objective scale development to the use of the instruments, equipment, and facilities normally found in training aircraft. However, some preliminary work was done exploring the possible use of two specially designed devices which might be used in evaluating instrument-flying proficiency.

Use of the work-adder in measuring flying technique.—A consistent and frequently repeated criticism by pilots of many objective measures of flying skill refers to the omission of adequate measures of technique. Most objective measures were taken from the end product of pilot performance with no evaluation of how these results were obtained. Flying instructors generally emphasized technique and judgment equally with, if not more than, precision limits of the plane in flight. In the absence of any adequate objective measure of technique the assumption was made that when testing groups of students, those with superior technique could maintain tighter limits than students with poor technique. Such an assumption is less tenable when evaluating the individual student.

One common error ascribed to poor technique is excessive use of controls, e. g., rudder, aileron, elevator, etc. Although the student achieves the desired control of heading, altitude, air speed, etc., he has to work too hard, wasting energy and gasoline by overcontrolling. The work-adder principle was applied directly to this problem yielding an objective record of the number and extent of control movements made. A more complicated device could probably be devised to measure pressures. In cooperation with Dr. Roger B. Loucks, Department of Psychology, AAF School of Aviation Medicine, an experimental model of a work-adder counter was developed and given preliminary try-outs in actual flight (TB-25). The attachment was made to the rudder cable in a manner that in no way affected the control of the plane in flight. Such a device could easily be attached to the aileron and elevator cables and to the throttle. The professional judgment of expert pilots was that this apparatus would be useful, particularly in evaluating instrument and formation flying.

For the purpose of exploring its use as a possible measure of instrument-flying proficiency, one model of the work-adder counter was installed in one of the instrument planes at Pampa Army Air Field to be used at the time of the final instrument check rides reported earlier in this chapter. A rudder action count was made for seven students on the basic maneuvers included on the final check-ride booklet. Table 8.12 gives the distribution of rudder-action counts for the 7 students tested and for the 10 basic instrument maneuvers.

TABLE 8.12.—*Average rudder action counts made by seven students on ten basic instrument maneuvers*

Maneuver	Average rudder count	Maneuver	Average rudder count
1. Instrument take-off	17.3	6. Steep turn	4.0
2. Spiral climb	2.9	7. Single-engine procedure	14.8
3. Level-out, partial panel	1.8	8. Power off stall	4.9
4. 180° full panel turn	1.7	9. Diving spiral	2.0
5. Partial panel 180° turn	5.9	10. Spiral climb, partial panel	7.0

Average accumulated rudder counts for the seven students: 1.7, 4.9, 5.2, 5.8, 7.3, 7.5, 10.6.

The results in table 8.12 reflect one of the flight characteristics of the TB–25 which is not a "rudder plane," i. e., coordinated turns can be made with very little use of the rudder. It is interesting to note that more than twice as much rudder was used when making turns or climbing without the use of the directional gyro and the flight indicator than when the full instrument panel was available to the student. This shows that the work-adder is sufficiently sensitive to discriminate at least one aspect of technique in maneuvers of varying difficulty.

A more crucial test of this apparatus could be made by using it during formation flying since amount of rudder "work" is one of the chief distinguishing characteristics of good and poor formation flying skill.

The apparatus worked very well. It was of simple construction, could be quickly installed, easily read by the check pilot, and provided an accurate count of any rudder movement as great as one-half inch.

Radio beam tracker.—A radio beam tracker was developed by Dr. Loucks, Department of Psychology, AAF School of Aviation Medicine, which recorded the flight path of an aircraft flying the radio range. This is an adaptation of the Link "crab" but was a complicated piece of apparatus and as such would not be suitable for large-scale use in the immediate future. However, the problem which it was designed to answer was crucial and its final development would be a distinct asset in objectively grading pilots' radio range flying proficiency.

SUMMARY

Instrument flying skill was of great importance in military aviation. Throughout the course of the war increasing emphasis was placed on instrument training including a greater stress on objective limits as a part of the final instrument check (AAF 50–3 "white card") required of all graduating pilots. Because of its importance and its adaptability to objective measurement, the Pilot Project placed a high priority on research in developing objective measures of instrument flying skill.

The completed studies in the instrument area can be summarized in four stages: (1) A "shotgun" administration of both contact and instrument measures to large groups of students to determine the best area of concentration; (2) a methodological analysis to determine the most effective way of scoring instrument measures; (3) actual construction of comprehensive objective check rides and administration to students for purposes of statistical analysis; and (4) construction of a large inventory of objective instrument measures for use in several different stages of training. The work was initiated in the two-engine area with the TB–25.

The first experimental use of instrument measures was made at the end of an extra 5 weeks' training period given to all upper-class students in the Advanced two-engine schools (see chapter 10). The results showed a significant improvement in precision instrument flying with the additional training. The results also indicated that much greater proficiency was attained when the additional training was given in the same type plane than when students were transferred from one type to another. It was concluded that pilots must be thoroughly familiar with the flying characteristics of the TB–25 before precision instrument flying is possible in this plane.

Three different methods of scoring instrument measures were compared. These were the range, the limits, and the time sample method. Three sample measures (altitude in instrument steep turn and heading and altitude in instrument straight-and-level) were scored by each of the three methods. Reasonably satisfactory results were secured by all three methods. For the three sample measures scored in the three different ways, the observer reliabilities ranged from 0.54 to 0.90, the reliabilities of test and retest in immediate succession from 0.37 to 0.82, and the correlations with instructor's ratings from 0.15 to 0.33. Although the time sample method had the highest observer reliabilities, its test-retest reliabilities and validities were no better than those of the range method. The range appeared to be somewhat better than the limits method. In general there were no great differences among the three methods.

A 34-item objective check ride was developed and administered on 6 instructor-student dual instrument flights. The odd versus even day reliabilities for the measures showed that there was a great deal of day-to-day variability. The coefficients ranged from −0.21 to 0.64, with a median of 0.25. For the 22 most reliable measures, the differences between scores after 2 and after 6 dual instrument flights were computed. Thirteen of the twenty-two measures yielded differences which were significant at the 1 percent level or above. This indicates that they were valid for the purpose of discriminating between students with different amounts of training.

A second and longer check ride of 70 items was constructed for use as a final check at the end of the instrument training period. It was found that the students who were given the daily check made somewhat better objective scores at the end of instrument training than those who were not given this test on each dual ride. After a 4-week interval of further contact and transition flying, the long instrument check was given again. No consistent change occurred in the level of instrument-flying proficiency.

A recommended final instrument check ride is presented as a sample of the type of standardized flight check which could be used to evaluate instrument-flying skill. Preliminary work was done at the four-engine level of training and sample objective instrument measures are briefly described.

The use of objective measures at the Basic level in the single-engine AT-6 is presented in the following chapter.

CHAPTER NINE

Objective Measures of Single-Engine Instrument Flying Skill at the Basic Level of Flying Training

Capt. William V. Hagin [*]

INTRODUCTION

As was explained in the previous chapter, most of the work of the Pilot Project on developing objective measures of flying skill beyond the Primary phase was concerned with instrument flying skill, since instrument flying was so important for military operations and was also especially adapted to objective measurement.

In the basic phase of flying training, the students learned how to control the airplane using only cues from the instrument panel as a guide to the flight attitude of the aircraft. A cloth hood was put over the cockpit to simulate conditions under which the pilot could not see the horizon or other natural points of reference. After the student had learned these fundamental elements of instrument flying skill at basic school, he graduated to advanced where he learned to use the additional cues supplied by radio and to apply the various elements of instrument skill to practical problems such as flying cross country and making instrument let-downs. The nature of instrument flying is described in detail in Chapter 3, An Analysis of the Pilot's Task.

It was important to develop a scale for measuring objectively instrument flying skill in the AT-6 (North American "Texan") since all students, both the prospective two-engine and single-engine pilots, learned the same elements of instrument flying skill in that aircraft during their Basic training.

PURPOSE AND PROCEDURES

Purpose

This study was initiated to develop and evaluate an objective scale for measuring instrument flying proficiency in the AT-6 at the Basic level of training. The specific sub-goals were:

(1) to construct a comprehensive battery of items for objectively measuring instrument flying skill;

[*] S/Sgt. Walter W. Ismael assisted in the construction and administration of this scale. The statistical analysis was planned with the assistance of S/Sgt. Ismael, T/Sgt. William O. Matheny, and Sgt. John O. Gleason.

(2) to evaluate each of these items by (a) exploratory try-out and expert criticism, (b) determining the test-retest reliability of each item, and (c) determining the validity of each item as indicated by its correlation with the instructor's over-all rating of instrument flying skill;

(3) to compare the reliability and validity of various methods of combining the scores of individual items into over-all scores for the maneuver involved;

(4) to compare the reliabilities of items and of maneuver scores when test and retest occurred within the same ride with those when test and retest were given on different days; and

(5) to evaluate the total score obtained from all of the items in the scale by determining (a) the split-half reliability of that score, (b) the reliability of test and retest on different days, and (c) the validity of the total score as indicated by its correlation with instructor's rating.

Developmental Procedures

First, a preliminary 85-item scale was constructed and administered to 29 students in Basic flying training. After testing, the instructors and students were interviewed with reference to the following points:

(1) Adequacy of the measures;
(2) Suggested revisions and additions to the scale; and
(3) Elimination of items which might contribute to an accident.

This pretest was an extremely important step in developing flying scales since many test items that appeared safe, sensible, and adequate on the ground were not suited to the actual flying situation. Though the groups tested were not large enough for extensive statistical analysis, the information gained was of value in constructing the final form of the scale and helped to establish rapport with school personnel through participation of instructors and supervisors in the construction of the scale they were to work with in the more extensive study.

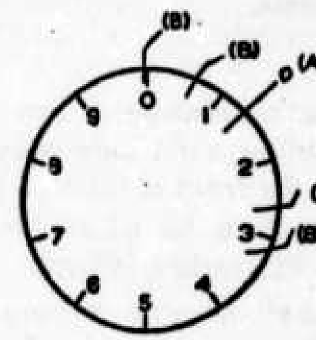
Description of the Check-Ride Booklet

The final form of the test booklet is reproduced in appendix A.9 (pages 433 to 441). It was printed on lightweight card stock. The entire first page was devoted to biographical information, general instructions to the students and check pilots, and tables showing the weights for deviations in heading, air speed, altitude, and time. These weights were derived from frequency distributions of the raw scores obtained on the preliminary administration. The class intervals were chosen to achieve good distributions and were weighted so as to approximate standard scores for deviations in heading, altitude, air speed, etc. Thus a score of (5) on headings was roughly equivalent to a score of (5) for any of the other variables. Point scores for the "Yes or No" type of item used for the instrument take-off were included with the item. The four Unusual Position Recovery measures were graded subjectively on a four-point multiple-choice scale printed opposite each of those maneuvers.

In all there were 29 maneuvers in the scale and a total of 81 items. The time required for administration was slightly over one hour.

Specific instructions to the check pilot were included with each maneuver; instructions to be given to the student were standardized by having the check pilot read them from the booklet.

The check pilot recorded the students' deviations on the score card dials using the markings shown in the sample below:



The starting point for each maneuver was marked with (A) and the deviations were marked with the curved lines (B). The deviations were marked continuously; the check pilot made new marks each time the students exceeded previous errors. Thus by subtracting the lowest figure reached from the highest, the student's range during the maneuver was easily obtained.

Administration of the Test

The final form was administered to class 45-G at Moore Army Air Field Basic Flying School during the period 30 August to 2 September 1945. All students had finished their Basic training in the AT-6 except for their final instrument check.

Sixty students were given two rides on successive days by different check pilots. Forty additional students were tested on one ride only during the second day of testing. Local conditions were such that it was impossible to schedule this group for both rides. The first administration of the test was made during the afternoon flying period, and the second on the following morning. No student was checked by his regular instructor.

Prior to actual test administration, each instructor rated his students' instrument flying ability on a four-group scale, using the best 25 percent of all the students he ever had, the second 25 percent, the third, and the poorest 25 percent as standards of comparison. This type of rating was used to force the instructors to make their evaluations thoughtfully and accurately. However, since only 11 of the 95 students tested were rated in the upper group and only 15 in the lower, it was necessary later to combine the two upper and the two lower groups into a dichotomy.

Each student was allowed to study the booklets before the flight. He was shown the weights of the items and was told how to make the best score. All students and instructors were briefed on the items and a general discussion of the scale was held in order to insure that they were thoroughly familiar with how to perform the maneuvers and how to score the booklets.

Scoring of Test

After the test had been administered, raw test scores (range of deviations in altitude, heading, etc.) were converted to coded scores using the table printed on the front of the test booklet. These coded scores served as the basic data for all analyses. A sample of the scoring key for deviations in heading follows:

Deviation	0° to 5°	6° to 10°	11° to 15°	16° to 20°	Over 20°
Coded score	5	4	3	2	1

It can be seen that if the range of heading deviations was 12°, the student would get a coded score of 3.

Statistical Analysis

The following statistical analyses of the data were made:

1. The reliability was determined for each item when test and retest were on different days by different check pilots.

2. The validity of each item was obtained by correlating item scores with the instructors' ratings of student performance.

3. The reliabilities and validities of three different methods of combining scores from individual items into over-all maneuver scores were compared. Three representative maneuvers were selected for this analysis.

4. The reliabilities of representative items and maneuvers when test and retest occurred within the same ride were compared with those obtained when test and retest occurred on different rides.

5. The total score of the scale, based on the 5 subjective and 76 objective items, was evaluated by determining:

   a. The split-half reliability of the total score on each administration.

   b. The reliability of test and retest on different days.

   c. The validity of the total score as indicated by its correlation with instructor's rating.

RESULTS

Item Analysis

The reliabilities and validities of the items are summarized in table 9.1. The day-to-day reliabilities were obtained by correlating scores from rides given on two different days by two different check pilots. The validity coefficients presented were correlations of item scores from the second day's ride with subjective ratings by the students' regular instructors. The second day's booklets were selected for validity analysis because 100 students were tested on that day while only 60 were tested on the first day.

TABLE 9.1.—*Summary of item reliabilities and validities for AT-6 instrument scale*

Kind of measures	Number in scale	Day-to-day test-retest reliability [1]				Validity [2]			
		Beyond 1 percent level		5 percent level to 1 percent level		Beyond 1 percent level		5 percent level to 1 percent level	
		n	%	n	%	n	%	n	%
Altitude	21	4	19	3	14	4	19	6	29
Air speed	25	3	12	7	28	1	4	9	36
Heading	19	0	0	3	16	2	11	3	16
Angle of bank	11	0	0	1	9	0	0	2	18
Subjective ratings	5	0	0	0	0	0	0	1	20
Total	81	7	9	14	17	7	9	21	26

[1] Based on 60 students tested on different days, in different planes, by different check pilots.
[2] Correlations of performance with ratings which were made by the student's regular instructor under controlled conditions which assured complete independence of validity criteria and check-ride measures. Ninety-five students were used in validity analysis.

The reliabilities ranged from −0.32 to 0.40 and the validities from −0.24 to 0.41. However, 21 of the 81 measures, or 26 percent of them were reliable at the 5 percent level of significance or better; 28, or 35 percent, were valid at the same level of significance. Altitude and airspeed measures were the most reliable. Although only 9 of the 81 items were both significantly reliable and valid, the correlation between the reliabilities and validities of all test items was 0.34 (significant at the 1 percent level or above) indicating that the more reliable items also tended to be somewhat more valid.

Five of the measures in the scale were subjective ratings by the check pilot. None of these items yielded statistically significant reliabilities. Only one, "Instructor Help on Take-off," was significantly valid at the 5 percent level or above. The reliability and validity, respectively, of each measure is presented in tables A9.1 and A9.2 of the appendix to this chapter.

Comparison of Three Methods of Combining Items into Maneuver Scores

Each item was a single measure (Heading, Altitude, etc.) taken during the performance of a maneuver such as climb, turn, or level

Table 9.4.—*Comparison of reliabilities when test and retest occur on the same check ride with those when test and retest occur on different days*

Altitude Measures in Turns and Legs of Pattern A

	N	r_{12}	M_1	M_2	SD_1	SD_2
First leg versus second leg:						
Same ride	57	0.43	2.96	2.75	0.96	1.17
Different days	58	.36	2.97	4.28	.96	.83
Second leg versus third leg:						
Same ride	57	.63	2.75	2.47	1.17	1.33
Different days	57	.05	2.75	4.14	1.17	.80
Third leg versus fourth leg:						
Same ride	57	.67	2.47	3.44	1.33	1.36
Different days	58	.20	2.47	4.07	1.32	.94
First turn versus second turn:						
Same ride	56	.62	3.18	2.30	1.13	1.32
Different days	57	.29	2.47	2.98	1.13	1.00
Second turn versus third turn:						
Same ride	56	.35	3.30	3.38	1.32	1.27
Different days	56	.18	3.30	4.14	1.32	.88
Third turn versus fourth turn:						
Same ride	57	.65	2.37	3.35	1.36	1.32
Different days	58	.26	3.35	4.16	1.36	.91
Mean (z combination):						
Same ride	57	.57				
Different days	58	.23				

Maneuver Scores in Turns of Pattern A

Scoring Method	N	r_{12}	M_1	M_2	SD_1	SD_2
Beta weight:						
Same ride	49	0.62	39.78	37.47	12.08	14.89
Different days	50	.24	40.58	46.54	12.13	10.80
Equal weight:						
Same ride	49	.64	10.37	10.08	2.91	2.48
Different days	50	.18	10.50	11.84	2.78	2.07
"Expert pilot" weight:						
Same ride	49	.68	5.92	5.50	2.09	2.57
Different days	50	.23	6.02	6.94	2.00	1.65

Evaluation of the Total Score

The coded scores on all of the items in the scale were summed to give three scores for analysis: odd half, even half, and total score. The scale was divided so that the two halves were equated with the same number of similar measures in each half. This was done for both administrations of the test. Thus it was possible to obtain split-half reliability for both rides as well as test-retest reliability of the total scale when the rides were administered on different days by different check pilots.

The reliability data in table 9.5 show that the internal consistency (split-half reliability) of the scale was 0.89 which is high and compares favorably with that of written tests. The test-retest reliability on different days was 0.46 and significantly higher than that of any of the items.

Similarly, the validity coefficients in table 9.5 were all significant at better than the 1-percent level and were larger than those from any single item in the scale. The correlation of the combination of the two administrations with instructor's ratings was 0.51. This is high in view of the probable unreliability of the latter criterion.

TABLE 9.5.—*Reliabilities and validities of total score derived from 81 measures in AT-6 scale*

[Based on 55 students from class 45-O]

A. Reliabilities

	r'_{II}	r_{II}	M_1	M_2	SD_1	SD_2
1. First half[1] *versus* second half[2] first administration	0.83	0.91	151.45	145.29	19.13	21.02
2. First half[1] *versus* second half[2] second administration	.80	.89	165.38	156.58	13.30	17.17
3. Total score[3], test *versus* retest	....	.46	290.75	321.96	38.40	28.95

B. Validities

	r_{bis}	P_s	M_s	M_f	SD_t
4. Total score first administration	0.44	0.49	311.33	292.08	38.40
5. Total score second administration	.40	.49	331.41	312.86	28.96
6. Total score first administration+second administration	.51	.49	642.76	505.84	57.86

r'_{II} Correlation between two halves.
r_{II} Correlation corrected to estimate reliability of total check ride for 1 and 2, and test-retest reliability of total check ride for 3.

[1] First half is total of scores 3, 4, 6, 7, 8, 9, 14, 16, 17, 18, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, O-1, O-2, 71, 75, 77, and 78.
[2] Second half is total of scores 5, 10, 11, 12, 13, 15, 19, 20, 21, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, O-3, O-4, 72, 73, 76, 79, and 80.
[3] Total score is total of first plus second halves as defined above.

SUMMARY AND CONCLUSIONS

Instrument flying was a most important aspect of flying skill during the war years. Every pilot had to be competent at instrument flying since combat operations were conducted despite bad weather. The importance of good instrument pilots and the adaptability of instrument flying to objective measures made it desirable to concentrate research in this area.

An 81-item scale for measuring instrument flying proficiency at the Basic level of training was constructed and administered at Moore Army Air Forces Basic Flying School on 2 successive days to 60 students and to an additional group of 40 students on the second day only.

Although the reliabilities of the test items ranged from −0.32 to 0.40 and the validities ranged from −0.24 to 0.41, 26 percent of the reliabilities and 35 percent of the validities were statistically significant at the 5 percent level or better.

Of the five subjective measures in the scale, none had a statistically significant reliability and only one had a significant validity.

A comparison was made of three different methods of combining the scores on individual items into a score for the maneuver involved. The methods were beta weights, equal weights, and weights based on professional judgment. This third method approximated cut-off scores. No great differences were found in the reliabilities or validities of these three methods. In general, the maneuver scores were somewhat more reliable and valid than the average of the items involved but poorer than the best item.

For both items and maneuver scores the correlation between test and retest within the same check ride was considerably higher than that between two tests on different days.

All items, both the 5 subjective and the 76 objective, were combined (with equal weights) into a total score. The odd-even reliability (corrected for double length) of a single check ride was 0.89. The correlation between check rides on different days was 0.46. From this it can be estimated that the test-retest reliability of the sum of the scores on two check rides would be 0.63. The correlation between the sum of the scores on the two rides and an over-all rating of instrument flying proficiency made immediately before these rides by the student's instructor was 0.51. These figures indicate that it is feasible to develop a reliable and valid objective scale of instrument flying skill.

CHAPTER TEN

A Large-Scale Objective Study of the Effects of Additional Training on Flying Skill[a]

Maj. Neal E. Miller, 1st. Lt. William E. Galt and S/Sgt. Charles P. Gershenson

PURPOSE AND PROCEDURES

The Training Freeze

In the fall of 1944 pools of surplus pilots had accumulated in the Training Command because combat needs were not as great as had been anticipated. In order to eliminate these pools, Headquarters, Army Air Forces, ordered training freezes.

Normally 10 weeks were spent at each of the three levels of training: Primary, basic, and advanced. At any one time there were two classes in the school at each level. The lower class received its first 5 weeks of training while the upper class received its last 5 weeks. At the end of this period, the upper class was graduated to the next level of training and the lower class was promoted to upper class.

The first 5-week training freeze began in the middle of October 1944. Instead of being graduated to the next level of training, the upper-class students in Primary, Basic, and Advanced schools were held over for an additional 5 weeks. During this time the students

[a] The entire staff of the Pilot Project was cooperatively involved in this study, with special assistance from the Psychological Section, Office of the Surgeon, Hq. AAF Training Command. Maj. Neal E. Miller was responsible for over-all direction and for conducting conferences to arrange the details of participation with schools in Central Flying Training Command. Capt. Richard P. Youtz was responsible for the development and pretesting of measures for the scale used in Primary and Advanced single-engine schools and for conferences to arrange for participation of single-engine schools in Western Flying Training Command. Maj. Anthony C. Tucker, assigned to temporary duty with the Pilot Project, was responsible for developing and pretesting measures for the scale used in Basic schools and for conferences to arrange for participation of single-engine schools in Eastern Flying Training Command. Capt. Stanford C. Ericksen was responsible for developing and pretesting measures for the scale used in Advanced two-engine schools and for the conference to arrange for the participation of all such schools. Lt. William E. Galt checked on the instrument measures with the standardization boards at the Instructors School (Instrument Pilot) at Bryan, Tex., and visited schools in Eastern and Central Flying Training Commands to check up on conditions of testing. T/Sgt. W. O. Matheny, S/Sgt. W. W. Ismael, S/Sgt. I. Robbins, Sgt. J. R. Rohrs, and Sgt. M. F. Connery supervised testing at different flying schools. Maj. Rains Wallace was responsible for coordinating directives in the Training Command Headquarters, arranging for the printing of scales and directions, and receiving the rosters from the schools. Lt. Col. Walter L. Deemer served as statistical consultant and supervised the transferral of the data to punch cards in the Statistical Unit of the Psychological Section. S/Sgt. Charles P. Gershenson supervised the analysis of the data on tabulating machines. He was assisted by Sgt. J. W. Stratton, J. J. Waeltermann and J. R. Rohrs. Prof. Phillip J. Rulon, Acting Dean of the Graduate School of Education, Harvard University, advised on the measures and on statistical design as a Special Consultant to the Secretary of War.

who had been in the upper class received an additional training [33] and those who had been in the lower class went on to receive the normal upper-class training. Thus, at the end of the 5-week freeze, the students who had been in the lower class of Primary, Basic, or Advanced had completed the normal 10 weeks of training and those who had been in the upper class had received an additional 5 weeks, making a total of 15.

After this first training freeze, the lower class which had already received the normal upper-class training, was trained for an additional 5 weeks and the next class began its initial 5 weeks of training. This second freeze put the two classes back into phase.

Purpose of the Study

Approximately 3 weeks before the beginning of the first training freeze, the Chief of Staff of the Training Command asked the Pilot Project to use this sudden opportunity to secure objective measures of the effects of the additional training. Specifically the purposes of this study were:

1. To measure objectively the effects on flying proficiency of the additional five weeks of training at each of the levels: Primary, Basic, and Advanced.[34]
2. To determine the effects of having received civilian flying training before starting military flying training.
3. To secure normative data on the level of skill of the pilots being trained.
4. To investigate the feasibility of developing and using objective measures of flying skill in large-scale studies.

Method Used in the Study

Decision to test in many schools.—It was first necessary to decide whether to plan an intensive or extensive study. The intensive approach would have been to concentrate all of the work in a few sample schools, supervising these very carefully and using a specially trained cadre of check riders analogous to the Primary Training Advisory Board. With the limited time available and the small staff of the Pilot Research Project, it would have been possible to conduct intensive experiments in only one school of each type: Primary, basic, advanced single-engine and advanced two-engine.

[33] In primary 25, and in all other schools 30 hours of additional flying training.

[34] The Chief of Staff of the Training Command was also interested in determining whether the introduction of the additional 5 weeks of training at one level would produce more effect upon the final product measured at graduation from advanced training than the introduction of 5 weeks of additional training at other levels. Although it was believed that changes in training conditions might make such a comparison impossible, data were collected on each of the classes as they graduated from advanced training. Since it was anticipated that such data would subsequently be collected, only half of the students were tested during the freeze at each of the levels of training so that the other half would not have had any previous experience with such measures when tested in advanced. Random selection of the students to be tested was insured by specifying that they be selected by army serial number. Only those students were tested whose serial numbers had the last two digits in the range 00 to 49. As had been expected, training conditions had changed so much by the time that the last students involved in the freeze (those tested in primary) graduated from advanced, that meaningful longitudinal comparisons were not possible.

Thus, under this plan, any unusual circumstance, such as a continued period of bad winter weather preventing flying at any one of the selected schools during the week allotted to testing, would have spoiled the experiment.

Since it seemed desirable to include more schools than could be tested by a specially trained cadre of check riders personally supervised by the Pilot Project, procedures had to be designed so that they could be administered by the regular staff of any school without special training. Once the procedures had been designed in this way, it was practically as easy to set up testing in all the schools within the Training Command as it would have been to concentrate on a medium-sized sample. Therefore, the most extensive possible sample was used in order to be sure to secure enough cases to take care of unforeseen emergencies such as weather, and to allow for splitting the data up into as many subgroups as might be required for thorough analysis.

Procedure used in developing measures.—At the time the training freeze study was assigned, the Pilot Project was in the early stages of developing objective measures of flying skill. Some measures had been devised and tried out on relatively small groups of students in the primary and advanced two-engine phases of training. No work had been done in basic or advanced single-engine schools. It was therefore necessary to devote considerable effort to the development of additional measures.

The measures for the different types of training (primary, basic, advanced single-engine, and advanced two-engine) were developed in collaboration with the Training Advisory Boards for those types at the Central Instructors School (Pilot). The Instrument Board at Central Instructors School (Pilot) assisted in the development of instrument measures. The Deputy for Training and Operations instructed the chairman of each board to assign one member whose primary duty was to cooperate with the Pilot Project in the training freeze studies. The other members of boards also assisted.

Since the objective measures were to be administered on short notice to large numbers of students in all of the schools in the Training Command, each individual item had to be simple enough so that it could be administered by the regular instructional staff without special training or complicated explanations, and the scale as a whole had to be short enough so that it could be fitted into the regular schedule and administered within a reasonable period of time.

Since the purpose of this study was to discriminate between the performance of groups (those with 10 and 15 weeks of training), it was possible to use a somewhat shorter check ride, containing a less comprehensive sample of objective items than would have been necessary if it had been desired to discriminate between individuals. A given individual might score high or low by chance; all of the mem-

703322—47——16

bers of a large group of students, however, would not get high or low scores by chance.

The first step in constructing the items was to study the curriculum prescribed for training during the additional 5 weeks and to determine what aspects of this curriculum could practically be measured under the conditions of this experiment. Items were then constructed, criticized, and flight-tested in collaboration with members of the Training Advisory Boards. Then those items which seemed most promising were tried out on groups of from 20 to 30 students in schools at each level in order to determine the feasibility of the item, the proper ranges to be selected for scoring, and to secure criticisms of the item from supervisory personnel at the schools. The schools selected for pretesting at each level were usually at a Wing Headquarters so that the criticism of Wing personnel could also be secured. Furthermore, instrument items were taken to the Instructors School (Instrument Pilot) at Bryan, Tex., and criticized by experts there.

Some items were found unsuitable when pretested and were eliminated. The best items surviving pretesting were assembled into check-ride scales for each level of flying training. These scales required approximately 45 minutes of administration time. They were printed on 9 by 12 inch card stock which could be folded to make check-ride booklets stiff enough to afford good backing for writing in the air. In all cases the scoring of items was by circling a number: 1, 2, or 3.[16]

The items used on the check-ride cards for each level of training are presented on pages 445 through 460 of the appendix. It shoulh be remembered that they were not designed to discriminate between individuals, but only to sample differences between groups.

General procedure followed in administering study.—First, a directive (letter, 702.5, Hq. AAF Training Command, 7 October 1944, subject: Objective Measures of Flying Skill During Temporary Training Freezes), was indorsed to the schools by each of the Flying Training Commands to inform them of the general nature of the study and the necessity for their maximum cooperation. Then, two sets of directions were made up.

One set of directions was for the administrative personnel at each school. It described the general plan of the study, when the check ride should be given, who should be tested, who should give the check ride, what field preparations should be made, and the procedures for briefing the check riders and the students. Specific directions were also given concerning how the score card should be filled out, what daily log should be kept of any special conditions that might affect

[16] Since different types of training planes had been used at different times in different schools in the Training Command, the check ride for each student contained a place for recording the types of planes he had been trained on in previous phases of training. The number of times which a student had been held over after graduating from preflight, his stanine, and whether or not he had soloed before entering primary school were also recorded.

the results, and how the school secretary should transcribe the scores to the printed rosters which were provided.

The second set of directions informed students and instructors as to the general purpose of the study and the way each item should be performed and scored. Arrangements were made so that a copy of this sheet of directions and a check-ride card would be distributed to each student before the check ride. This was done so that each student would know exactly what he was expected to do and how he would be scored on it.

Every effort was made to make the two sets of instructions so comprehensive and clear that the whole procedure could, if necessary, be run off on the basis of these instructions alone. In addition, special conferences were held at the headquarters of each of the three Flying Training Commands. These conferences were attended by three representatives from each school: the Commanding Officer or the Director of Training, the Director of Flying, and a member of the Standardization Board. Conferences for each type of school were held on separate days. At these conferences a representative of this Project explained the whole procedure in detail and discussed with representatives of the schools special problems which might arise. For two-engine schools, one conference was held at Randolph Field instead of a separate one in each of the Flying Training Commands.

Finally, an effort was made to visit as many schools as possible during the course of the testing to see how it actually was accomplished. Those schools that could not be visited during the course of testing were visited shortly thereafter. Key personnel were interviewed to determine how the testing was conducted, what special conditions might have influenced the scores, and what criticisms and suggestions the schools had to offer for improving and adding to the objective measures.

Specific conditions for administering check rides.—The procedures which have just been described (written directions, conferences, and visits to schools) were all designed to insure careful administration of the objective items under standard conditions. The most important conditions of testing were as follows:

1. All male, white students within the specified ranges of army serial numbers [64] were tested with two exceptions: student officers and students from foreign air forces.

2. All check rides were flown during the week 13–18 November 1944. This was the last week of 10 weeks of normal training for one class, which was called the 10-week group; it was the last of 15 weeks of training (10 normal plus the 5 additional weeks during the freeze) for the other class, which was called the 15-week group.

3. Equal percentages of the 10- and the 15-week groups were tested

[64] See footnotes 54 and 67.

during any given half day in order to equalize, as far as possible, weather and other conditions of testing.

4. No student was tested by a check pilot from his own squadron. The best available check pilots were selected by the supervisory personnel at each of the schools. These check pilots flew, as nearly as possible, equal percentages of students from the two groups.

5. The maneuvers were performed under as good conditions of smooth air, visibility, etc., as possible, but climbs to excessively high altitudes to obtain smooth air were avoided.

6. The check pilots and administrative staff were told that the rides were for research purposes only, that no student would be eliminated from training on the basis of his performance, and that the results of all schools would be combined in the final report so that there would be no possibility for invidious comparisons.

Handling of data.—Each school transcribed to rosters the single-digit numbers which had been encircled on the check-ride cards. These rosters were spot-checked and all of the data transferred to punchcards at the Psychological Section of the Training Command Headquarters. Then enlisted personnel from the Pilot Project analyzed these data on tabulators and sorters. First, distributions of all the cards were tabulated in order to get a general picture of the data. On the basis of these distributions it was possible to decide which groups should be broken down for final analysis into categories homogeneous enough for a meaningful interpretation and yet large enough for statistical stability.

Because the training freeze happened to coincide with a change from lighter planes to the TB-25 in the advanced two-engine schools, the conditions for that part of the study were different from those in the single-engine schools. The further details of analysis and the results will therefore be presented separately for these two types of schools.

EFFECT OF ADDITIONAL TRAINING ON SINGLE-ENGINE FLYING PROFICIENCY

Analysis of Data

Students included in the final analysis.—In order to avoid artifacts and to make each group more homogeneous, certain students whose selection or training was atypical were excluded from the final analysis.

All students who had been held over from one class to the next, for sickness or other reasons, at any time after preflight were eliminated from the study. Students were excluded if they had not had all of their training for the level at which they were tested in the type of plane most commonly used: i. e., Stearman for Primary, BT-13 or BT-15 for Basic, and AT-6 for Advanced single-engine. Since

Central Flying Training Command was in the process of changing from the Fairchild to Stearman at the Primary level, this meant excluding students from all of the schools in that command from the analysis of the Primary level of training.[*]

Students were also excluded if they were part of specially selected groups. For the analysis of the advanced single-engine level of training, the students in Western Flying Training Command were excluded since the upper class was specially selected by having poorer students sent to another type of training instead of being retained for extra training during the freeze. Some of the Basic students trained on BT planes had to be excluded because they were the poorer students who were left over after the better ones in their school had been selected for Basic training on the AT-6.

After the students whose military flying training or selection was atypical had been sorted out, the sample of single-engine students was reduced from 6,438 to 4,757 cases.

The groups at each level of training, however, were still not homogeneous. Some of the students had received enough civilian flying training to solo before they entered Primary and commenced their military flying training. Most of the students had not soloed before entering Primary. It was therefore necessary to divide the sample into two groups: those who had soloed and those who had not soloed before entering Primary. Since the students who had not soloed before entering Primary had received the most uniform and representative training they were the sample from which the most meaningful conclusions concerning the effects of the five weeks of additional training during the freeze could be drawn. Limiting conclusions concerning the freeze to these students reduced the sample to 3,833 single-engine students.

Groups compared for each level of single-engine training.—After the sample had been refined by discarding atypical students, it was divided into the following groups in order to determine the effects of the five weeks of additional training given during the freeze and the effects of having had enough civilian flying training to have soloed before entering Primary.

1. *10-week, nonsolo group.*—Students who were tested during the last week of 10 weeks of normal flying training and who had not soloed prior to Primary training.

2. *15-week, nonsolo group.*—Students who were tested during the last week of 15 weeks of flying training (made up of the regular 10

[*] Since the freeze also coincided with the closing of approximately half of the Primary schools in the three Flying Training Commands, the upper class from the schools which had just been closed was transferred to those still open. Therefore, only students with army serial numbers ending in the range 00-24 were tested; those in the range 25-49 were saved for a later control test on the effects of overcrowding. This test was not made since visits to the schools indicated that they had been able to cope with the double load. See footnote 54 for an explanation of why students with army serial numbers ending in the range 50-99 were saved for possible further study.

plus 5 additional weeks during the freeze) and who had not soloed prior to Primary training.

3. *10-week, solo group.*—Students who were tested during the last week of 10 weeks of normal flying training and who had soloed prior to Primary training.

4. *15-week, solo group.*—Students who were tested during the last week of 15 weeks of flying training (made up of the regular 10 plus 5 additional weeks during the training freeze) and who had soloed prior to Primary training.

The number and type of students in these groups at each level of training is presented in table 10.1.

TABLE 10.1.—*Groups used in the final comparisons at each level of single-engine training*

Group number	1	2	3	4
Level of training, plane, and command	10-week nonsolo	15-week nonsolo	10-week solo	15-week solo
Primary students trained on Stearman plane in eastern and western flying training commands	211 Class 45-C	954 Class 45-B	60 Class 45-C	228 Class 45-B
Basic students trained on BT-13 or 15 plane in eastern, central, and western flying training commands	509 Class 45-A	710 Class 44-K	234 Class 45-A	221 Class 44-K
Advanced students trained on AT-6 plane in eastern and central flying training commands	428 Class 44-J	931 Class 44-I	62 Class 44-J	119 Class 44-I

10-week: Students tested during the last week of 10 weeks of normal training in primary, basic, or advanced.
15-week: Students tested in primary, basic, or advanced during the last week of training made up of the normal 10 weeks at that level of training plus the 5 additional weeks during the training freeze.
Nonsolo: Students who had not had enough civilian flying training to solo before entering primary.
Solo: Students who had had enough civilian flying training to solo before entering primary.

Results for Single-Engine Students

The results for primary, basic, and advanced are presented in tables 10.2, 10.3, and 10.4, respectively. Each of these tables presents the percent performing correctly for the groups with the four different amounts of training, and the reliability of the differences among these groups.

When a measure shows a significant difference between the groups compared, this means not only that the additional training produced an improvement but also that the objective measure employed was good enough to indicate that improvement. When no significant difference is shown, this may indicate any one or more of three things: the particular type of training produced no improvement, the measure employed was not reliable and discriminative enough to indicate the improvement, or that the measure did not test the particular type of improvement that occurred.

Effects of 5 weeks' additional training during the freeze.—The most relevant comparisons for determining the effects of the additional training at each level during the freeze are those between Group 1, 10 weeks nonsolo, and Group 2, 15 weeks nonsolo. It can be seen that when these two groups were tested under exactly equivalent conditions, the students with the additional training gave a measurably better performance at all three levels of training. The 15-week group was better than the 10-week one at primary on all 13, at basic on 15 out of 17 and at advanced on 10 of the 16 objective measures used.[16]

In primary, most improvement was shown in the following measures: (1) Maintaining constant altitude while making a 360° steep turn to the right immediately followed by a 360° steep turn to the left; (2) Performing an Immelmann without stalling, starting with an air speed no greater than 150 m. p. h. and coming out lined up within 45° of the correct heading; (3) Performing a loop without stalling, starting with an air speed no greater than 110 m. p. h. and coming out within 45° of the correct heading. In these three measures the superiority of the 15-week group was of a size which would be expected by chance less than 1 time in 100.

TABLE 10.2.—*Objective measures of the effects of additional training on the flying skill of primary students*

	Percent of students performing correctly				Probability of obtaining by chance a difference in the expected direction as large as the one observed between the groups			
Group number	1	2	3	4	2>1	4>3	3>1	4>2
Weeks of primary flying training	10	15	10	15				
Soloed before entering primary	No	No	Yes	Yes				
Number of students tested	211	954	60	228				
Immelmann.—Starts with air speed less than 150 m. p. h., completes without stalling and comes out within 45° of correct heading.	45	64	52	90	0.01	0.01	0.17	0.08
Loop.—Starts with air speed less than 110 m. p. h., completes loop without stalling and comes out within 45° of original heading	40	55	47	55	.01	.13	.17	(1)
Steep Turns.—Student rolls directly from 360° left turn into 360° right turn:								
Altitude deviation no more than 60 feet	40	64	70	71	.01	.44	.01	.02
Reached 60° bank in each turn	73	79	77	77	.04	(1)	.28	(2)
Roll-out within 45° original heading	98	99	100	96	.18	(2)	.12	(2)
Forced Landing.—Plane 1,200 feet right above field headed downwind:								
Student made a base leg	82	88	88	90	.02	.33	.12	.10
Would make first ⅓ field ¼ mile square.	65	70	80	77	.08	(2)	.01	.01
First Accuracy Landing.—Power-off 90° approach to zone 200 feet wide by 600 feet long:								
Landed in zone	45	51	61	52	.06	(2)	.01	.30
Landed in 3-point attitude	54	60	67	65	.04	(2)	.03	.08
Did not drop in or bounce 3 feet	55	59	57	63	.15	.20	.39	.14
Second Accuracy Landing.—Power-off 90° approach to zone 200 feet wide by 600 feet long:								
Landed in zone	56	60	66	63	.14	(2)	.08	.20
Landed in 3-point attitude	57	65	80	73	.02	(2)	.01	.01
Did not drop in or bounce 3 feet	59	61	75	63	.30	(1)	.01	.20

[1] No difference between the groups.
[2] Difference between the groups not in the expected direction.

[16] Attempts to conclude that the additional training at one level was of more value than that at another are not justified. The check rides used at the different levels are not necessarily equivalent. The 5 weeks was a 50-percent increase in the total weeks of flying training for the primary students, and 16⅔-percent increase for those in advanced. Furthermore, it is possible that a marked improvement showing up at the end of the primary phase may be entirely obliterated by the time the student graduates from advanced training.

In the basic schools the most improvement was shown in the measures of instrument flying. These were: (1) maintaining constant heading, air speed and altitude in straight-and-level instrument flight; (2) maintaining constant altitude and bank, and coming out on the correct heading in a 360° steep instrument turn; and (3) maintaining constant heading and air speed, and coming out at the correct altitude in a constant speed climb and descent. In each of the 9 measures involved in these 3 maneuvers, the superiority of the 15-week group was of a size that would be expected by chance less than once in 100.

TABLE 10.3.—*Objective measures of the effects of additional training on the flying skill of basic students*

	Percent of students performing correctly				Probability of obtaining by chance a difference in the expected direction as large as the one observed between the groups			
Group number	1	2	3	4	2>1	4>3	3>1	4>2
Weeks of Basic flying training	10	15	10	15				
Soloed before entering Primary	No	No	Yes	Yes				
Number of students tested	599	710	234	221				
Steep Turns.—Student rolls directly from 180° left turn into 180° right turn:								
Altitude deviation no more than 50 feet	33	38	41	44	0.03	0.26	0.02	0.06
Reached 60° bank in each turn	73	77	68	75	.05	.05	[2]	[2]
Accuracy Landing in zone 200 feet wide by 600 feet long using throttle and flaps as desired:								
Landed in zone	74	74	77	74	[1]	[2]	.18	[1]
Landed in 3-point attitude	70	75	74	71	.03	[2]	.13	[2]
Did not drop in or bounce 3 feet	42	44	46	47	.24	.42	.15	.22
Accuracy Landing in zone 200 feet wide by 600 feet long using flaps as desired:								
Landed in zone	57	56	64	65	[2]	.41	.03	.01
Landed in 3-point attitude	65	67	68	71	.23	.25	.21	.13
Did not drop in or bounce 3 feet	41	43	44	49	.24	.15	.22	.08
Straight and Level Instrument Flight.—Student flies straight and level at 130 m. p. h. for 3 minutes with directional gyro and artificial horizon caged:								
Heading deviation no more than 10°	46	58	49	55	.01	.10	.22	[2]
Air speed deviation no more than 5 m. p. h.	53	60	51	56	.01	.15	[2]	[2]
Altitude deviation no more than 50 feet	41	48	46	47	.01	.42	.10	[2]
360° Steep Instrument Turn.—Makes a 360° right turn with a bank of over 30° at an air speed of 130 m. p. h. and comes out on heading. Uses full instrument panel:								
Heading deviation no more than 5°	60	69	59	66	.01	.06	[2]	[2]
Altitude deviation no more than 100 feet.	63	76	70	74	.01	.18	.03	[2]
Maintained at least a 30° bank	87	91	86	93	.01	.01	[2]	.16
Instrument Constant-Speed Climb and Descent.—Flies straight and level at 110 m. p. h., climbs 1,000 feet at 500 feet per minute, and immediately starts descent to level off at starting altitude:								
Heading deviation no more than 10°	75	87	78	87	.01	.01	.18	[1]
Air speed deviation no more than 5 m. p. h.	27	37	29	41	.01	.01	.29	.18
Level off within 50 feet correct altitude top of climb and end of descent	56	66	60	60	.01	[1]	.16	[2]

[1] No difference between the groups.
[2] Difference between the groups not in the expected direction.

For Advanced, the group with more training showed the greatest superiority in the measures on the instrument let-down and low approach; they were better at hitting the edge of the field "on the beam" and at the proper altitude. On these two measures, the difference in favor of the 15-week group was of a size which would be expected by chance less than 1 time in 100.

TABLE 10.4.—*Objective measures of the effects of additional training on the flying skill of advanced single-engine students*

	Percent of students performing correctly				Probability of obtaining by chance a difference in the expected direction as large as the one observed between the groups			
Group number	1	2	3	4				
Weeks of advanced single-engine flying training	10	15	10	15	2>1	4>3	3>1	4>2
Soloed before entering primary	No	No	Yes	Yes				
Number of students tested	428	931	62	119				
Two Maximum Performance Loops.—Starts with air speed less than 160 m. p. h., 1,925 r. p. m., and manifold pressure of 25″ Hg., and completes loop without stalling out:								
First loop completed without stalling or adding throttle	57	62	56	64	0.04	0.15	([2])	0.34
Second loop completed without stalling or adding throttle	67	71	60	76	.07	.02	([2])	.12
Steep Turns.—Student rolls directly from 180° left turn into 180° right turn:								
Altitude deviation no more than 100 feet	53	57	57	69	.05	.06	.23	.01
Reached 60° bank in each turn	86	88	84	88	.21	.24	([2])	([1])
Accuracy Landing in zone 200 feet wide by 600 feet long using throttle and flaps as desired:								
Landed in zone	75	73	64	73	([2])	.11	([2])	([1])
Landed in 3-point attitude	69	67	67	76	([2])	.11	([2])	.02
Did not drop in or bounce 3 feet	39	45	53	49	.02	([2])	.02	.21
Straight and Level Instrument Flight.—Student flies straight and level at 130 m. p. h. for 3 minutes with directional gyro and artificial horizon caged:								
Heading deviation no more than 10°	54	57	61	56	.16	([2])	.18	([2])
Air speed deviation no more than 5 m. p. h.	53	51	51	48	([2])	([2])	([2])	([2])
Altitude deviation no more than 50 feet	52	48	43	53	([2])	.10	([2])	.18
360° Steep Instrument Turn.—Makes a 360° right turn with a bank of over 30° at an airspeed of 130 m. p. h. and comes out on heading. Uses full instrument panel:								
Heading deviation no more than 5°	76	73	75	81	([2])	.18	([2])	.02
Altitude deviation no more than 100 feet	81	82	79	81	.33	.38	([2])	([2])
Maintained at least a 30° bank	90	89	97	83	([2])	([2])	.01	([2])
Instrument Let-Down and Low Approach.—Makes a regular let-down and low approach; shakes stick at the proper number of seconds after the low cone. Starts on the beam at proper altitude about 1 minute before the high cone and headed toward it:								
Reached the edge of the field no more than 400 feet outside of the beam	58	68	63	64	.01	.45	.22	([2])
Reached the edge of the field within 5 seconds of prescribed time after the low cone	44	47	54	48	.16	([2])	.07	.42
Completed let-down within 60 feet of prescribed altitude at edge of field	71	77	79	75	.01	([2])	.08	([2])

[1] No difference between the groups.
[2] Difference between the groups not in the expected direction.

For the students who had soloed before entering Primary, the additional five weeks of training produced some improvement, but the differences were not as large or consistent as they were with those who had had no previous flying experience.

Effect of having had enough civilian flying training to solo before entering Primary.—At the end of 10 weeks of Primary training the students who had soloed before starting military flying training were better than those who had not received this training in all 13 of the objective measures used. This superiority may have been due to the additional training or it may have been due to a selective factor, with the students who had the most interest and aptitude taking the trouble to get this additional training.

The superiority of the students who had had enough previous flying training to solo before starting military flying persisted for a while but became progressively smaller in the more advanced phases of army training. The students who had soloed were better than those who had not in 10 out of the 13 objective measures at the end of 15 weeks of Primary flying and in 13 out of 17 at the end of 10 weeks of Basic flying. By the time they had completed 15 weeks of Basic flying training the consistent difference had disappeared; there was no consistent difference between the solo and nonsolo groups in the tests administered at the end of 15 weeks of Basic flying or 10 and 15 weeks of Advanced single-engine training.

These results are in line with what is known about the way having soloed before starting army flying training affects the likelihood that the student will be eliminated. The students who had already soloed were considerably less likely to be eliminated from Primary training than those who had not. The elimination rates for a typical class of approximately 7,000 students were 6 percent for students with previous solo and 33 percent for those without it. By the time these students reached Basic schools this difference was greatly reduced, becoming 8 percent and 14 percent, respectively. Furthermore, as will be shown in the next chapter, the students with previous solo are no better than those without it in the air-to-air and air-to-ground fixed gunnery practice which is given immediately after graduation from Advanced training.

EFFECT OF ADDITIONAL TRAINING ON ADVANCED TWO-ENGINE FLYING PROFICIENCY

Groups Compared

The fact that the advanced two-engine schools were in the process of changing over from the AT–10 and UC–78 planes to the much higher-powered TB–25 introduced certain different features into the design of the two-engine section of the training freeze study.

Students were tested in only those schools which were using the TB–25. In all of these schools, the students with 10 weeks of Advanced training had received all of it on the TB–25. In two of the schools, the students with 15 weeks of Advanced training had received all of it on the TB–25; in two other schools, however, students who had received 10 weeks of training on the lower-powered AT–10 or UC–78 received 5 weeks of additional training on the much higher-powered TB–25 during the first training freeze. This meant that, although they had a total of 15 weeks Advanced training, only 5 weeks of this training had been on the higher-powered TB–25.

So few Advanced two-engine students had soloed before entering Primary flying training that they were not sorted out of the sample for separate analysis. All students were tested on the TB-25. Hold-overs were excluded.

The composition of the three groups compared in the Advanced two-engine study was as follows:

1. *10-5-week group.*—The students in this group had received 10 weeks of Advanced two-engine training on the UC-78 or AT-10 and 5 weeks of training on the TB-25. The group consisted of 290 students in Class 44-I.

2. *0-10-week group.*—The students in this group had received the normal 10 weeks of Advanced two-engine training on the TB-25. The group consisted of 730 students in Class 44-J.

3. *0-15-week group.*—The students in this group had received the normal 10 weeks of Advanced two-engine training on the TB-25 plus 5 weeks of additional training on the same plane during the training freeze. The group consisted of 275 students in Class 44-I.

Results of Comparing Different Types of Advanced Two-Engine Training

Measures on a sample maneuver.—First, the results for one sample maneuver, single-engine procedure, will be considered. In this maneuver, when the student was flying under standard conditions, the instructor cut one throttle completely and thus simulated the condition of one engine failing. The student's task was to keep the plane from swerving toward the dead engine, go through the proper procedure of feathering the dead propeller and increasing power in the good engine, and hold the loss of air speed and altitude to a minimum.

The results are presented in figure 10.1. It can be seen that the three groups were not different in their ability to hold heading. In the important aspects of holding airspeed and altitude, however, there was progressive improvement with additional training on the TB-25. It is especially interesting to note that students with only 10 weeks of Advanced training, all on the TB-25, were definitely better than those with a total of 15 weeks Advanced training, 10 of which were on a lower-powered plane. Apparently the transfer of training was imperfect enough so that the 10 weeks on the lower-powered plane did not help as much as 5 weeks on the higher-powered TB-25.

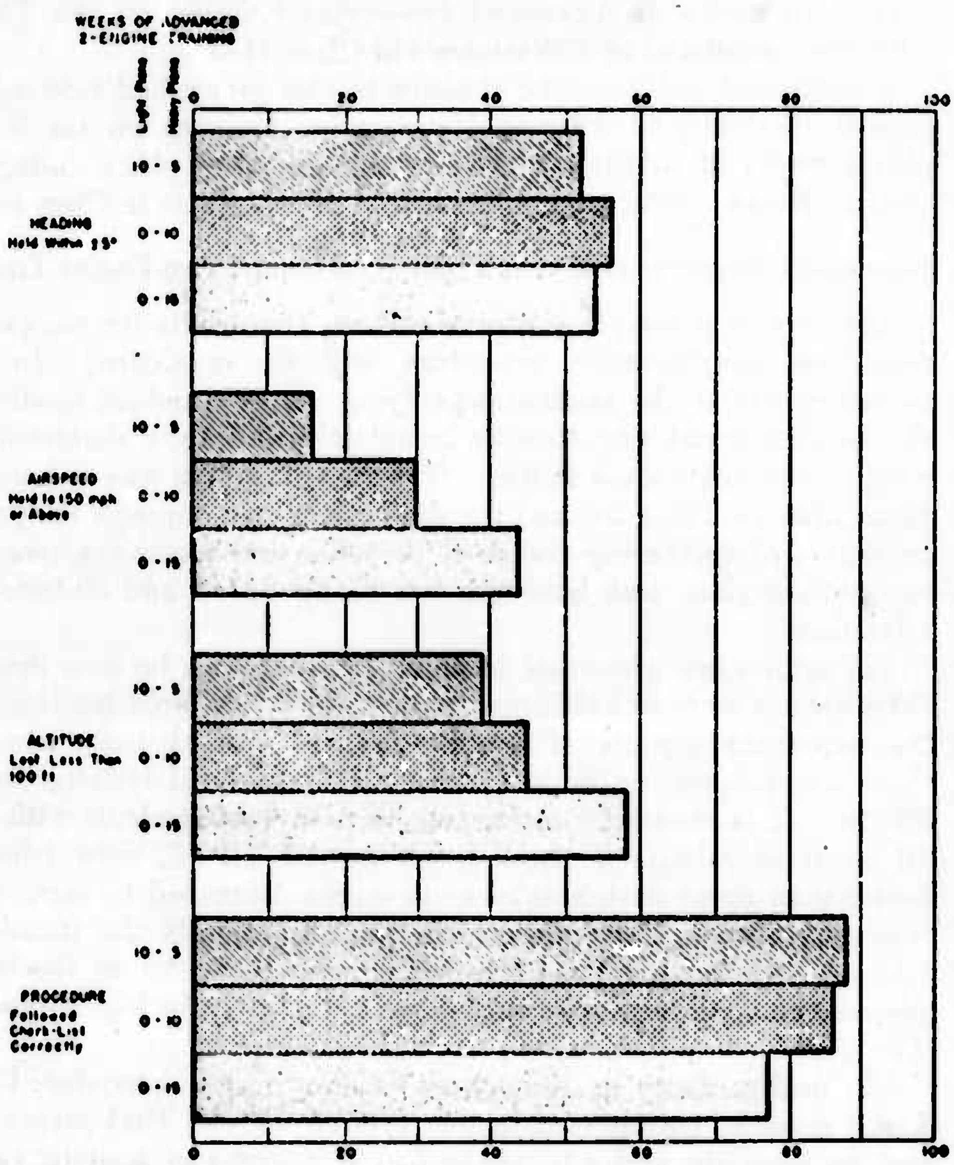
The final measure on the graph, "Following the Procedure Check List," must be interpreted in the light of the fact that proper procedures were stressed during the first few weeks of training on the TB-25 and relatively neglected after that. The progressive forgetting shown in the graph clearly indicates the need for more review on this aspect of training.

Fig. 10.1

RESULTS ON ONE OF THE MANEUVERS USED TO MEASURE OBJECTIVELY EFFECTS OF ADDITIONAL TRAINING

SINGLE-ENGINE PROCEDURE IN THE B-25

With plane flying on a cardinal heading at an airspeed of 160 mph, wheels down, instructor completely cuts one throttle, simulating engine failure, and measures the student's performance for 2 minutes
290 students in 10-5 week group, 730 in 0-10 week group, 275 in 0-15 week group



PROBABILITY OF OBTAINING, BY CHANCE ALONE, A DIFFERENCE AS LARGE AS THAT OBSERVED.

	Heading	Airspeed	Altitude	Procedure
10-5 vs 0-15	.64	less than .001	less than .001	less than .001
10-5 vs 0-10	.26	less than .001	.08	.40
0-10 vs 0-15	.58	less than .001	less than .001	less than .001

Psychological Research Project (Pilot), Randolph Field, Texas

Measures on all maneuvers.—The results on all of the measures used at the advanced two-engine level of training are shown in Table 10.5. The chief findings were:

1. In general, the students with 15 weeks of advanced two-engine training, but only 5 of these on the TB-25, gave a poorer performance than those with only 10 weeks of advanced two-engine training, all of which was on the TB-25. The 10-5-Week group was poorer than the 0-10-Week one on 18 of the 20 objective measures used. Apparently the transfer of training was imperfect enough so that the 10 weeks on the lower-powered plane did not help as much as 5 weeks on the higher-powered one.

TABLE 10.5.—*Objective measures of the effects of additional training on the flying skill of advanced two-engine students*

	Percent of students performing correctly			Probability of obtaining by chance a difference in either direction as large as the one observed between the groups [2]		
Group number	1	2	3			
Weeks of advanced flying training on light two-engine planes (UC-78, AT-10)	10	0	0	3 versus 1	2 versus 1	3 versus 2
Weeks of advanced flying training on heavy two-engine plane (TB-25)	5	10	15			
Number of students tested	[illegible]	730	275			
Single-Engine Procedure.—Starting conditions: wheels down, cardinal heading, air speed 160 m. p. h., 2,200 r. p. m. and enough manifold pressure to maintain 160 m. p. h.:						
Maintained heading within 5°	52	56	54	0.64	0.26	0.56
Held air speed to 150 m. p. h. or over	16	30	44	.01	.01	.01
Altitude loss less than 100 feet	39	45	58	.01	.09	.01
Procedure check-list followed correctly	68	86	77	.01	.40	.01
Single-Engine Landing.—Check pilot cuts inside engine after completion of the GUMP check on downward leg:						
Lined up with runway after turn into approach	55	56	62	.09	.78	.09
Held 140 m. p. h. air speed on approach within 10 m. p. h	53	55	61	.06	.57	.09
Checked gear on approach	53	54	49	.35	.78	.16
Cut power at roundout or landing	60	58	60	(1)	.57	.57
Landed in 500 foot target zone of runway	29	23	24	.79	.06	.13
Did not land in skid	52	57	62	.02	.16	.16
Straight and Level Instrument Flight.—Student flies straight and level and normal cruising air speed for 3 minutes with directional gyro and artificial horizon caged:						
Heading deviation no more than 5°	29	32	46	.01	.35	[illegible]
Air speed deviation no more than 5 m. p. h	52	65	48	.35	.01	.01
Altitude deviation no more than 50 feet	46	52	45	.81	.09	.06
Instrument Let-Down and Low Approach.—Makes a regular let-down and low approach; shakes the stick at the proper number of seconds after the low cone. Starts on the beam at proper altitude about one minute before the high cone and headed toward it:						
Reached the edge of the field no more than 400 ft. outside of the beam	43	55	68	.01	.01	.01
Reached the edge of the field within 5 seconds of prescribed time after the low cone	45	51	51	.16	.09	(1)
Completed let-down within 40 ft. of prescribed altitude at edge of field	47	56	64	.01	.01	.03
Power-off Approach and Landing.—Makes a power-off approach and landing. Cuts power completely at traffic altitude when in position to land in the target zone with power off all the way down:						
Altitude at roll into approach deviated no more than 100 ft. from prescribed altitude	74	84	72	.60	.01	.01
Did not add power on approach	63	70	74	.01	.04	.21
Landed in target zone without adding power	27	31	37	.04	.31	.16
Landed in target zone after adding power	28	40	43	.04	.03	.66

[1] No difference between the groups.
[2] In computing these probabilities both tails of the distribution were used because there was no *a prior* reason for expecting the difference between groups to be in a particular direction.

2. Although the students with 15 weeks of training on the TB-25 in general tended to be the best, the additional advanced two-engine training did not improve skill on all of the aspects of performance measured.

a. Most improvement was shown in hitting the edge of the field "on the beam" at the correct altitude on the instrument let-down and low approach, and in maintaining altitude and air speed on the single-engine procedure.

b. In the single-engine procedure, however, the students with more training were progressively more careless about performing the whole procedure check list correctly. Those with 15 weeks of training on the TB-25 were also somewhat more likely to omit the landing gear check on the single-engine landings; in fact, a surprisingly high proportion, more than 45 percent, of all students failed to check their landing gear after lowering it on approach. Procedure items were stressed during the first part of training and apparently were not reviewed enough during the latter part of the course to make up for the effects of forgetting.

SUMMARY AND CONCLUSIONS

While it was still in the early stages of developing objective measures of flying skill, the Pilot Project was asked by the Chief of Staff of the Training Command to use the opportunity afforded by temporary training freezes to develop objective measures for all levels of training and to measure the effects of additional training at each level.

With the aid of pilots on Training Advisory and Standardization Boards, objective check-ride scales were developed for each of the phases of training: Primary, basic, advanced single-engine and advanced two-engine. The measures in these check rides were designed to sample differences between groups; they were not designed to be reliable enough to discriminate between individuals.

These objective measures were administered during a single week to over 8,000 students in the four different phases of training. When the students whose selection or training was atypical were sorted out to purify the samples, the population was reduced to 6,052 students.

In the primary, basic, and advanced single-engine schools, students who had received the normal 10 weeks of training and students who had received an additional 5 weeks during the freeze (making a total of 15 weeks) were tested at the same time under comparable conditions. The results showed that at each of these three levels of single-engine training, adding an additional 5 weeks produced a definite improvement in the performance of the students. At none of these levels had the students reached their limit of improvement within the normal period of training.

At the advanced two-engine schools three different groups were tested on the TB-25. One of these had received 10 weeks of normal advanced two-engine training on the lower-powered UC-78 or AT-10 followed by 5 weeks on the faster and heavier TB-25. Other students had received 10 and 15 weeks of advanced two-engine training, respectively, all on the TB-25. The results of testing showed that in general 10 weeks of advanced two-engine training, all on the TB-25, was better than 15 weeks of training, 10 of which was on lower-powered two-engine planes and only 5 on the TB-25. Apparently, the transfer of training was imperfect enough so that 5 weeks on the larger plane was of more value than 10 on the smaller one.

In general, the students with 15 weeks of training, all on the TB-25, were the best of all, but the additional training did not help all aspects of performance. During the additional training there was a tendency to forget those aspects of procedure which were emphasized during the first weeks and relatively neglected later on.

An investigation was also made of the difference between those single-engine students who had had enough civilian flying training to solo before entering primary and those who had not. It was found that during the earlier phases of military training the students who had soloed before primary were definitely superior to those who had not. But the difference between these two groups did not seem to persist beyond the basic level of flying training. These results are in line with what is known about the effects of previous civilian flying upon the likelihood that the student will be eliminated at the different levels of training.

The fact that it was possible to achieve the results reported when objective measures were used at 36 different schools to test the flying skill of over 8,000 students in a single week, indicates that it is practicable to employ such measures in evaluating the results of a large-scale pilot training program.

CHAPTER ELEVEN

Fixed Gunnery as an Objective Measure of Flying Skill

Sgt. John G. Gleason

JOB ANALYSIS OF THE FIXED GUNNERY TASK IN THE TRAINING COMMAND [a]

An important aspect of a fighter-pilot's task was to register hits on air and ground targets by firing guns fixed to his plane. Therefore, the ability to hit similar targets while in training constituted an objective measure of one important area of fighter-pilot proficiency.

The pilot's task in fixed gunnery firing involved two major skills: (1) perceptual skill in evaluating the aim and (2) skill in flying the plane so as to obtain and keep the correct aim while maintaining smoothly coordinated flight. A high degree of both of these skills was needed; the development of either without the other was valueless.

The sight picture in fixed gunnery is different from the sight picture obtained with any conventional gun sight. A summary of the sight picture, as the fighter-pilot sees it when using the optical gun sight, is presented in the following paragraphs. No attempt is made in this discussion to explain the optics of this sight.

As the pilot is seated in the cockpit, he looks straight ahead through a tilted glass reflector plate. The glass appears perfectly clear, except for the lighted image of the sight. The sight appears as a ring with a bead as its center.

The sight is so installed in the plane that it is harmonized with the path of the bullets and the path of the plane. That is, the sight, the airplane, and the gun are all aligned in such a manner that, within the effective range of the gun, the bead of the sight indicates the path the bullets follow. The image of the sight appears out in space at the distance at which the pilot's eyes are focused. The pilot may move his head about in any direction and the sight image which appears out in space will continue to indicate the same point of aim.

The ring of the sight measures angular distance, i. e., the distance between the bead and any point on the ring of the sight is a measure of angular distance. This measure of angular distance serves two functions: if the pilot knows the size of a target he can estimate its range; if he knows its speed, he can estimate the angle by which he must lead it.

[a] Work in this section was done by Sgt. John G. Gleason.

The apparent diameter of the ring used in the training sight was 84 feet at a range of 1,200 feet, the maximum range for effective firing. Knowing this, the pilot needed only know what portion of his sight ring the target should fill in order to determine whether he was in or out of range. For example, in air-to-air firing, a tow target 20 feet long was employed. When this target was viewed broadside, it filled one quarter of the ring at a range of 1,200 feet. If the target were closer than 1,200 feet, it would appear to fill more than one quarter of the ring; or, if the target were more than 1,200 feet away, it would appear to fill less than one quarter of the ring.

A target moving straight away from the pilot does not appear to move across his line of sight and can be hit by aiming directly at it. For any target which is moving across the line of sight, however, the pilot must aim ahead of it to hit it. There is no accurate way of measuring lead with the conventional sight. The ring and bead of the optical sight give a basis for estimating angle of lead. Since the angle of lead remains constant (for any given target speed) regardless of changes in range, and since the optical sight measures the angle of lead, it is unnecessary to know the distance to the target in order to determine the proper lead.

To obtain maximum effectiveness against a moving target the pilot had to be able to fire at it for a period of time. The only way in which a pilot could do this, without undue exposure to defensive fire from the attacked aircraft, was to attack from the side, turning with the target and firing while turning. Such a path of flight, which keeps the target constantly under fire, is called a "curve of pursuit."

In firing on a target, while flying a "curve of pursuit," both the angle of fire on the target and the angle of lead changes. The maximum angular lead is required when firing at right angles to the path of the target and varies to no lead when the attack is from behind or ahead.

The angle determined by the bead and ring of the sight used in training was such that when firing at a 90° angle to a target moving 65 miles per hour, the correct lead was 1 radius, or in other words, the angle formed between the bead and any point on the ring. If the target moved twice as fast, or 130 miles per hour, as did the tow target used in training, the lead also had to be doubled and therefore became 2 radii.

The pilot's angular lead varies as the sine of the angle of attack. Since it was obviously impossible for the pilot to estimate each change in angle of attack with extreme accuracy, he learned the correct lead for several angles of attack that were easy for him to estimate accurately. Gunnery students usually learned the sight picture for every 15° as follows:

Angle of attack:

	Radius lead
90°	2.0 radii
75°	1.9 radii
60°	1.7 radii
45°	1.4 radii
30°	1.0 radii
15°	0.5 radii

The pilot then flew the plane as smoothly as possible from one known sight picture to the next so that the lead was gradually but constantly changing.

As can be seen, the pilot had to make two related judgments. He had to recognize the sight picture which gave each of the listed radii lead. This required that the pilot have good perceptual memory. He also had to be able to judge accurately his angle off, i. e., judge the angle between the path of his bullets and the path of the target, in order that he would use the correct sight picture. If he erred in either of these two judgments he would not hit the target. With practice, these two judgments were reduced to one.

With the preceding description of the optical gun sight as a background, we can now consider the specific task presented to the gunnery student in training.

Air-to-Ground Fixed Gunnery

This was the less difficult of the two types of fixed gunnery task. Air-to-ground training was given to familiarize the student with attacking ground targets. In combat, such targets included tanks, trains, grounded aircraft and ships. In training a 6 by 6 foot rifle target inclined at a 60° angle was used.

The attack consisted of a shallow dive (30°) with the bead of the sight centered on the target. The dive was started from an altitude of 800 feet above the ground and carried the plane down to approximately 300 feet. The dive was made at a speed of approximately 150 miles per hour. Firing was started when the plane came within range, as indicated by a white line on the ground 1,200 feet from the target. Firing stopped before the plane passed the foul line, indicated by a second white line on the ground, 600 feet from the target. On passing this point the pilot immediately began a climbing turn back to normal altitude in preparation for the next attack on the target.

Hitting the target on a calm day required the pilot to fly the plane in a perfectly coordinated dive at the target aiming the plane so that the bead of the sight was on the bull's-eye. When firing in any wind, a correction in sighting had to be made. The sight had no adjustment for wind correction so the allowance for windage was made by aiming slightly off the bull's-eye. When flying in a crosswind, the pilot aimed slightly upwind of the bull's-eye to allow for the downwind drift of the bullets. In order for the pilot to hold such a point of aim in a crosswind, it was necessary for him to fly a very slight, coordinated turn. In a headwind, the pilot aimed high and in a tailwind he aimed low.

The flying task in air-to-ground fixed gunnery was largely a matter of coordinated flying. The slightest slip or skid of the plane while sighting on the bull's-eye caused the bullets to go wide of the target. At a range of 1,200 feet, a skid of only 5° caused the bullets to miss the target by 13 feet.

The perceptual aspect of the air-to-ground gunnery problem was relatively simple. The pilot needed only to see whether the bead was on the target or not. Errors due to the effect of the wind were immediately noted from spurts of earth thrown up by the bullets and corresponding corrections in aim could be made.

Air-to-Air Fixed Gunnery

This was the more difficult type of fixed gunnery firing and required good judgment of relative speed and distance, accurate judgment of angles, perception of the correct sight pictures and precision flying. It was in this type of firing that the optical sight, described earlier, was most valuable.

The purpose of air-to-air firing was to familiarize the student with the task of attacking another aircraft. The training closely simulated an attack on a large aircraft whose ability to perform evasive action was negligible. A banner type target measuring 4 feet by 20 feet was used and was towed behind an AT-6 type aircraft.

The procedure in air-to-air fixed gunnery was for a flight of seven planes to be assigned a specific range and altitude for firing. The tow ship towed the target straight and level at an air speed of 130 miles per hour, back and forth over the range. The target was towed in this manner for all air-to-air fixed gunnery practice so that the students could devote all of their attention to timing their approach and determining the correct position to open fire.

In flying an air-to-air gunnery mission the pilot approached the target by flying parallel to its path but in the opposite direction. When the target appeared off at a 45° angle to his line of flight, the pilot began to roll into his curve of pursuit. As the pilot continued the turn he reached an angle of fire of 90° when he was approximately 1,200 feet from the target. By the time this angle of fire had been closed to 60° the target was well "sighted" and firing begun. It continued until the angle of fire approached 30° when the pilot was forced to break off the attack in order not to endanger the tow ship. The pilot then regained his position in traffic to prepare for the next attack on the target.

Hitting the towed target required a number of highly developed skills. First, the pilot needed to judge the proper distance from the path of the target to begin his approach. This was difficult because of the high rate of closure (280 miles per hour) between the attacking plane and the target. Although difficult, the task was crucial, since, if his path were too close to the target when the turn was started, it forced the pilot into an impossibly tight curve of pursuit resulting in only a brief instant in which the proper lead could be obtained before the pilot could no longer turn fast enough to keep the target in his sights. Should his path be too far out from the target's path when the turn was started, the pilot found himself unable to get in range to fire. An allied problem of judgment was the timing of the start of the roll into the curve of pursuit. The penalties for error here were approximately the same as those for misjudging distance. If the roll-in were started too soon, the pilot would not be able to turn with the target since he was too close in by the time it reached him, or if the turn were started too late, he was never able to get within range to open fire.

The pilot also had the perceptual task of judging his range and his angle of fire, so that he could apply the correct sight picture. He had to visualize the correct sight picture for each angle of fire, and hold the plane on the course which maintained the correct lead.

Throughout all this action the pilot had to maintain perfectly coordinated flight. As in air-to-ground firing, the slightest slip or skid of the aircraft threw the bullets wide of their mark.

In brief, then, in air-to-ground fixed gunnery, the chief problems were the coordinated flying of a 30° dive and correcting for drift. In air-to-air firing, the task of coordinated flying was again of basic importance but the level of difficulty of the maneuver was greatly increased. The curve of pursuit was a steep turn requiring the pilot to fly the plane near its limits of performance. In addition, the pilot had to make accurate judgments of speed and distance in order to place and time his curve of pursuit properly. Finally, in order to fly the curve correctly, the pilot had to judge his angle of fire with accuracy while simultaneously achieving the proper sight picture for each angle.

EVALUATION OF FIXED GUNNERY SCORES AS CRITERION DATA [*]

Importance of Evaluating Fixed-Gunnery Scores as Criterion Data

The job analysis of the student's task in the two types of fixed gunnery training shows that flying skill is an important factor in both. It is quite possible, however, for a measure to appear objective and logical and yet to be so distorted by extraneous factors, over which little or no control may be exercised, that the scores are so unreliable that they do not measure the level of ability of the students.

---

[*] Except where otherwise specified, work in this section was done by Maj. Glen L. Heathers, Lt. William E. Galt, Lt. Ralph M. Rust, and Sgt. John O. Gleason.

Reliability of Air-to-Ground and Air-to-Air Fixed Gunnery Scores

The term reliability refers to the consistency of measurement. The gunnery scores would be said to be highly reliable if the student pilots who scored high on the odd-numbered gunnery missions also scored high on the even-numbered missions and if the students who scored low on the odd-numbered missions also scored low on the even-numbered missions. If the scores on the odd missions did not agree with the scores on the even missions, it could hardly be expected that they would be related to anything else. If the pilots who scored highest on one mission are just as likely as any others to score the lowest on the next, the scores would not be a measure of their fixed gunnery ability.

It was possible that the results of a few rounds of firing would give an unreliable picture of the proficiency of the pilot, while those of many rounds fired during a number of successive missions would give a reliable measure. It was therefore desirable to know how many rounds had to be fired in order to secure a reliable score.

Since, in general, the reliability of a score is increased by increasing the number of measurements included, the determinations of reliability were based on the same number of rounds as were the gunnery scores used in the validation of the classification tests. By inspection of the data, it was determined that the greatest number of rounds fired by all students was 1,200 rounds air-to-air and 400 rounds air-to-ground. The reliability measures were therefore based upon these numbers of rounds.

The reliabilities were determined by product-moment correlation of scores on odd- and even-numbered missions. Normally from two to three missions were flown in a day. A mission equalled 100 rounds fired in all cases except air-to-air firing in the Eastern Flying Training Command. There the first air-to-air mission was 100 rounds and all succeeding missions were 200 rounds. In instances where a student fired less than the 100 or 200 rounds prescribed, the scores were made comparable by the use of the percent of hits made. The mission was omitted entirely from the calculations when less than 25 rounds were fired.

The means, sigmas, and reliability coefficients as determined for Class 44-1-G from Western Flying Training Command, 44-1-H from Central Flying Training Command, and Class 44-1-H from Eastern Flying Training Command are presented in table 11.1 for Air-to-Ground Fixed Gunnery, and table 11.2 for Air-to-Air Fixed Gunnery. The reliability coefficients were corrected for double length by use of the Spearman-Brown formula. The average reliability for all schools and commands combined by use of Fisher's z-transformation was 0.59 ($N=932$) for air-to-ground and 0.63 ($N=1064$) for air-to-air fixed gunnery.

Applicability of Spearman-Brown Formula.—The reliability figures presented in tables 11.1 and 11.2 are, of course, specific to the number of rounds on which they are based. A determination based on fewer rounds fired would be expected to be lower and one based upon more rounds would be expected to be higher. If one can assume that the same skills were involved throughout the course of learning fixed gunnery so that no qualitative change occurred in the task measured, the way in which reliability varies with the number of rounds fired should be exactly predicted by the Spearman-Brown formula.[61]

TABLE 11.1.—*Reliability of 400 rounds of air-to-ground fixed gunnery*

Classes 44-1-G in the Western and 44-1-H in the Central and Eastern Flying Training Commands

Transition school	N	M [1]	SD [2]	r'_{11} [3]	r_{11} [4]
Western:					
Williams [5]	87	88.94	32.05	0.51	0.67
Williams	58	91.57	30.00	.51	.67
Eastern:					
Craig	60	87.57	29.12	.22	.35
Do	45	82.51	22.50	−.09	....
Spence	66	127.52	41.49	.38	.55
Do	47	97.69	38.27	.21	.35
Napier	90	84.03	36.94	.50	.67
Do	101	98.35	38.02	.42	.59
Central:					
Foster	55	130.05	44.29	.49	.66
Do	61	131.16	50.55	.45	.62
Aloe	74	121.66	44.04	.62	.76
Do	62	138.22	47.03	.64	.78
Moore	84	123.18	42.52	.33	.50
Do	52	146.13	45.03	.49	.66
z-Average	932	....	....	....	.59

[1] Mean of cumulated mission scores.
[2] SD of cumulated mission scores.
[3] Odd-even reliability of four 100-round missions.
[4] Odd-even reliability corrected for double length (400 rounds) by the Spearman-Brown formula.
[5] The first of each pair is the Pre P-40 section; the second is the Post P-40 section.

TABLE 11.2.—*Reliability of 1,200 rounds of air-to-air fixed gunnery*

Classes 44-1-G in the Western and 44-1-H in the Eastern and Central Flying Training Commands

Transition school	N	M [1]	SD [2]	r'_{11} [3]	r_{11} [4]
Western:					
Williams [5]	87	166.93	57.05	0.69	0.81
Williams	58	137.45	53.35	.55	.71
Eastern:					
Craig	60	143.05	62.95	.54	.70
Do	50	104.58	38.72	.23	.38
Spence	64	144.72	61.73	.46	.63
Do	47	105.77	45.28	.39	.56
Napier	93	107.90	44.38	.31	.47
Do	103	126.70	64.65	.49	.66
Central:					
Foster	65	121.42	49.28	.61	.76
Do	90	119.58	47.19	.46	.63
Aloe	85	114.93	51.30	.45	.62
Do	80	125.36	52.47	.41	.58
Moore	84	130.51	56.00	.56	.72
Do	86	119.69	45.30	.46	.63
z-Average	1,064	....	....	....	.63

[1] Mean of cumulated mission scores.
[2] SD of cumulated mission scores.
[3] Odd-even reliability of twelve 100-round mission scores.
[4] Odd-even reliability corrected for double length (1,200 rounds) by the Spearman-Brown formula.
[5] The first of each pair is the Pre P-40 section; the second is the Post P-40 section.

[61]
$$r_{11} = \frac{A r'_{11}}{1-(A-1)r'_{11}}$$
where r_{11} is the reliability of the lengthened test, r'_{11} is the reliability of the shorter test, and A is the factor by which the shorter test is lengthened.

Available evidence on this point seems to indicate that the relationship between the reliability of fixed gunnery scores and the number of rounds involved actually does vary in the way predicted by the Spearman-Brown formula. The reliability of 400 rounds of air-to-ground firing was found to be 0.59 when the coefficients for all schools and commands were combined by Fisher's z-transformation technique. From the Spearman-Brown formula, it was predicted that the reliability of 600 rounds would be 0.68. For 265 students in Class 44-1-J in Central Flying Training Command who had fired 600 rounds of air-to-ground gunnery, the reliability was found to be 0.69.

In a study at Psychological Research Unit No. 1, based on 523 students in Class 43-F in the Eastern Flying Training Command, it was found that the reliability of the 822 rounds fired on the last 2 days of air-to-air training was 0.52. On the basis of this figure, the estimated reliability for all of the rounds fired during training, a total of 1776, should be 0.70. The actual reliability of the 1,776 rounds was found to be 0.69. The fact that these two figures agree closely indicates that the assumptions of the Spearman-Brown formula were satisfied by the conditions of air-to-air fixed gunnery learning in the Training Command. The fact that, as will be shown, the students continued learning and improving their scores throughout training did not seem to affect the estimates of reliability.[a]

In the study by the Pilot Project, the reliability of 1,200 rounds air-to-air gunnery was 0.63. On the basis of this figure, the reliability of 1,776 rounds was predicted to be 0.72. The agreement of this estimate with the actual figure of 0.69 is well within the limits of sampling errors to be expected in the populations of the two studies.

To summarize, the reliability of fixed gunnery scores varied with the number of rounds involved in the determination. The relationship between the number of rounds and reliability was approximately what the Spearman-Brown formula predicted. This indicates that scores at different levels of the learning curve are not qualitatively different. They measure the same basic kind of skill, though of course the level of skill depends upon the amount of training.

Results of other studies.—Reliabilities in the same general range as those found in the studies by the Pilot Project were reported by other organizations, but no exact comparisons are possible because the numbers of rounds involved were not reported.

Psychological Research Unit No. 1 reported the odd-even reliability of air-to-air fixed gunnery scores for three classes in the Eastern Flying Training Command as follows: (1) Class 43-J, for 459 students the reliability was 0.31 uncorrected and 0.47 corrected; (2) Class 43-K, for 403 students the reliability was 0.48 uncorrected and 0.65 cor-

[a] This study was performed by Maj. N. E. Miller, Lt. S. M. Roshal, and Lt. W. H. Angoff. It was the first investigation by aviation psychologists of fixed gunnery scores in the AAF.

rected; and (3) Class 44-1-A, for 119 students the reliability was 0.63 uncorrected and 0.77 corrected.[63]

A further study by the same unit reported the odd-even reliabilities for class 44-1-D in the Eastern Flying Training Command by transition training school. Following are the odd-even reliabilities:

School	r'_{11}	r_{11}
Craig	0.47	0.64
Mariana	.65	.79
Napier	.52	.68
Spence	.58	.73

The Research Section, Office of the Surgeon, Headquarters First Air Force, reported the following odd-even reliability coefficients for air-to-ground fixed gunnery and low-altitude-aerial fixed gunnery at the tactical training level:

Base	Air-to-ground			Low-altitude aerial		
	N	r'_{11} [1]	r_{11} [2]	*N*	r'_{11} [1]	r_{11} [2]
Millville	91	0.13	0.23	135	0.38	0.55
Norfolk	139	.38	.55	144	.50	.67
Dover	175	.44	.61	184	.31	.47

[1] Odd-even reliability.
[2] r_{11} corrected by Spearman-Brown formula.

The research section, Headquarters Second Air Force reports an odd-even reliability of 0.61 for 14 air-to-air missions of at least 100 rounds each. Corrected for double length, this becomes 0.76.

Intercorrelation of Air-to-Air and Air-to-Ground Fixed Gunnery Scores

How similar are the skills involved in air-to-air and air-to-ground fixed gunnery? If the two scores measure separate skills, it may be desirable to use them as separate criteria; if they measure the same skills, the combined score is all that need be used.

The correlation was calculated between the mean mission score for 4 air-to-ground missions, and the mean mission score for 12 air-to-air missions. One exception to this was the data for class 44-1-G in the Western Flying Training Command in which the air-to-ground mean mission score was based upon six missions.

Means, sigmas, and correlations for the classes studied in each command will be found in table 11.3. The average correlation, as determined by Fisher's z-transformation for all classes studied in the Eastern Flying Training Command was 0.22 (N=1175) and for all classes studied in the Western Flying Training Command it was 0.33 (N=383).

[63] This study was performed by Lt. Wilse B. Webb, Sgt. Roland E. Johnston and Sgt. Samuel B. Lyerly, and the following one by Lieutenant Webb and Sergeant Johnston.

Table 11.3.—*Correlation between air-to-ground and air-to-air fixed gunnery scores*

Class	N	M_1[1]	M_2[2]	SD_1[1]	SD_2[2]	r
Eastern Flying Training Command:						
44-1-I. Pre P-40	261	21.67	8.71	8.32	4.14	0.28
44-1-I. Post P-40	148	31.91	8.54	11.40	3.95	.27
44-1-J. Pre and Post P-40	498	27.28	10.79	9.91	5.24	.23
44-1-K. Pre and Post P-40	268	21.13	10.52	8.14	5.22	.13
z-Average	1175					.22
Western Flying Training Command:						
44-1-O. Pre P-40	87	24.02	13.89	7.89	4.73	.37
44-1-O. Post P-40	58	24.50	13.89	7.40	4.73	.38
45-A. Pre and Post P-40	119	25.64	15.92	10.14	6.08	.25
45-B. Pre and Post P-40	119	25.29	16.67	12.76	7.66	.34
z-Average	383					.33

[1] The subscript 1 indicates air-to-ground.
[2] The subscript 2 indicates air-to-air.

In order to obtain an estimation of the causal factors mutually shared by the two types of fixed gunnery, the correlations were corrected for attenuation. This correction was made in order to estimate the true relationship when the diluting effects of chance differences due to the unreliability of the scores had been removed. The corrected correlations were 0.36 for Eastern Flying Training Command (N=1175) and 0.54 for Western Flying Training Command (N=383). These indicated that approximately 13 percent and 29 percent, respectively, of the causal factors in the two fixed gunnery tasks were mutually shared. In other words, 87 percent and 71 percent of the factors in the two tasks were not shared. In the light of these findings, it seemed advisable to use the two gunnery scores as separate criteria.

Learning in Air-to-Air and Air-to-Ground Fixed Gunnery Training

The amount and course of learning in air-to-air and air-to-ground gunnery were determined in order to indicate how learning must be taken into account in using these scores as criteria and also to show whether or not the students were reaching the point of diminishing returns within the number of missions used in training.

The learning curve for each type of gunnery training was obtained from the mean score for each successive mission. It was assumed that a relatively accurate picture of the learning process could be obtained if enough cases were included so that other factors would be randomized. Table 11.4 presents the means by mission for each command and for all commands combined for air-to-ground fixed gunnery. Similar data are presented in table 11.5 for air-to-air fixed gunnery.

It can be seen from the data presented in tables 11.4 and 11.5 that considerable learning occurred in each type of gunnery, but that in air-to-ground gunnery there were definite unsystematic differences in the learning curves for successive groups of students. Most striking was the difference in the learning in air-to-ground gunnery

Table 11.4.—*Learning in air-to-ground fixed gunnery*

Class 44-1-G in the Western, 44-1-F in the Eastern, and 44-1-J in the Central Flying Training Command

Hundred rounds	Western		Eastern		Central				All combined	
					Pre P-40		Post P-40			
	N	*M* [1]	*N*	*M* [1]	*N*	*M* [1]	*N*	*M* [1]	*N*	*M* [1]
First	145	19.8	489	18.9	314	21.0	215	20.0	1,163	19.8
Second	145	22.1	485	22.2	312	21.8	228	32.3	1,170	24.0
Third	145	22.3	479	27.1	311	21.3	199	39.0	1,134	27.0
Fourth	145	26.3	473	29.9	308	23.1	175	39.9	1,101	29.1
Fifth	145	28.1	419	37.9	299	28.4	122	41.3	985	34.0
Sixth	145	26.5	284	39.7	256	33.7	50	41.4	735	35.1

[1] *M* is the average percentage of hits.

Table 11.5.—*Learning in air-to-air fixed gunnery*

Class 44-1-G, 44-1-H in the Eastern and 44-1-G, 44-1-J in the Western Flying Training Commands Combined

Hundred rounds	*N*	*M* [1]	*SD*	Hundred rounds	*N*	*M* [1]	*SD*
First	861	6.59	7.83	Eighth	458	11.64	9.85
Second	466	8.83	8.14	Ninth	867	12.68	10.44
Third	842	8.60	8.21	Tenth	464	13.25	11.47
Fourth	471	10.27	9.15	Eleventh	844	13.13	9.95
Fifth	869	10.29	9.17	Twelfth	461	14.06	11.68
Sixth	467	11.78	10.97	Thirteenth	822	14.47	11.27
Seventh	853	11.05	9.58				

[1] M is the average percentage of hits.

for the Pre and Post P-40 sections [64] of class 44-1-J in Central Flying Training Command. Since the number of students in each class was limited to roughly 200 to 300 in Central and Eastern Flying Training Command, not enough cases were available for the analysis of the variables affecting each mission. It is probable, however, that the most accurate approximation of the "true" gunnery learning curves is obtained by basing the curve on as large a sample as possible so that uncontrolled variables are randomized. Figures 11.1 and 11.2 present these curves for air-to-ground and air-to-air fixed gunnery, respectively.

As can be seen from these learning curves it was important, if either air-to-ground or air-to-air fixed gunnery scores were to be used as criteria, that they be based upon a specific number of missions flown. Defining a training mission as 100 rounds, each successive mission average was higher than the preceding one. To make scores for all students comparable they all had to be based upon the same number of missions.

[64] In the light of findings presented later in this chapter on the effect of P-40 transition training on fixed gunnery scores, it was unlikely that the differences found here could be due to one group having had and the other not having had the P-40 training.

Fig. 11.1

LEARNING CURVE FOR AIR-TO-GROUND FIXED GUNNERY

(Based on data in Table 11.4)

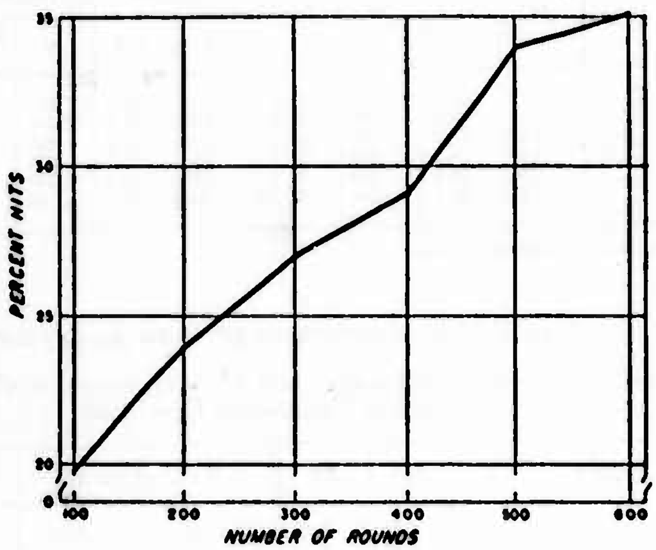
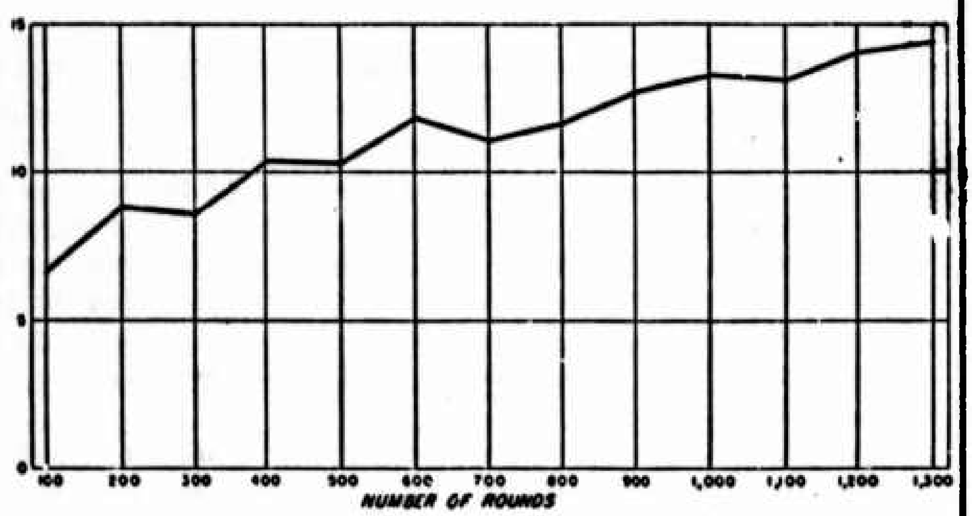


Fig. 11.2

LEARNING CURVE FOR AIR-TO-AIR FIXED GUNNERY

(Based on data in Table 11.5)



Since the learning curve was still rising approximately as steeply as ever at the end of training, it could be seen that the amount of firing given was only introductory in nature, with only a few, if any, students ever reaching their maximum proficiency.

A similar curve of learning, which was still rising steeply at the end of training, was secured for air-to-air firing on 523 students in class 43-F in Eastern Flying Training Command by Maj. N. E.

Miller, Lt. S. M. Roshal, and Lt. W. H. Angoff at Psychological Research Unit No. 1.

In a study made by the Research Section, First Air Force, of 19 instructor aerial gunnery learning curves, it was found that the increase in proficiency proceeds at a fairly uniform rate up through the sixtieth mission, where apparently a plateau was reached.

VARIABLES AFFECTING FIXED GUNNERY SCORES[11]

Variation in Training Procedures

The study of gunnery training procedures in the three flying training commands indicated considerable variation at both the advanced and transitional training levels.

In two of the three commands, gunnery training at the advanced training level consisted only of ground-school courses, while in the third command, two hours of gunnery missions were flown using gun cameras. The ground-school curriculum varied both in extensiveness and intensiveness from command to command and to a somewhat lesser extent between advanced schools in the same command.

Actual air-to-air and air-to-ground gunnery practice with live ammunition took place only during the transition phase of training. Here again the ground-school curriculum varied from command to command both in content and time devoted.

In two commands the students were sent from their base transition training school to a separate gunnery school for their fixed gunnery firing. In the third command the students and their transition training instructors were kept together for the fixed gunnery phase of training.

In one command every mission was flown with a gun camera in addition to live ammunition. These gun camera films were then assessed and criticized prior to the student's flying his next mission. In a second command only about one-half of the missions were flown with the gun camera while the third flew no such missions because of lack of assessing equipment.

In two of the commands students generally flew the same plane on successive missions, while in the third the students generally flew different planes.

Variations in penalties for "fouls" were found for both air-to-ground and air-to-air fixed gunnery from command to command.

It can be seen from this brief description that while fixed gunnery training followed the same general form in the three flying training commands, differences did exist in procedure, amount and type of training, and scoring. Differences also occurred in procedures and training for successive classes in the same command, and minor

[11] Work in this section was done by Maj. Olen L. Heathers, Lt. William E. Galt, Lt. Ralph M. Rust, and Sgt. John O. Okeson.

differences existed between schools in the same class and command. In the light of these differences, it seemed necessary to analyze data separately by command and class.

A more precise description of these variations in fixed gunnery training will be found in Appendix 11.

Effect of 10 Hours P-40 Training

One half of each class was given fixed gunnery training before receiving P-40 transition training; for the other half the order was reversed.

To determine the effect of this P-40 training, the data used in the determination of the reliability of gunnery scores were analyzed by school for differences between Pre and Post P-40 groups. The statistical reliability of the difference between mean scores was determined by use of the critical ratio. These comparisons indicated that the P-40 training itself did not affect the gunnery scores. Fourteen comparisons were made. In 7 cases differences were found of a magnitude such that there were less than 5 chances in 100 that a difference as large as this would arise by chance in comparable samples. Of these, four cases showed the Pre P-40 section to have the higher mean score while in the remaining three it was the Post P-40 section. Since the number of rounds fired was held constant in all cases, learning was also eliminated as the source of variance in mean scores. Therefore the variability of the differences indicates uncontrolled factors affecting fixed gunnery scores.

Tables A11.1 and A11.2, in the appendix, present these comparisons for air-to-ground and air-to-air fixed gunnery, respectively.

Data presented by Lt. Wilse B. Webb and Roland E. Johnston of Psychological Research Unit No. 1 also indicated no consistent trend in Pre and Post P-40 mean scores. In their study of air-to-air firing in Classes 44-1-B (N=487) and 44-1-D (N=647), it was found that the mean score for the Pre P-40 section was significantly higher (CR=7.97) than the Post P-40 section in Class 44-1-B and that the reverse situation held in Class 44-1-D (CR=5.30).

Turbulence, as Measured by Instructors' Ratings

The operation of uncontrolled variables had been suggested by the results of the studies of learning and of the effect of P-40 training. Both students and instructors agreed that turbulence was one such factor. They further agreed that air-to-ground firing was more affected by it than air-to-air, because turbulence was most severe at the low altitudes where air-to-ground gunnery was fired.

The degree of turbulence was rated during each mission, on a two-point scale, by the instructor flying with each flight of students in Class 44-1-J (Pre P-40), Central Flying Training Command. The two categories were: "rough," defined as turbulence judged to be greater than average, and "smooth" meaning less than average turbulence.

Mean scores by mission were compared for the two conditions by use of the critical ratio. These data are presented in the appendix in table A11.3. No reliable difference (CR=75 based on 1,198 cases), was found between the two conditions of flight. Therefore, it cannot be said that turbulence (as rated by the instructors) had any effect on air-to-ground gunnery scores. It should be noted that this does not eliminate turbulence as a variable in air-to-ground gunnery. It only indicates that either (1) turbulence may not have been an important factor or, as is more probable, (2) that the instructor ratings did not actually differentiate between the two categories of turbulence.

Time of Day

Gunnery instructors agreed that different times of day were not equally favorable for air-to-air and air-to-ground gunnery. The prevailing opinion was that, for air-to-ground gunnery, early morning and late afternoon were most generally favorable since the air was most likely to be smooth at these times. For air-to-air gunnery the best times for firing were considered to be near the middle of the day, when haze and glare were at a minimum. Turbulence was not judged to be an important factor in air-to-air gunnery since it was seldom encountered at the higher altitudes where such firing was done.

Mean scores by time of day, corrected for learning, were computed for both air-to-ground and air-to-air gunnery using time of take-off as the time of the mission. Tables A11.4 and A11.5, in the appendix, present these data together with a fuller description of the method of the studies.

From the air-to-ground data studied, it seemed that there were factors associated with time of day which affected these fixed-gunnery scores. However, it appeared that the effect varied from command to command and from class to class. It seemed probable, therefore, that the factors themselves differed from one location to another, at least in degree if not in kind. To determine to what extent this was true required a more detailed analysis than was then practicable.

From the air-to-air data studied it appeared that those fixed-gunnery scores were relatively independent of variation in conditions associated with time of day.

Wind Direction and Velocity

This variable affects only air-to-ground gunnery since, in air-to-air gunnery, both the target and the attacking plane are in the same moving air mass. As was stated in the job analysis of air-to-ground fixed gunnery, one problem is the correction for wind drift. Since all missions were not flown under the same wind conditions, the difficulty of the task of air-to-ground gunnery might vary from one mission to the next.

Air-to-ground fixed gunnery scores, corrected for learning, were tabulated by wind direction (headwind, crosswind, and tailwind) and velocity and the mean score for each direction and velocity computed. Figure 11.3 presents these data graphically for class 44-1-J in the Central Flying Training Command. A more detailed description of the methods used in this study is given in the appendix, and Table A11.6 presents the data separately for the Pre and Post P-40 sections of the class.

As figure 11.3 shows, there was a progressive decline in percent hits with increasing wind velocity from any direction. Also, it appeared that the net effect of tail and cross winds was similar. Missions flown in headwinds, at all velocities studied, yielded approximately 15 more hits per 100 rounds fired than did missions flown in either of the other two wind directions.

A comparison of the wind direction and velocity data for the two sections of the class shows the general points obtained from the combined curve held true. In each section, there was a decline in percent hits with increased wind velocity, and the scores obtained while flying into a headwind were higher than those obtained while flying in either of the other two wind directions; however, neither the rate of decline in hits with increasing wind velocity, nor the relative differences between wind directions remained constant. This indicated the interaction of uncontrolled variables masking the true picture of the effect of the wind.

The true trend was probably most closely approached in the combined data since the larger number of cases increased the likelihood of other variables being more completely randomized.

Correction of Air-to-Ground Scores for the Effect of Wind Direction and Velocity.—Although it was true that the general trend of effects of wind direction and velocity were similar in the two sections of class 44-1-J, there were considerable variations in absolute values. This fact made it seem less likely that any correction for wind conditions would materially improve the reliability of the gunnery scores. However, since this variable was the best defined and most accurately measured of the several known variables, it was decided to attempt the correction of the gunnery scores for the effect of that factor. Should no improvement in the reliability of the scores be obtained with this correction, then certainly no improvement could be expected as a result of correcting for the less well defined and measured variables.

Using the data for the Pre P-40 secfion of class 44-1-J, a set of conversion tables was made to correct individual scores obtained while flying in any given wind direction and velocity to the scores that would be expected had the mission been flown in the optimum headwind. These conversions were equal to the actual difference in percent hits between the expected score for any given velocity and direction of wind and the optimum headwind.

The scores made by individual students during the Post P-40 section of this class were corrected for the effect of wind direction and velocity by using the conversion tables based on the Pre P-40 section. If wind direction and velocity were stable variables within each day of missions in air-to-ground fixed gunnery, the reliability of the raw scores should be spuriously high because of the correlated factor. Correction to eliminate effect of wind would therefore yield a lower reliability coefficient. Such proved to be the case. The reliability of the corrected scores was only 0.47, while the reliability of the raw scores was 0.53.[66]

Since no greater change in results could be expected in correcting for other such correlated variables than was found above, it seemed impractical to attempt to reduce the class-to-class variations by computing corrections for them. All data, which had been studied, indicated that the variables affecting fixed-gunnery performance varied from place to place and probably from season to season. Therefore, the best control of them was to analyze the data separately by class and command. The most effective way of achieving this control was to convert the scores for each class and command into standard scores and to use these standard scores in further analysis.

GENERAL CONCLUSIONS REGARDING THE PRECAUTIONS TO BE FOLLOWED IN HANDLING OF FIXED GUNNERY SCORES[67]

Reliability

Provided a sufficient number of rounds is involved, the reliability of fixed-gunnery scores was satisfactorily high. Anyone using this type of Training Command fixed-gunnery scores in a study could be reasonably certain that, for any class involved, the reliability of 400 rounds of air-to-ground firing would be approximately 0.59, and that of 1,200 rounds of air-to-air firing would be approximately 0.63. Since the reliability varied with the number of rounds in approximately the manner predicted by the Spearman-Brown formula, it was important to consider the number of rounds on which the reliability was based in evaluating results.

Learning

The learning curves for both air-to-ground and air-to-air fixed gunnery are so steep, with no indication of a plateau being reached within the limits of the training outlined here, that it is important

[66] Table A11.7, in the appendix, presents these data more completely. In connection with this table will be found a more complete description of the method of making these corrections.

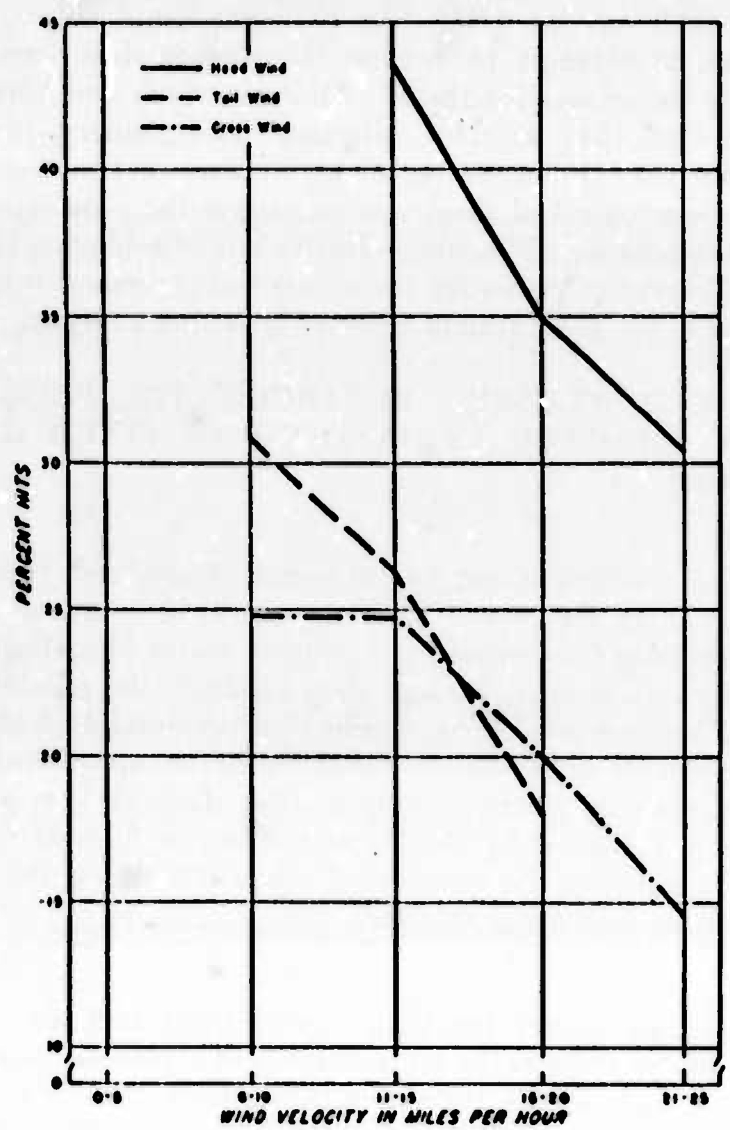
[67] Work in this section was done by Maj. Glen L. Heathers, Lt. William E. Galt, Lt. Ralph M. Rust, and Sgt. John O. Gleason.

to control this factor in the use of gunnery data. In order to control learning, data for each student to be used in a study must be based on the same number of rounds fired during the same stage of learning.

Fig. II.3

RELATION BETWEEN WIND DIRECTION AND VELOCITY AND AIR-TO-GROUND FIXED GUNNERY SCORES

(Based on data in Table AII.6)



P-40 Training

It appears reasonably certain that the 10 hours of P-40 Transition training received by one-half of each class prior to gunnery training had no appreciable effect on the gunnery scores obtained. Therefore, the Pre and Post P–40 sections of any given class can be combined when other factors remain constant. Since the two subsections of the class are subjected to somewhat different conditions, it is well to analyze them separately when a sufficient number of cases are available in each.

Other Variables

The effects of such variables as turbulence, time of day, weather, etc., either were not accurately measurable or were so variable in effect from one class or location to another that it was impossible to make any simple correction for their effects. Therefore, it was advisable to handle data including more than one class or command in one of the two following manners: either the computations should be carried out separately for each group (class or command) and the results combined, or scores of individual students should be converted to normalized scores by class or command before analyzing the data.

Another variable which was present in the gunnery training situation, but which was not subject to accurate measurement was the plane flown by the student. A preliminary analysis, not reported in detail here, showed that there was a relationship between the plane flown and the gunnery scores obtained by the students flying it. It was difficult to work out any systematic correction for this factor because the effect of the same plane varied from time to time depending upon (1) the state of maintenance, and (2) the harmonization of the sight and gun. Furthermore, the number indicating the plane might change either during or between classes, since each field had only a limited number of plane numbers assigned to it and replacement planes were given the same number as one previously assigned to the field. In any study of fixed gunnery, it should be stated whether or not each student flies the same or different ships on successive missions in order to indicate whether or not possible plane differences have added to or subtracted from the apparent reliability of the data.

FURTHER EVALUATION OF FIXED GUNNERY SCORES [66]

Ability of Instructor Ratings of Pilot Proficiency in Advanced Training to Predict Subsequent Fixed Gunnery Scores

In order to evaluate fixed gunnery scores more accurately as a possible criterion of pilot proficiency, instructors' ratings of students' flying proficiency were correlated with the students' subsequent gunnery scores.

During the final week of advanced single-engine training for class 45–B at Aloe Army Air Field, each of the 174 students were rated in flying proficiency by his flying instructor. The instructors were asked

[66] Work for this section was done by Lt. William E. Galt and Sgt. John O. Gleason.

to rate their students on a quartile rating basis, using all of their previous students as the standard of judgment. These ratings were then assigned numerical values of from 1 to 4, with 4 including the top quarter. Of the 174 students rated, 58 were assigned to and completed fixed gunnery training. The larger proportion of these students had been rated in the best two quarters of flying ability. None of the students assigned to gunnery training was rated in the poorest quarter. Product-moment correlation coefficients were computed between instructor ratings at the Advanced school and air-to-air and air-to-ground fixed gunnery scores for the 58 students assigned to gunnery training. The air-to-ground gunnery score used for each student was his average percent hits for four 100-round missions and the air-to-air score used was the average percent hits for twelve 100-round missions.

The instructor ratings correlated 0.35 [69] with both the air-to-ground scores and the air-to-air scores. In evaluating these results it is necessary to remember that the group of students used in this study represent a very restricted range of ability; a restriction made even more severe by the final selection in Advanced schools of only those men judged to be the best potential gunnery students. The correlations, while not high, gave added support to the professional judgment of gunnery-training personnel that ability in fixed gunnery was a good criterion of pilot proficiency.

Ability of Fixed Gunnery Scores in the Training Command to Predict Those in the First Fighter Command

The air-to-air and air-to-ground fixed gunnery scores, separately and combined, in the Training Command were correlated with the corresponding fixed gunnery scores in the First Fighter Command. This was done to determine the degree of relationship between the task of fixed gunnery in the slower AT-6 plane used in the Training Command and the faster P-47 used in the Fighter Command.

The sample of Training Command scores came from the same 1,308 students described in the following section of this chapter. Of these 1,308 students, only 322 were eventually assigned to the First Fighter Command and fixed gunnery records were available for only 179 of these. The Training Command gunnery scores used in this study were described in the preceding section. The Fighter Command scores used were standard scores based on the mean and sigma of each class and school. They were supplied by the Research Section, Office of the Surgeon, First Air Force.

Table 11.6 presents the intercorrelations of the air-to-air and air-to-ground fixed gunnery scores in the two stages of training. The correlation between the combined scores is 0.38.[70]

[69] Table A11.8, in the appendix, present these data more fully.

[70] Based on 179 cases, mean and sigma of the Training Command data are 4.7 and 1.3 and of the Fighter Command 51.9 and 8.0, respectively.

It can be seen that fixed gunnery scores in the slower, lighter planes used in the Training Command did predict performance in the faster, heavier ones used in the First Fighter Command, but that the prediction was far from perfect even when corrected for the attenuation produced by the fact that the scores in each command were not perfectly reliable. The relatively low correlations indicate that the tasks were quite a bit different at these two levels of training.

TABLE 11.6.—*Correlation between fixed gunnery scores in the training command and those in the First Fighter Command based on 179 students*

	First Fighter Command [2]			
	Air-to-ground	Air-to-air	M	SD
TRAINING COMMAND [1]				
Air-to-ground	0.12 (0.21) [3]	0.29 (0.49)	4.91	2.00
Air-to-air	.18 (0.31)	.24 (0.43)	5.06	1.75
M	50.87	51.72		
SD	10.00	11.00		

[1] Single-digit normalized scores, single-engine transition training in the AT-6.
[2] Standard scores, single-engine tactical training in the P-47.
[3] Correlations in parentheses are corrected for attenuation in order to estimate the degree of relationship which would be expected if perfectly reliable measures could be secured in each command.

It will be remembered that the two types of gunnery, air-to-air and air-to-ground, correlated with each other only 0.22 in Eastern Flying Training Command, which was the source of all of the students sent to the First Fighter Command. Even after correction for attenuation, this correlation was only 0.36. This low value indicates that the skills involved in these two types of gunnery were considerably different. From this it might be expected that the correlation between the same type of gunnery in the two commands would be higher than that between different types: For example, one would expect the correlation between Training Command air-to-air and Fighter Command air-to-air gunnery to be higher than that between Training Command air-to-air and Fighter Command air-to-ground gunnery. Table 11.6 shows that this was not the case; the correlations between the different types are as high as those between the same types. None of the elements of skill which were general enough to apply to both the slower, lighter planes in the Training Command and the faster, heavier ones in the Fighter Command were specific to either air-to-air or air-to-ground gunnery.

Ability of Classification Test Scores to Predict Fixed Gunnery Performance [71]

The individual classification tests were validated against gunnery scores to determine the extent to which these tests, administered prior to any pilot training, could predict fixed gunnery ability approximately one year later.

[71] Work in this section was done by Sgt. John O. Olsson and S/Sgt. Charles P. Gershenson; classification test scores on punch cards were supplied by Statistical Unit, Psychological Section, Office of the Surgeon, Hq. AAF Training Command.

Fixed gunnery scores were obtained for 1,308 students in classes 44-1-I, 44-1-J, 44-1-K, and 45-A in the Eastern Flying Training Command. Two scores were obtained for each student: (1) The average percent hits for the first four 100-round air-to-ground missions, and (2) the average percent hits for the first twelve 100-round air-to-air missions. These scores were normalized and transformed into single-digit scores by the following method. Distributions of the air-to-ground and air-to-air fixed gunnery scores were made separately by class, and broken into nine class intervals containing respectively 4, 7, 12, 17, 20, 17, 12, 7, and 4 percent of the students' scores. A single digit score was substituted for the actual percent hits in each class interval, 1 being assigned to the lowest 4 percent, 2 to the next higher 7 percent, and so on until 9 was assigned to the top 4 percent. The single-digited, normalized, air-to-air and air-to-ground scores were averaged to yield a third score reflecting the student's over-all fixed gunnery ability. All fractions in these averages were 0.5 and were rounded to the next lower whole number. Thus, the combined score was also a single-digit number.

These three scores were used separately as criteria of gunnery proficiency and the classification tests validated against them. Table 11.7 presents the correlations of the tests in classification batteries 3 and 4, and of the three stanines with the three criteria of gunnery proficiency.

It can be seen that each of the three stanines, determined prior to the beginning of flying training, predict gunnery performance at the end of training. The pilot stanine yields the best prediction. While the correlation is relatively low, it is of a size which would be expected to occur by chance less than 1 time in 10,000. In evaluating these correlations, it should be noted that the range of talent is markedly restricted by the time the students reach gunnery training. No corrections for restriction of range were made.

The fact of having had flying experience prior to starting military flying training, though highly related to success in primary flight training, appears to have no relation to proficiency at this advanced stage of training.

Those tests which best predict air-to-ground fixed gunnery scores also seem to be the ones which best predict the air-to-air gunnery scores, within the limits of expected sampling variation. In general, the tests which best predict pass-fail in primary flight training also yield the best prediction of the combined fixed gunnery score. Two possible exceptions to this generalization are the general information and rudder-control tests, which seem to be less related to gunnery performance than to success in primary-flight school.

The main results of this study are confirmed by those of such other investigations as have been made in this area. As is shown in table 11.7, Psychological Research Unit No. 1 secured a correlation of 0.18

Table 11.7.—*Correlations of all classification tests and stanines with three criteria of gunnery proficiency*

Number of Cases: Battery 3, 319; Battery 4, 681; Total 1,000

	Gunnery scores				Primary				
	Ground	Air	Combined	Air [2]	Pass-fail [1]		N [4]	M [4]	SD [4]
1. Bombardier stanine	0.12	0.13	0.16	......	0.21	[3]0.36	1000	6.50	1.69
2. Navigator stanine	.10	.12	.15	......	.21	.34	1000	6.88	1.73
3. Pilot stanine	.13	.16	.19	0.18	.35	.48	1000	7.01	1.36
4. Pilot credit	−.02	−.02	−.03	......	.23	......	1000	.40	1.00
5. Rotary pursuit, CP410B	.05	.12	.12	......	.11	......	1000	63.83	10.16
6. Two-hand coordination, CM101A	.10	.08	.12	......	.11	......	1000	65.25	9.22
7. Complex coordination, CM701A	.13	.07	.14	.17	.17	......	1000	66.72	9.73
8. Rudder control, CM120B	.06	.05	.07	......	.22	......	1000	64.92	8.17
9. Discrim. reaction time, CP611D	.07	.09	.10	......	.18	......	1000	63.73	7.40
10. Finger dexterity, CM116A	.07	.02	.07	.11	.04	......	1000	62.95	9.35
11. Biographical data—(N), CE602D	.05	.02	.04	......	.00	......	1000	22.80	2.94
12. Biographical data—(P), CE602D	.06	.07	.10	......	.10	......	1000	30.88	6.13
13. Spatial orientation, I, CP501B	.03	.11	.10	.12	.07	......	1000	30.00	6.60
14. Spatial orientation II, CP503B	.07	.11	.12	.13	.10	......	1000	22.14	6.46
15. Dial and table reading, CP621-22A	.07	.12	.13	......	.23	......	1000	35.05	8.32
16. Mathematics A, CI702F	.03	.09	.08	......	.11	......	1000	18.11	12.84
17. Mathematics B, CI206C	.08	.05	.08	......	.10	......	1000	18.13	9.37
18. Reading comprehension, CI614H	.07	.05	.08	......	.11	......	681	23.90	14.05
19. Mechanical principles, CI903B	.12	.10	.14	......	.15	......	681	36.56	8.96
20. Instrument comprehension I, CI-615B	−.07	−.15	−.14	......	−.22	......	681	9.32	2.90
21. Instrument comprehension II, CI-616B	.12	.13	.16	......	.26	......	681	37.86	10.27
22. General information, CE505E	.06	.03	.07	......	.26	......	681	63.78	12.29
23. Reading comprehension, CI614G	−.09	.11	.02	......	.18	......	319	24.60	12.27
24. Mechanical principles, CI903A	−.02	.14	.07	......	.34	......	319	63.66	17.21
25. Numerical operations (P), CI702B	.02	.00	.04	......	.01	......	319	37.73	11.24
26. Numerical operations (B), CI702B	.01	.00	.03	......	.02	......	319	32.92	10.90
27. Speed of identification, CP610A	.00	.09	.03	......	.12	......	319	35.26	7.10
28. General information—(N), CE505D	−.03	.08	.04	......	.07	......	319	20.75	6.00
29. General information—(P), CE505D	−.11	.05	−.03	......	.23	......	319	36.88	6.10

[1] Primary pass-fail validity coefficients for variables 1 through 22 based on 6,366 students in classes 45-O through 45-F, and for tests 23 through 29 based on 4,213 students in class 44-E. Correlations computed by Statistical Unit, Psychological Section, Office of the Surgeon, Hq. AAF Training Command.

[2] Correlations with air-to-air scores for 523 students from class 43-F of Eastern Flying Training Command reported by Maj. N. E. Miller, Lt. S. M. Roshal, and Lt. W. H. Angoff from Psychological Research Unit No. 1.

[3] Corrected for restriction in range produced by fact that students with low stanines were not sent to pilot training.

[4] Referring to stanines and test scores of students involved in correlations with three criteria of gunnery proficiency.

between the average percent hits on 1,800 rounds of air-to-air fixed gunnery and the pilot stanine for 523 students from an earlier class, 43-F. Since adding credit to the stanine for having had flying experience before starting military flying training did not improve the prediction, it may be inferred that such experience was no help by the time the students reached fixed gunnery training. The individual tests which were validated also yielded coefficients similar to those secured by the Pilot Project. Finally, the Research Section, Office of the Surgeon, Hq. Second Air Force, has correlated the pilot stanine of 427 students with the average percent of hits on 14 air-to-air missions of at least 100 rounds each. The correlation at this later level of tactical training was 0.14, again in the same range. Furthermore, adding credit to the stanine for flying experience before starting military training did not improve the prediction, it yielded a correlation of only 0.13, again confirming earlier results.

PROBLEM FOR FUTURE RESEARCH

The most important problem for further research is believed to be determining how much different amounts of fixed gunnery training on the lower-powered AT-6 will help the student in the higher-powered planes used in combat. The learning curves presented in the first part of this chapter were still going up about as steeply as ever at the end of the amount of practice given in the Training Command. Apparently the students did not approach their limit of learning. Although it is quite certain that with more practice the students would have become considerably more accurate in firing with the AT-6, it is by no means certain that this would have helped them in learning to fire with higher powered tactical planes. In fact, it is even possible that the training on the slower and lighter AT-6 taught the students habits which interfered with gunnery in the faster, heavier tactical planes.

The correlations between the gunnery scores in the Training Command and those in the First Fighter Command are low enough (even when corrected for the attenuation produced by the unreliability of the measures) to indicate that the two types of firing are far from identical tasks. Laboratory experiments and practical experience indicate that it is unsafe to make *a priori* assumptions concerning the type of transfer which will occur in complex situations of this kind. The only way to determine the most efficient distribution of fixed gunnery training is by an experimental investigation of the amount of transfer to tactical planes which is produced by different amounts of training in the AT-6. Some students should be sent to the higher powered tactical ships with no gunnery training in the AT-6, others should be sent with considerably more training than is now given, and still others should be given various intermediate amounts of practice in the Training Command. The groups receiving different amounts of fixed gunnery training in the AT-6 should not be selected by chance; the students in them should be matched on the basis of their flying proficiency in advanced training. When these students are sent to the Air Forces an equal number should be assigned from each group to each instructor or squadron. Then the curves of learning in the Air Forces should be examined for each group to discover differences in initial performance and whether or not these differences persist throughout tactical training.

SUMMARY

These studies were undertaken to determine the suitability of fixed gunnery scores as objective measures of pilot proficiency at the single-engine transition training level.

It was found from the job analysis of air-to-ground and air-to-air

fixed gunnery, that considerable flying skill was required to fly the plane in such a way that the bullets from its guns would hit the target.

Since these tasks did include important elements of flying skill, the next step in evaluating fixed gunnery scores as an objective measure was to determine their reliability and other statistical characteristics. The following facts were determined:

(*a*) The odd-even mission reliability was 0.63 (N=1064) for 1,200 rounds fired in air-to-air fixed gunnery and 0.59 (N=932) for 400 rounds fired in air-to-ground fixed gunnery. These reliability coefficients were satisfactorily high to warrant the use of gunnery scores as criterion data. The reliability of these scores varied with the number of rounds on which they were based, and the Spearman-Brown formula was found to give accurate predictions of this relationship.

(*b*) The intercorrelation of air-to-air and air-to-ground fixed gunnery scores was relatively low (0.22 based on 1,175 cases in the Eastern Flying Training Command, 0.33 based on 383 cases in the Western Flying Training Command) so that it seemed advisable to use the two types of scores as separate criteria.

(*c*) The learning curves for both air-to-air and air-to-ground fixed gunnery were still rising rapidly at the end of training. This indicated that the limit of learning had not been reached during the period allotted to training.

Further study of fixed gunnery data indicated several possible uncontrolled variables including: variations in training procedures, turbulence, time of day, plane flown, and wind direction and velocity. Since the exact amount of effect produced by each of these variables was not easily determined, the students' individual scores could not be easily corrected for their effects. It was therefore necessary to control for these variables by analyzing all data separately by class and command.

Recognizing the above-listed characteristics of fixed gunnery scores and taking the proper precautions for their control, it was found that subjective ratings of flying ability, made by instructors at the advanced stage of training, correlated 0.35 with both air-to-air and air-to-ground gunnery scores in the transition stage. Furthermore, the combined air-to-air and air-to-ground fixed gunnery score in the Training Command correlated 0.38 with a similar score obtained in the First Fighter Command. The correlation between different types of gunnery in the two commands (i. e., air-to-ground in Training Command *v.* air-to-air in Fighter Command) was as high as the correlation between the same type. This indicated that none of the elements of skill which were general enough to apply to both the slower, lighter planes in the Training Command and the faster, heavier ones in the Air Force were specific to either air-to-air or air-to-ground gunnery.

Finally, it was found that those tests in the classification battery which had the highest validity as measured by the pass-fail criterion in primary training, also tended to have the highest correlations with fixed gunnery scores in transition training.

The Pilot Project recommended that an experiment be conducted to determine how much different amounts of fixed gunnery training on the lower powered AT-6 will help students in the higher powered planes used in combat.

CHAPTER TWELVE

Printed Tests of Flying Information

S/Sgt. Irving Robbins and Sgt. Robert Levine[73]

INTRODUCTION

Reports of Aircrew Evaluation and Research Detachments sent overseas by the Office of the Air Surgeon and information gained from interviews with experts recently returned from combat and assigned to organizations such as the AAF Tactical Air Center and the Proving Ground Command emphasized the fact that the task of a successful combat pilot involved much more than the motor and perceptual skills required to handle a plane. The pilot had to have a thorough, practical knowledge of the aerodynamics and flying characteristics of his airplane in order to get maximum performance under combat conditions. Since he was responsible for seeing that the crew maintained and used the plane's equipment properly and since enemy fire created sudden emergencies in the air, he had to have a comprehensive understanding of what his plane and its equipment were designed to do and how they were affected by unusual conditions and emergency situations.

One of the most important requirements for airplane commanders and lead-crew pilots was a wide background of practical knowledge to serve as a basis for good judgment in making decisions bearing on the success of the mission and the safety of the personnel involved. For example, the pilot had to balance the probabilities of icing and headwinds at certain altitudes against the true air speed and rate of fuel consumption. Though he received advice from his navigator and engineer, the final decision rested with him. Having the proper information may not insure sound judgment, but without a certain minimum of knowledge good judgment is impossible.

The printed tests of flying information were designed for two primary purposes:

(1) To *select* pilots for special assignments, such as lead-crew or instructor training, on the basis of their achievement in advanced stages of training.

(2) To *diagnose* the strengths and weaknesses of the pilots as a check of the effectiveness of the curriculum and as a possible means of deciding what refresher courses they needed.

[73] Work on the test-scoring machine was supervised by Lt. William E. Harris.

Most effort was devoted to the development of these tests for selection purposes. The scores from these tests, of course, were only one of the factors to be taken into account in selecting pilots on the basis of their achievement in advanced stages of training. Other factors, such as flying skill, interest, and attitude had also to be considered. The scope of testing was limited to the measurement of those aspects of flying information which were essential to the good pilot and which could be thoroughly measured with greater economy on the ground than in the air. The tests were designed to measure the type of information needed by the pilots, whether or not it was included in the curriculum.

GENERAL APPROACH AND TECHNIQUES OF TEST CONSTRUCTION

Selection of Areas To Be Tested

The purposes of testing determined the areas to be sampled. The contents of these areas were based upon the opinions of experts, technical manuals, the curricula of pilot schools, and accident statistics. Since areas of flying information were well defined and categorized in the Army Air Forces, it was a relatively easy process to select suitable areas for testing. The main areas selected were analysis of maneuvers, navigation, aeroequipment, instrument flying, weather, and the use of oxygen equipment.

Writing of Test Items and Consultation With Experts

Items for the various tests were based on the latest authoritative regulations, manuals, and references in use in the Army Air Forces. Experts were consulted to insure that the test items were functional, important, and not likely to become obsolete as frequent technical changes were introduced. Consideration was also given to the probable difficulty and clarity of the test items. Persons in responsible positions were consulted as direct sources for test items. In some cases, old test forms used in ground schools and on flight lines were scanned for ideas and items. All items were rewritten in multiple-choice form in order to reduce the effects of guessing and to increase test reliability. Catch questions or ambiguous items were weeded out. The final criterion for the inclusion of a test item was, "Would the pilot be likely to get into trouble if he didn't have this information?"

A 4-point rating scale was employed with expert pilots to evaluate the importance of each item. The uncorrected correlation between the average rating of one group of five experts with another of seven, was 0.50. This permitted the selection of those items on which there was substantial agreement among experts.

This technique was employed in the development of three of the subsections of the test (instrument flying, weather, and personal equipment), since groups of experts were available only in these areas.

Because the experts did not agree on the relative importance of different areas as well as they did on items within an area, this technique could not be used in weighting the various subsections of the test. This difficulty was not a reflection upon the technique, but rather symptomatic of the subject matter. Depending upon the situation, normal or emergency, the importance of one area may be temporarily increased over that of all others. Therefore, the inclusion of a particular test item was based not on the *a priori* significance of an area, but rather on the importance of the specific point of information.

When available, evaluation boards performed technical verification of test items and official coordination for the Army Air Forces. The AAF Instrument Flying Standardization Board at Bryan Army Air Field, Bryan, Tex., was consulted, and other groups offered expert help in the development of the printed tests. The organizations represented included:

(1) Basic and twin-engine groups, AAF Central Instructors School, Randolph Field, Tex.

(2) Ground Training Technical Advisory Department, Randolph Field, Tex.

(3) Ground-school personnel, Lockbourne Army Air Field (Central Instructors School for B-17 pilots).

(4) Ground-school personnel, Smyrna Army Air Field (Central Instructors School for B-24 pilots).

(5) Returned combat pilots in the Flight Test Section and Staff of the Physiological Section, Proving Ground Command, Eglin Field, Fla.

(6) Multiengine squadron of returned combat pilots, AAF School of Applied Tactics, Orlando, Fla.

(7) Physiology Department, School of Aviation Medicine, Randolph Field, Tex.

(8) Station Standardization Board for instrument flying, ground-school instructors who were rated pilots, and the AAF Instrument Flying Standardization Board, Bryan Army Air Field, Bryan, Tex.

(9) Psychological Section, Office of the Surgeon, Hq. AAF Training Command, Fort Worth, Tex.

(10) A-3 Division, Hq. AAF Central Flying Training Command, Randolph Field, Tex.

Trial Administration and Statistical Analysis

After a series of test items had been examined for technical correctness and clarity of wording, they were mimeographed in test booklets and administered to samples drawn from populations whose achievement level was comparable to the level of pilot performance ultimately to be tested. Generally, similar analyses were made for all tests, though the size of the population and practical considerations sometimes dictated a less complete analysis of data. When a test was fully analyzed, the following calculations were made:

(1) Means and standard deviations of subtests and total score.

(2) Test reliability (odd even technique using Rulon formula).[73]

(3) Subtest intercorrelations (Pearson product-moment).

(4) Determination of difficulty level of each item.

(5) Item analysis of internal consistency according to the upper-lower 27 percent technique.[74]

(6) Correlation with grades in ground school and flight line.

Revision on the Basis of Empirical Findings

The results of the preliminary administration provided a basis for eliminating those items which dealt with controversial material, were poorly worded, ambiguous, and either too easy or too difficult. The first point to be considered was the level of difficulty of the question and the attractiveness of each of its alternative choices. An item answered correctly by only 10 percent or by more than 90 percent of the students was considered undesirable because it did not add much to the total test reliability. For the same reason any alternate choice which failed to attract responses was revised. Those incorrect alternate choices which received too many responses were also revised.

The measure of internal consistency (agreement of each item with the total score on the test) was used chiefly to spot possible ambiguities in the questions. If the 27 percent of the students who got the highest total scores on the test did not do better on a given question than the 27 percent who got the lowest total scores on the test, that question was scrutinized for content and ambiguity. If the item looked doubtful, it was revised or discarded; if not, it was kept.

Internal consistency was *not* used as a decisive criterion because there is no logical reason for expecting the knowledge of different types of information in this area to be highly correlated. An item which is not related to the total score on the rest of the test, but is well worded and covers an important point, adds to the comprehensiveness of the test by tapping an independent area of information.

The general procedure was to show specialists in each area of flying the distributions from both the level of difficulty and the internal consistency analyses, to discuss these data, and encourage them to suggest revisions.

Administration of Revised Test

The revised test was then administered to a second sample similar to the first so that results would be comparable. The results of this second administration were then used for further analysis.

---

[73] Rulon, Phillip J., A Simplified Procedure for Determining the Reliability of a Test by Split-Halves, *Harvard Educational Review*, 1939, 9: 99-103.

[74] Flanagan, John C., Item Analysis by Test Scoring Machine Graphic Item Counter, *Proceedings of the Educational Research Forum*, International Business Machines Corporation, August 1940, pp. 80-94.

Statistical Analysis and Further Revision

A set of complete statistics was obtained on the second administration of the test in order to make a final evaluation of the test before printing for actual use.

Validation

Because so much of the information tested was of a type which would be useful only in tactical operations or combat and not during the earlier stages of training, there did not seem to be much point in validating the tests against the criteria available in the somewhat protected environment of the Training Command. The tests were put into use after experimental administration had demonstrated that they yielded reliable measures of those points of information which experts considered to be relevant.

This procedure is less safe than validation against a good criterion, where one is available. Plans were formulated to follow the crew members tested in the Training Command through subsequent training in the Continental Air Forces and performance in the combat theaters with the cooperation of the psychological organizations in those places. Such validation, though difficult, would have been a better proof of the tests but the end of the war occurred before it could be carried out. Even had it been possible to make this follow-up, two difficulties would have been encountered: the measures of combat performance would have been relatively unreliable, and the specific points of information which were most important at the time when the study was made might have become rapidly obsolete as techniques were rapidly improved.

THE PRACTICAL USE AND STATISTICAL ANALYSIS OF THE TESTS CONSTRUCTED

In all, work was done on five tests. Of these, two, the Combat Flying Judgment Test and the B-17 Check-Out Test, remained in undeveloped stages and three were completed. One of the three tests was used for selection purposes, a second on a try-out basis at AAF Redistribution stations, and the third was used throughout the Training Command for examining instrument pilots. Since these tests affected Army Air Forces' practices, a description of the content, use, and statistical data on each is presented.

Pilot Information Test, Form 3 (Multi-Engine), Booklets I and II

The Pilot Information Test, Form 3, was developed to aid in the selection of potential lead-crew pilots in the Very Heavy Bombardment program. The score on this test in combination with previous pilot experience, pilot stanine, and proficiency grades was weighted in selecting the best pilots to be grouped with the best navigators, bombardiers, and flight engineers for training as potential lead crews.

The test was composed of 276 multiple-choice items. The areas measured were analysis of maneuvers, navigation, aeroequipment, instrument flying, weather, and use of oxygen equipment. Illustrative items follow:

In order to hold a constant rate of turn when air speed is building up, angle of bank should be—

A. Increased.
B. Maintained.
C. Reduced.

If the regular static line as well as the alternate source are blocked, static pressure can be obtained—

A. From alternate vacuum source.
B. By breaking the glass of the instrument.
C. From alternate source venturi.
D. By switching to an engine-driven pump.

Variation is 6° E. Wind is from 50°. In order to make good a true course of 90°, one should—

A. Crab to the left.
B. Crab to the right.
C. Fly a heading of 84°.
D. Fly a heading of 96°.

Total administration time, which included directions and actual testing time was 3 hours; of this, actual testing time for booklet I was 70 minutes, for booklet II was 85 minutes. This time allowance permitted practically all men to finish the test. Since the correlation between test score and test time was 0.12, it was decided not to make this a speed test. The test score was the total number of correct answers.

The basic statistics of the Pilot Information Test, Form 3, are presented for 371 randomly selected test papers supplied by Psychological Research Project (Combat Crew) at Lincoln, Nebr. The average of the levels of difficulty (percent answering correctly) for the items in this test is 68 percent.

TABLE 12.1.—*Intercorrelations and reliabilities of the subsections of the pilot information test, form 3 (N=371)*

Name of test	1	2	3	4	5	6	r_{11} [1]	Number items	M	S. D.
1. Flying	......	0.38	0.41	0.41	0.34	0.20	0.49	50	32.01	3.68
2. Navigation	0.38	......	.34	.49	.51	.24	.64	40	25.76	4.18
3. Aero-equipment	.41	.34	......	.45	.42	.41	.63	46	33.11	3.99
4. Instrument flying	.41	.49	.45	......	.48	.35	.70	57	39.78	5.65
5. Weather	.34	.51	.42	.48	......	.30	.63	53	39.26	5.03
6. Personal equipment	.20	.24	.41	.35	.30	......	.35	30	18.41	2.80
Total	......	......	......	......	......	......	.85	276	188.33	18.06

[1] Computed by Hoyt method which includes Spearman-Brown correction.

The data in table 12.1 show that the test provided an adequate distribution of scores for use in a selection battery. The fairly high reliability of 0.85 for the total score means that those who scored high on the test did not do so because of chance factors like guessing or luck; therefore a pilot's score was probably based on what he knew about the areas being tested.

From table 12.1 it can be seen that the intercorrelations among various subsections of the test are relatively low. In table 12.2 these

TABLE 12.2.—*Independence of the subsections of pilot information test, Form S*

Intercorrelations Corrected for Attenuation; Multiple Correlations and Coefficients of Multiple Non-Determination Based Upon the Corrected Intercorrelations

Item of test	1	2	3	4	5	6	*R*	K^2
1. Flying	......	0.68	0.73	0.69	0.61	0.49	0.91	0.18
2. Navigation	0.68	......	.53	.73	.80	.51	.87	.24
3. Aero-equipment	.73	.53	......	.67	.67	.87	.96	.09
4. Instrument flying	69	.73	.67	......	.72	.71	.87	.24
5. Weather	.61	.80	.67	.72	......	.64	.86	.26
6. Personal equipment	.49	.51	.87	.71	.64	......	.94	.12

R is the multiple correlation of the subsection with all of the other subsections, i. e., R1.23456, R2.13456, etc.

K^2 is the coefficient of multiple nondetermination ($1-R^2$) and is a conservative estimate of the percentage of variance in a perfectly reliable measure of the factors involved in each subsection which could not be accounted for by perfectly reliable measures of the factors involved in all of the other subsections. Thus it is an index of the degree to which each area of information is independent of the others.

intercorrelations have been corrected for attentuation in order to estimate how high they would be if each type of information were measured by a perfectly reliable test. The fact that the correlations are still below 1.00 indicates that no two tests measure exactly the same thing. The question of how well perfectly reliable tests in five of the areas would completely account for the sixth area was answered by computing coefficients of multiple nondetermination based upon multiple correlations. These are presented in the last column of table 12.2. It can be seen that the most independent areas are Navigation, Instrument Flying, and Weather. Approximately one quarter of the variance in each of these areas would be left unaccounted for by perfectly reliable tests in the other five areas. This is a conservative estimate since multiple correlations tend to shrink upon cross-application of beta weights, leaving somewhat more variance unaccounted for.

Biographical data accompanying the test papers supplied by Psychological Research Project (Combat Crew) were analyzed in a number of ways at Psychological Research Project (Pilot).

The scores were tabulated separately for pilots and copilots and for those who had served as instructors and those who had not been instructors. The results are presented in table 12.3. There was no significant difference between the test scores of those assigned as pilots or as copilots.

TABLE 12.3.—*A comparison of the scores on the pilot information test (Form 3) of pilots, copilots, instructors, and noninstructors*

Sample	N	M	SD
Pilot	238	180.01	17.45
Copilot	128	187.64	18.90
Noninstructor (pilot)	52	178.40	17.82
Noninstructor (copilot)	123	187.28	19.00
Instructor (pilot)	186	191.97	16.15
Instructor (copilot)	5	196.60	13.40

When the data were broken down further, it was found that the copilots who had not served as instructors obtained better scores than the pilots who had not served as instructors. The difference was of a size ($t=2.91$) which would be expected by chance less than one time in 100. The noninstructor copilots have an average of 400 hours flying time while the noninstructor pilots had 1,500. The most plausible interpretation of the difference between the test scores of these two groups is that, since the pilots with more flying time had graduated earlier, they had received the less comprehensive ground school training which was given during that earlier period, and had had more time to forget what they had learned. The more recently graduated copilots had received the benefit of the improvement in training and had had less time to forget.

This interpretation tends to be confirmed by the fact that the instructor pilots, who presumably are kept more up to date by the nature of their job, scored better on the test than either the noninstructor pilots or the noninstructor copilots, the differences being of a size ($t=4.89$ and 2.24) which would be expected by chance less than 1 and less than 5 times out of 100, respectively. These differences may have also been caused by the selection of the better pilots to become instructors. They could not have been caused by differences in the amount of flying time since the instructor and noninstructor pilots were approximately equal in this respect.

The correlation between number of years of education and score on this test was found to be 0.05 ($N=362$), indicating that the factors measured by this test were not related to the formal education of these students.

Finally, Psychological Research Project (combat crew) has shown that this test correlated 0.16 ($N=480$) with the subjective grade of flying skill given at the transition schools and 0.32 ($N=544$) with the pilot stanine. It is interesting, though not surprising, to note that the pilot stanine predicts the score on this pilot information achievement test better than does previous education.

The P. D. C. (Personnel Distribution Command) Pilot Information Test Form 1

At the request of the redistribution psychological program in the AAF Personnel Distribution Command, a Pilot Information Test of

150 multiple-choice items was prepared for possible use in the instructor selection battery administered to returnee pilots. This test was designed to replace the Analysis of Maneuvers Test described in Chapter 14 on Instructor Selection and was given the Personnel Distribution Command test code number DS4044. The test contains 25 items on analysis of maneuvers, 25 on navigation, 25 on aeroequipment, 25 on instrument flying, 25 on weather, and 25 on personal equipment. These 150 test items were taken directly from the Pilot Information Test, Form 3.

The test items were adapted for both single- and multi-engine pilots and the whole examination required 80 minutes of time to administer as a power test. The total score is the sum of the number of correct answers.

The Psychological Branch at AAF Redistribution Station No. 2, Miami Beach, Fla., administered this test experimentally and worked up the data on 371 pilots. These are presented in table 12.4.

TABLE 12.4.—*Odd-even reliability of the pilot- information test (PDC Form 1)*

Sample	N	M	SD	r_{tt}
Odd items	371	43.17	6.29	
Even items	371	44.54	6.19	
Total items	371	87.70	11.40	0.81

It can be seen that the distribution of scores on the test and the odd-even reliability of 0.81 are satisfactory. The test was considered to be readily administrable and suitable for use in the redistribution stations.

The mean item level of difficulty of the Pilot Information Test, Form 1, P. D. C., is 58 percent. All items were answered correctly by more of the students whose total scores were in the top 27 percent than by those in the bottom 27 percent. In other words, none of the items were inconsistent with the rest of the test.

In table 12.5, a comparison of the mean item level of difficulty of 150 items common to the Pilot Information Test, Form 3, and the Pilot Information Test, Form 1, P. D. C., shows that in each subsection of the test the transition-school pilots are better than the returned combat pilots. This superiority was probably not due to the fact that the returned combat pilots were both single and multiengine pilots while the transition-school graduates were all multiengine, very heavy bombardment pilots; the Psychological Branch of Redistribution Station No. 2 at Miami Beach, Fla., found that the difference, if any, between these two types of pilots on the scores in a similar test, Analysis of Maneuvers, was in favor of the single-engine pilots. Furthermore, in testing advanced-school students trained at single- and twin-engine schools, no differences were found on similar test items. Thus, it seems likely that the superiority of the more recent graduates was due to improvement in the curriculum on which they were trained and less time for forgetting. However, since the stanine level of the two groups was not controlled, the obtained difference might be due, at least in part, to a difference in stanine for the two groups. Such a difference could easily occur, if the returned combat pilots had been tested appreciably earlier than the transition pilots.

Table 12.5.—*A comparison of scores on the pilot information test (PDC Form 1) made by returnee and by transition-school pilots*

Subsection	Very heavy bombardment transition pilots ($N=371$)	Returned combat pilots ($N=371$)
	Percent	*Percent*
Flying information	66	63
Navigation information	64	57
Aeroequipment information	72	66
Instrument-flying information	67	56
Weather information	73	62
Personal-equipment information	55	47
Total test	67	58

The Pilot Instrument Flying Information Test

As has already been pointed out, the extreme importance of instrument flying was one of the reasons why the Pilot Project concentrated in developing objective measures of flying skill for that type of flying. But, as was shown in chapter 3, the good instrument pilot needs knowledge as well as skill. The Pilot Project, therefore, also developed tests aimed at those types of knowledge involved in instrument flying which could be measured more efficiently on the ground than in the air.

The Pilot Instrument Flying Information Test (Preliminary Form) was composed of 145 items with a scope including the following categories: Principles of operation and limits of the flight instruments, radio compass operation and procedures, AAF instrument approach system, radio range, flight rules, instrument procedures, instrument flying problems, weather fronts, thunderstorms, icing, wind, fog, atmospheric stability, weather reports, flight planning, weather maps, weather flight procedures, and judgment based on these facts.

This test was constructed with the cooperation of expert personnel of the AAF Instrument Flying Standardization Board. This board verified technical information in the test and recommended its use as a prerequisite to the oral examination for the white card which authorizes a pilot to fly under instrument conditions. Sample items from the test follow:

If the ball moves to the low side of the turn, it indicates a slip due to

A. overbanking.

B. underbanking.

C. rate of turn.

Exclusive of take-off and landing the minimum altitude above terrain on instrument flights shall be

A. 500 feet.

B. 1,000 feet.

C. 1,500 feet.

D. 2,000 feet.

The general limits of operation of the artificial horizon are

A. 60° pitch and 90° bank.

B. 70° pitch and 55° bank.

C. 80° pitch and 100° bank.

D. 90° pitch and 100° bank.

In recovery from a diving spiral, where air speed is excessive, first

A. reduce power, level wings, and correct pitch altitude.

B. stop turn and check excessive speed.

C. pull back on stick and coordinate rudder and ailerons.

D. stop turn, reduce power, and lower nose.

The examination required 90 minutes of time for administration. The total score is the number of correct responses.

Mimeographed copies of the Pilot Instrument Flying Information Test (preliminary form), were administered by personnel in the Pilot Project to 151 advanced pilot students (Class 45D) at Douglas Army Air Field, twin-engine school, Douglas, Ariz., and to 164 advanced pilot students (Class 45D) at Luke Army Air Field, single-engine school, Phoenix, Ariz. Sufficient time was allowed for all students to complete the test.

The means, standard deviations, and reliabilities of the total test and each of the subsections together with reliabilities of the total test for the Douglas Army Air Field sample, the Luke Army Air Field sample, and the combined samples are shown in table 12.6.[78]

TABLE 12.6.—*Mean, standard deviation, and odd-even reliability of the pilot instrument flying information test (preliminary form)*

Pilot school	Instrument flying			Weather		Total test		r_{11}
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Douglas	151	60.46	6.61	44.03	5.68	104.49	10.96	0.92
Luke	164	59.72	7.25	45.07	5.32	104.79	11.18	.94
Douglas and Luke	315	60.06	6.96	44.47	5.53	104.65	11.09	[1].93

r_{11} Reliability computed by the Rulon formula. (The Spearman-Brown correction for double length is an inherent part of this formula.)

[1] Combined by Fisher's z-transformation.

The Pilot Instrument Flying Information Test (preliminary form) has a satisfactory range warranting its use for selection purposes.

[78] Pfc Stuart Foulke and Pfc William H. Wiesenberg were responsible for most of the analysis of data reported in tables 12.6, 12.7, and 12.8.

Results from the two schools are similar. The agreement between the odd and the even sections of the test indicates that it gives an accurate total score not greatly influenced by chance factors. The correlation between the two subsections (Instrument Flying and Weather) of the Pilot Instrument Flying Information Test (preliminary form) is 0.57. As table 12.7 shows, there is a positive relationship between the total test score and the pilot stanine.

TABLE 12.7.—*Correlation of pilot instrument flying information test (preliminary form) with the pilot stanine*

Pilot school	*N*	r_{11}	r_{ts}	cr_{ts}	Stanine	
					Mean	*SD*
Douglas	143	0.82	0.25	0.35	7.28	1.40
Luke	160	.84	.23	.31	7.28	1.43
Douglas and Luke	303	.83	.24	[1].33	7.28	1.43

r_{11} Odd-even reliability of Pilot Instrument Information Test computed by the Rulon Formula (the Spearman-Brown correction for double length is an inherent part of this formula).
r_{ts} Uncorrected correlation of test with the pilot stanine.
cr_{ts} Correlation of test with pilot stanine corrected for restriction of range by use of formula presented by Kelley, T. L., Statistical Method, Macmillan, New York, 1924, pp. 225–227.
[1] Combined by Fisher's z-Transformation.

The correlations of each sub-section of the test with grades in the two ground school subjects, flight instruments and weather, are shown in table 12.8.

TABLE 12.8.—*Correlations between ground-school grades and scores on subsections of the pilot-instrument flying-information test (preliminary form)*

Pilot school	Ground-school course	*N*	Mean	*SD*	Pilot-flying instrument information test	
					Instrument	Flying weather
Douglas	Flight instruments	151	81.41	7.89	0.22	
	Weather	151	82.43	5.90		0.50
Luke	Flight instruments	163	2.77	.86	[1].16	
	Weather	163	2.39	.63		[1].48

[1] Sheppard's correction for grouped data was used on the Luke data only, because the scores were recorded by letter grade while the Douglas ground school grades were raw numerical scores. The uncorrected figures for the Luke flight instrument and weather correlations are 0.14 and 0.44 respectively.

The correlation between the instrument flying section of the Pilot Instrument Flying Information Test (preliminary form) and the ground school grades in flight instruments are positive but low. In contrast, the correlation between the Weather section of the Pilot Instrument Flying Information Test (preliminary form) and ground school grades in weather are significantly higher. This difference may be due to the fact that only 7 hours are devoted to flight instruments in ground school while 17 hours are devoted to weather.

This form of the Pilot Instrument Flying Information Test was revised to a 150-item printed test called Form A. An alternate form of the test, Form B, was completed and made ready for use as a "selective or diagnostic instrument." No statistics were available on these two forms by the end of the war, though it was highly probable that they were comparable to each other and to the preliminary form of the Pilot Instrument Flying Information Test described above. The AAF Instrument Flying Standardization Board recommended these forms for use throughout the Air Forces as one of the requirements for the card which signifies that a pilot may be cleared for instrument flights and the Training Command adopted them for this use.

SUMMARY

Reports from theaters of operation stressed the importance in combat flying of good judgment based on a thorough knowledge of practical aero-dynamics, navigation, aero-equipment, personal equipment, weather, and principles of instrument flying. Knowledge in these areas was considered to be a necessary, though not a sufficient, condition for good judgment in combat flying. Experts in each of these areas were asked to select those points of information most useful in keeping pilots out of trouble or helping them solve problems. A battery of subtests was constructed to give reliable measures of these points of information. This battery was administered experimentally and refined on the basis of item analysis. These tests, along with other variables such as flying grades, have been used in selecting the best pilots to be grouped with the best navigators, bombardiers, and flight engineers for training as potential lead crews.

Two other tests have been constructed. One of these was developed for possible use as one of the tests in an improved pilot instructor selection battery to be administered at AAF Redistribution stations. The other test covered information relevant to instrument flying, flight planning, and weather. The AAF Instrument Flying Standardization Board has recommended this test as one of the requirements for securing the instrument card which authorizes a pilot to fly under instrument conditions.

CHAPTER THIRTEEN

Training Experiments

1st Lt. William E. Galt

INTRODUCTION

Value of Scientifically Designed Experiments

Planning the best program of pilot training involves many problems which can be solved conclusively only by training experiments. The more scientifically these experiments are designed, the easier it is to get accurate, decisive results.

Many of these problems involve fundamental factors in human learning, such as distribution of practice and transfer of training, which psychologists have already investigated in the laboratory. For example, fixed gunnery training is given first on the less expensive and lower powered AT-6 with the expectation that the pilot will benefit by the transfer of that training to higher powered tactical planes such as the P-51. Similarly, students are first given training on smaller single-engine planes to make it easier for them to learn to handle two- and four-engine aircraft. Some of the habits learned on the lighter plane apply to the heavier one and are a distinct advantage; others, however, may not apply and may actually cause a certain amount of habit interference. The net effects of the positive and negative transfer can be expected to vary, depending on the similarity of the two situations and the type and amount of training given. Under some conditions the effects of the first training may be beneficial; but it is also possible in some cases for such training to be largely wasted or even harmful; or a limited amount of training may be valuable, but more than this amount a waste of time. While the results of laboratory experiments cannot predict exactly what will happen in a complex situation like that involved in flying training, they can tell one what to look for and how to go about setting up a series of training experiments to decide what kinds and amounts of training at each level will be most efficient in the long run.

Scientific Experiment Involves More Than Just Trying Something Out

While the essential idea behind the experimental approach is trying something out to see how it works, a truly scientific experiment involves far more than this. Its great power lies in the fact that conditions are arranged so that the effects of disturbing factors are controlled or cancelled out and the results are decisive and unambiguous. It is also necessary to devise some way of measuring these results so that they can be stated clearly and interpreted in the same way by everyone.

Little Psychological Work on Training Experiments During the War

Serving as consultants on the scientific design of experiments to evaluate training methods and devices should ultimately be one of the more useful functions of psychological personnel working with the Army Air Forces. Under the stress of work to be accomplished during the years of the war, however, the activities of aviation psychologists were so concentrated on other problems more closely related to the original research on classification tests that little opportunity was left to cooperate in training experiments.

Some of the work of aviation psychologists during the war years was relevant to training experiments. The research on methods of measuring flying skill should help to serve as a foundation for experimental work, since no experiment can be better than the yardstick used to measure its results. The evaluation of effects of additional training during the training freeze (chapter 10) might be considered to be semi-experimental, but since the conditions of this additional training were designed for another purpose, there was no opportunity for establishing suitable control groups. Finally, there were the two studies which will be described in this chapter. Because of the conditions under which they were conducted, these studies cannot be taken to illustrate the best principles of experimental design. Their description, however, including the faults which could have been corrected under better circumstances, will serve to illustrate what is meant by the experimental approach to training problems.

THE EFFECT OF DIFFERENT TRAINING PLANES IN BASIC SCHOOL ON FLYING SKILL IN ADVANCED TRAINING [76]

The first "experiment" to be described involved comparing two groups of students who had received their Basic flying training in different planes. All students received 15 weeks of training in Basic since they were involved in the temporary training freezes. The groups were:

1. Students who had received five weeks of single-engine training on the BT-13, five weeks of two-engine training on the AT-10, and five weeks of single-engine training on the AT-6. Since these students received some of their Basic flying training on a two-engine plane, they will be called the *two-engine group*.
2. Students who had received all 15 weeks of their basic phase of flying training on a single-engine plane, the AT-6. They will be called the *single-engine group*.

---

[76] Capt. S. C. Ericksen devised the rating forms on which the ratings of students were secured and coordinated the experimental procedure employed with supervisory personnel at Moody Army Air Field. The statistical analysis of the data was under the supervision of Sgt. Charles P. Gershenson.

The students in the two-engine group received their training at one basic school, Shaw, while those in the single-engine group received theirs at another, Cochran. Both groups received their advanced two-engine training on the TB–25 at the same school (Moody Field). There were 165 students in the two-engine and 84 students in the single-engine group.

Starting with these two groups the study was set up to answer the following questions:

1. How will these two groups, trained on different types of planes in basic school, compare in advanced two-engine training on the TB–25?

2. If one group shows an advantage over the other in the early phases of training, will this superiority be maintained throughout the 10-week training course?

Uncontrolled Variables

This experiment suffered from certain defects because of the fact that the two groups simply happened to be formed during the course of changing planes in the training program and were not deliberately set up as experimental groups. In the first place, the comparison between students with all of their training on the AT–6 and those with training on three different types of planes, the BT–13, AT–10, and AT–6, is not the most relevant comparison that could be made. The effects of training on a two-engine plane are mixed up with those of having training divided among three different types of planes. It would probably be more meaningful to compare a group with all of its training on the two-engine AT–10 with another one with all of its training on the single-engine AT–6.

The two groups are not matched. The fact that all of the students at one school were given one type of training and all of the students at the other school were given another type of training is good because it rules out the possibility that either of the schools selected a special type of student for the special type of training. Insofar as could be determined there were no special factors causing better students to be sent to one school than to the other. Experience has shown, however, that segregation of students into groups in the Training Command is not purely a random matter. There is some tendency for students entering at a given time and place, or having other features in common, to remain together so that groups may differ far more than would be expected by chance. Therefore, in designing an experiment, the safest thing to do is to select the two groups in such a way that they are either perfectly randomized or, better still, matched for any measurable factors that might affect the results. In this case, the ideal thing would have been to build up the two groups out of pairs of students who had no flying experience before primary, had equal stanines, and had received the same flying grades at the same primary school. The correlation between the scores of the members of each matched pair would reduce the standard error of the difference between the two groups, and hence increase the discrimination of the experiment.

Since the two groups came from different schools, there is a possibility that differences between them may have been caused by differences in the level of instruction at these schools rather than by the types of planes employed. It would have been desirable to have divided one Basic school into two squadrons with instructors as comparable as possible. Then Squadron A should have taught the two-engine group and Squadron B the single-engine group for one class and the two groups have been reversed (Squadron A teaching single-engine and Squadron B two-engine students) for the next class. Although control procedures of this kind may seem inconvenient, they save time and money in the long run by allowing decisive results to be achieved more quickly with fewer students tested.

Procedure Followed at the Advanced Two-Engine School

The way in which the two groups were originally derived was not under experimental control, but after they arrived at the Advanced two-engine school every effort was made to study them under comparable conditions.

Each instructor was assigned insofar as possible the same proportion of students from each group. The performance of the students was measured in the following way: At the conclusion of each week of training, each instructor was required to rank-order his students on 12 aspects of proficiency. The categories used were taken from the grade slip and hence were ones which the instructors were accustomed to observe. They covered 8 general aspects of flying and 4 general characteristics of each student's work as follows:

Taxiing	Formation
Take-off and climb	Night landings
Stalls	Attitude toward work
Slow flying	Technique
Single-engine work	Progress
Landings	Judgment

The instructors were warned to guard against the common error of overlooking the few weak points of good students or the few strong points of poor students in making their ratings. They were instructed to rate their students on each item separately guarding against the common tendency (called "halo" effect) to let the rating given a student on one item affect the rating given him on another. They were also instructed to keep the items on the scale in mind when flying each mission so that they would observe as many things as possible on which to base the ratings given at the end of each week.

By assigning equal proportions of each group to each instructor and requiring him to rank-order his students, the most sensitive possible subjective measure was obtained. This procedure avoided one of the chief difficulties inherent in subjective grades, namely the difference in the absolute standards of various instructors, and allowed each one to make relative judgments. It also forced an adequate distribution. The rank-order of any given instructor will be influenced by whether he happens to get the best or poorest students from one of the groups; the average rank-order of the whole group, however, will reflect its proficiency.

In addition to rank-ordering their students each week on the items contained in the rating form, the instructor and supervisory personnel were required to fill in a short questionnaire at the end of the 10-week course in which they gave their over-all impressions of the relative proficiency of the two groups of students with special reference to aspects of training in which either group was superior and any change in the relative performance of the two groups from beginning to end of the course.

Results

The data derived from this study were analyzed in the following way. Each instructor's students were divided into two groups: those in the higher half of the rank-order and those in the lower half of the rank-order. When an instructor had an odd number of students, the middle case was discarded. In table 13.1 are presented figures contrasting the percentage of students in each of the two groups (single-engine group and two-engine group) who were rated in the upper half by their instructor on each of the items rated at the completion of each week of advanced training. The reliability of the differences on each of the items at each week of training is also presented. The same data for two items--Single-Engine Work and Technique—are presented graphically in figure 1.

It can be seen that having had basic school training in the three different types of planes, one of which was two-engine, is a definite advantage during the first week of advanced two-engine training in taxiing, take-off and climb, single-engine work, progress, and technique. In general, these are the aspects of flying in which two-engine training differs most from single-engine training. Furthermore, it will be seen that these differences have been substantially reduced by the end of the ninth week of Advanced two-engine training, although the advantage is still in favor of those students who were trained in a two-engine plane in Basic school. For the first week of training the differences on all five of these items are of a size which would be expected by chance alone less than one time in 100. At the end of the ninth week the difference in taxiing would still be expected by chance less than one time in 100 and that in single-engine procedure only three times in 100.

TABLE 13.1.—*Effect of using different types of airplanes in basic on performance in advanced flying training*

Item	Group	Week of advanced flying training																	
		First		Second		Third		Fourth		Fifth		Sixth		Seventh		Eighth		Ninth	
		%	P	%	P	%	P	%	P	%	P	%	P	%	P	%	P	%	P
Taxiing	T	60	0.01	60	0.01	60	0.01	54	0.01	59	0.01	54	0.05	56	0.01	59	0.01	57	0.01
	S	29		32		32		42		33		42		39		34		38	
Take-off and climb	T	59	.01	61	.01	56	.01	60	.01	58	.01	52	.25	54	.07	56	.01	51	.34
	S	32		30		40		35		36		47		43		39		48	
Stalls	T	52	.28	58	.01	59	.01	57	.05	50	.50	59	.02	54	.21	59	.14	59	.10
	S	46		36		32		39		50		35		43		38		38	
Slow flying	T	50	.50	66	.01	57	.02	55	.16	53	.27	55	.19	54	.25	58	.10	74	.01
	S	50		22		36		43		43		43		43		38		13	
Single engine	T	61	.01	60	.01	56	.01	56	.01	59	.01	53	.15	58	.01	57	.01	55	.03
	S	31		33		38		38		33		45		36		37		41	
Landings	T	56	.08	50	.01	57	.01	55	.03	57	.01	56	.01	54	.06	48	x	55	.03
	S	38		32		31		41		38		39		42		54		41	
Formation	T			No training				55	.10	60	.50	59	.01	52	.27	54	.08	55	.03
	S			No training				41		50		35		47		43		41	
Night landing	T					No training								61	.01	57	.04	58	.01
	S					No training								25		39		36	
Attention to work	T	48	x	58	.01	50	.80	52	.25	49	x	51	.35	54	.07	51	.30	52	.21
	S	54		35		50		47		52		48		43		49		46	
Technique	T	60	.01	61	.01	55	.01	58	.01	56	.01	53	.14	53	.14	56	.01	54	.05
	S	33		31		34		35		38		45		45		40		42	
Progress	T	56	.01	54	.04	58	.01	57	.01	56	.02	51	.34	51	.34	53	.14	52	.25
	S	38		41		40		37		39		48		48		45		47	
Judgment	T	55	.04	61	.01	56	.02	56	.01	54	.08	55	.02	57	.01	55	.03	55	.03
	S	41		31		39		40		42		40		37		41		41	

T Two-engine group, consisting of 165 students. These students had received the following training in basic school: 5 weeks of single-engine training on the BT-13, 5 weeks of two-engine training on the AT-10, and 5 weeks of single-engine training on the AT-6.

S Single-engine group, consisting of 84 students. These students had received the following training in basic school: 15 weeks of single-engine training on the AT-6. Since all students were not always rated on all aspects of flying each week, the number of cases involved in the comparisons varied a little from week to week.

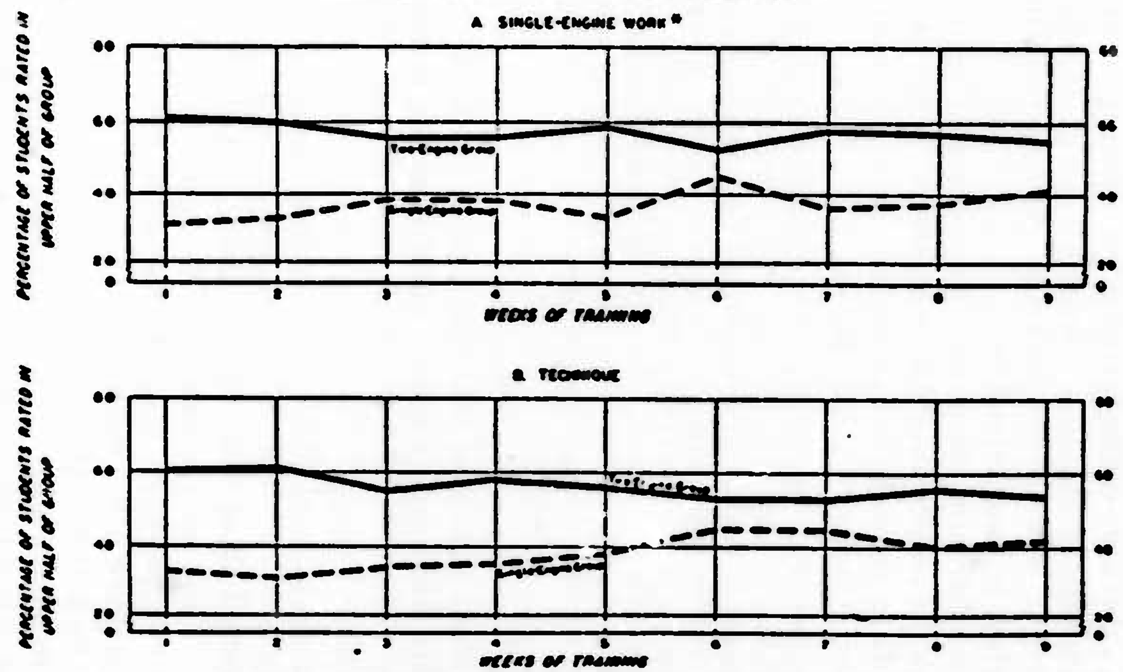
% Percentage of students in upper half of rank-order ratings.

P Probability of obtaining by chance alone a difference of this size in favor of the two-engine group. A "P" of 0.01 means that the observed difference would be expected to occur by chance only one time in 100.

x The percentage of single-engine students in the upper half was greater than the percentage of two-engine students.

FIGURE 13.1

Students with Single-Engine and Two-Engine Training in Basic Schools Compared on Flying Proficiency in Successive Weeks of Advanced Two-Engine School



* Instructor simulates engine failure by cutting throttle on one engine and forcing student to maintain flight with remaining engine.

Students in the Two-Engine Group had received 5 weeks of single-engine training on the BT-13, 5 weeks of two-engine training on the AT-10, and 5 weeks of single-engine training on the AT-6 during their basic flying training.

Students in the Single-Engine Group had received 10 weeks of single-engine training on the AT-6 during their basic flying training.

On the other aspects of flying and the other two general characteristics of the students' work rated, the differences during the first week of two-engine advanced training were not statistically reliable. The differences for the same items were also not statistically reliable during the ninth week of advanced two-engine training except for the item on slow flying. For this item, the two groups with diverse Basic training showed no difference during the first week of Advanced two-engine training but showed a large advantage during the ninth week of training in favor of the students who had been trained in a two-engine plane in basic school.

The subjective opinions of supervisors and instructors coincided well with the results obtained from analyzing the data. Some of the comments on general pilot ability and change in relative performance of the two groups are quoted below.

> The students of the two-engine group acquired the techniques and procedures more rapidly and with less apparent effort. They had no trouble with single-engine operation, flew patterns better and more accurately and had no trouble with checks while in the pattern. In general, they acquired twin-engine "know-how" more rapidly and this reflected on their general pilot ability.
>
> Students of the two-engine group started out with more on the ball and checked out sooner. Students of the single-engine group used up maximum permissible dual time, but did satisfactory work later.

Conclusion

This study indicated that the type of plane in which students receive their basic training does affect their performance in advanced training. Students who received some of their training in a two-engine plane (the AT-10) in basic school were more proficient in taxiing, take-off and climb, single-engine work, and technique during the early weeks of their advanced two-engine training than students who received all of their basic training in a single-engine plane (the AT-6). These are the aspects of flying in which two-engine training is most different from single-engine training. With continued two-engine training, the initial superiority of the students who were trained in a two-engine plane in basic school decreased but still tended to remain above the proficiency of students who had received no training in a two-engine plane at the basic level.

Though these results seem reasonable and are probably correct, it was impossible to be completely certain that the two groups were either completely randomized or matched before the experiment or to control the quality of instruction at the basic schools. This leaves the possibility that the difference between the two groups may have been produced either by systematic differences in the aptitude of the students originally assigned to them or to differences in the quality of instruction at the two schools where they received their basic flying.

Effect of AT-6 at Basic on Proficiency at Single-Engine Advanced

A similar experiment was set up at the advanced single-engine school at Craig Field, Selma, Ala. Some students entering class 45-A at that school had received their basic training on the BT-13 while others had received their basic training on the AT-6. Unfortunately, there were only 16 students in the AT-6 group in that class and special conditions at the school made it impractical to conduct an experiment on the following and only remaining mixed class. In general, the results showed the same trend as that found in the study reported above. Initially, the students trained on the AT-6 at basic school were considerably superior to those receiving their basic training on the BT-13. As training progressed the superiority of these students decreased although performance still remained above that of the other group on most items rated. The number of students involved, however, was too small to make the results conclusive.

EVALUATION OF THE SELF-REFLECTING OPTICAL SKEET SIGHT AS A FIXED GUNNERY TRAINING AID [7]

Under the supervision of the Training Aids Officer of Aloe Army Air Field, Victoria, Tex., a self-reflecting optical sight was designed for attachment to the conventional shotgun for use in teaching the sight picture for air-to-air fixed gunnery. The sight designed gave the student, firing skeet, the same sight picture (described in the first pert of chapter 11) that he sees later in fixed gunnery training when making passes with an AT-6 plane at an aerial towed target. Increasing this important aspect of the similarity between the two tasks should increase the transfer of training from shotgun to fixed gunnery firing.

In order to evaluate this newly developed sight as a training aid, the Training Aids Officer had each of three successive classes (44-1-G, 44-1-J, and 44-1-K) split into two randomly selected groups: a control group and an experimental group. The control group was given its skeet training as usual with the conventional shotgun. The experimental group fired the same number of rounds of skeet using shotguns equipped with the optical sight. This training took place during the advanced phase of pilot training. When these students graduated to gunnery training, three students from the control group and three students from the experimental group were assigned to each gunnery instructor. The selection of these groups of three for assignment was random.

Results

After the results were secured, personnel of the Pilot Project were asked to analyze the air-to-air fixed gunnery scores. Means, standard deviations, and critical ratios were determined for each subsection of

[7] Lt. William F. Galt and Sgt. John G. Oleason are responsible for the statistical analysis of the data from this experiment and for the evaluation of the results.

each class and for all classes combined. Table 13.2 presents these data.

TABLE 13.2.—*Effect of using different types of shotgun sight during skeet shooting training on subsequent performance in air-to-air fixed gunnery*

Class	Sight used in skeet training	*N*	Average gunnery score	*S. D.*	Difference between averages	*C. R.*	*P*
44-1-G	Optical	36	28.25	12.[illegible]	3.18	1.28	0.10
44-1-G	Conventional	44	25.07	8.[illegible]			
44-1-J (1)	Optical	38	29.71	8.[illegible]	5.34	2.43	.01
44-1-J (1)	Conventional	30	24.37	9.06			
44-1-J (2)	Optical	29	35.83	10.40	4.77	1.83	.03
44-1-J (2)	Conventional	33	31.06	10.11			
44-1-K (1)	Optical	18	32.33	11.42	2.27	.62	.27
44-1-K (1)	Conventional	17	30.06	10.16			
44-1-K (2)	Optical	24	35.67	13.12	−4.38	.90	([1])
44-1-K (2)	Conventional	20	40.05	15.67			
Total	Optical	145	32.16	10.90	3.19	2.40	.01
Total	Conventional	144	28.97	11.68			

P Probability of obtaining by chance alone a difference of this size in favor of the group trained on the optical sight. A "P" of 0.01 means that the observed difference would be expected to occur by chance only one time in 100.

[1] The difference was in favor of the group trained on the conventional sight.

It will be seen from the table that the mean percent hits in air-to-air fixed gunnery of the 145 students who were trained with the self-reflecting optical skeet sight is higher than the mean percent hits of 144 students who were trained with the conventional shotgun. The difference in favor of the group trained with the optical sight is of a size which would be expected to occur by chance less than one time in 100. The difference of 3.19 in the means of the two groups gives the group trained with the optical sight an average of 11 percent more hits in air-to-air gunnery than the control group. These data strongly indicate that the self-reflecting optical skeet sight is a valuable fixed gunnery training aid.

How Design of Experiment Might Have Been Improved

In general, the design of this experiment was quite good. Still better results might have been secured, however, if the plan had been refined in two ways.

The groups should have been matched on the basis of flying skill. In this experiment the flying ability of the experimental and the control groups was probably made approximately equal by assigning the students to the two groups in a purely random manner. When selection is made in this way, one will expect a certain amount of difference in the average level of flying ability of the two groups to occur by chance. If the groups are small, like those involved in each of the separate classes, these chance differences may be appreciable; if they are large, like the total for all of the classes involved in the experiment, the chance differences will become relatively much less important. Since the results for all of the classes combined showed a difference of the magnitude which would be expected to occur by chance less than one time in 100, it is safe to conclude that the different types of training produced a real difference in the two groups of students.

If one looks at the separate results on each class, however, one will see that there is a reversal in class 44-1-K (2). In this one out of the five classes tested, students who had been trained with the conventional shotgun did better than those who had been trained with the shotgun equipped with the optical sight. Further analysis of these results makes it seem likely that this reversal was produced by a few exceptionally good students who happened to be included in the group trained with the conventional shotgun.

The chance differences in the ability of the students assigned to the two groups might have been reduced by matching them on the basis of flying grades or instructor ratings instead of relying on random selection to produce groups of equal flying ability. If the groups had been matched before the beginning of the experiment, it seems likely that statistically reliable differences could have been secured without having to test so many students.

The fixed gunnery scores should have been based on the same number of rounds for all students. In this experiment the percent hits was based on all of the rounds fired. Different students fired different numbers of rounds. Since learning occurs throughout fixed gunnery training, scores of the students who fired more rounds will on the average be expected to be somewhat higher than those who fired fewer rounds. This will be a disturbing factor diluting the results of the experiment. Since the number of rounds fired is a random factor, it cannot account for the differences observed; it only tends to reduce the reliability of those differences. This difficulty could have been avoided by the technique (described in chapter 11 on Fixed Gunnery as an Objective Measure of Flying Skill) of basing the scores of all students on the same number of rounds. This would have increased the sensitivity of the experiment and the reliability of the results.

SUMMARY

Serving as consultants on the scientific design of experiments to evaluate training methods and devices should ultimately be one of the more useful functions of psychological personnel working with the Army Air Forces. Such emphasis was not possible during the war because of the concentration of psychological personnel on tasks more immediately related to the classification testing program. Two minor experiments, however, were undertaken in the area of pilot training. These are briefly described to illustrate the experimental approach to training problems even though, because of their incidental nature, it was not possible to control conditions in such a way as to secure ideal experimental designs.

One of these experiments indicated that students who had received part of their basic flying training on a two-engine airplane performed better in advanced two-engine training than those who had received all of their basic training on a single-engine airplane.

The other experiment compared the normal procedure of firing skeet with an ordinary shotgun with the procedure in which a newly devised self-reflecting optical sight was attached to the gun. The optical sight simulated the sight picture which the students would later employ in air-to-air fixed gunnery. It was found that the students trained on the skeet range with the gun equipped with the optical sight subsequently made reliably better scores in air-to-air firing than the students who received their skeet training with the conventional type of shotgun.

CHAPTER FOURTEEN

Evaluation and Selection of Flying Instructors

1st Lt. William E. Galt and Sgt. Daniel J. Grier

INTRODUCTION

The instructor occupied a key position in the pilot-training program because the qualities of one instructor influenced many students. The Pilot Project was therefore directed to concentrate research on the evaluation and selection of flying instructors. The main objectives of this research were: the determination of the basic interests and aptitudes of good flying instructors; the construction of experimental batteries of instructor selection tests measuring these basic interests and aptitudes; the development of scales for measuring instructor proficiency, and the use of these scales in validating the tests and in studying the relationship of background factors to teaching proficiency. The studies dealing with this research are described in the following seven subsections:

1. Analysis of the Abilities Characteristic of Good Flying Instructors.
2. Studies of Existing Methods of Instructor Selection.
3. Survey of Possible Criteria for Validation of Instructor Selection Tests.
4. Development of a Scale for Student Rating of Flying Instructors.
5. Analysis of Background Factors Related to Instructor Success.
6. Construction and Validation of Batteries of Instructor Selection Tests.
7. Summary of Results and Recommendations for Future Work.

ANALYSIS OF THE ABILITIES CHARACTERISTIC OF GOOD FLYING INSTRUCTORS

The first phase in the development of techniques of instructor selection was a job analysis of the most important characteristics of the good flying instructor. The studies made in this area are outlined below.

Some Early Work

In a field study early in 1943, Maj. Neal E. Miller and Capt. Donald E. Super, both assigned to Psychological Research Unit No. 1 at the time, spent 6 weeks living, studying, and flying with the students in a primary flying school. Later on, they followed up the students by visiting and interviewing them toward the end of their basic and advanced training. This work furnished background information on the job of the flying instructor.

Prior to the establishment of Psychological Research Project (Pilot), field studies of the characteristics of good civilian flying instructors were made by Psychological Research Unit No. 2 under the supervision of Capt. Richard P. Youtz. The first step in these studies was to interview supervisory personnel and experienced instructors on the qualities which they considered the most important characteristics of good instructors. Supervisors were also asked to rate their instructors and then describe the ways in which the best differed from the poorest. In this way a list of characteristics considered important in differentiating good and poor instructors was evolved.

This list of characteristics of good instructors was used as a basis for developing a scale on which supervisors could rate various observable aspects of instructional ability. Most of the work of developing this rating scale was done at Psychological Research Unit No. 2,[18] the officers of the Pilot Project supplying criticisms and a few suggestions for modification. Most of the items on this scale were qualities of good instructors but a few referred to the results of good instruction.

The Navy approached the problem similarly, except that naval aviators who had been through pilot training were asked to write out descriptions of the characteristics of their best and poorest instructors. In this way, data were obtained on 93 good instructors and 78 poor instructors. The characteristics considered important in differentiating the good from the poor instructors were rank-ordered in accordance with the percentage of times that each trait was mentioned.

Examples of Good and Poor Techniques of Teaching Flying

A considerable number of examples of good techniques in teaching flying was available from a collection made by Maj. Neal E. Miller in the spring of 1942 from instructors attending a refresher course at the Central Instructors School, Maxwell Field, Ala. This collection was rounded out by Lt. Wallace Nygard and Sgt. Daniel J. Grier, who collected further examples of both good and bad techniques from instructors in the field and from recently graduated pilots attending the Central Instructors School at Randolph Field. The organized file of such examples furnished the materials for a 50-item check list of ways in which teachers could most improve their flying instruction. The items most frequently checked on the list were used as a basis for constructing instructor rating scales.

[18] Capt. John T. Dailey was the major contributor in this work. He was assisted by the following officers and enlisted men: Maj. M. P. Crawford, Capt. Richard P. Youtz, Capt. W. A. McClelland, Lt. J. T. Cowles, Lt. J. Weitz, Lt. C. P. Froehlich, Lt. R. J. Keller, Lt. I. E. Farber, Lt. C. H. Patterson, and S/Sgt. I. Robbins.

The most frequent suggestions of recently graduated students centered around the following general aspects of instructing:

1. *Not using ridicule, sarcasm, or abusive language.*—Use of oaths, ridicule, or sarcasm at a time when the student was already underconfident, tense, or confused was a frequent criticism by students of their instructor's behavior. Students reported that this type of behavior tended to increase their difficulties.

2. *Ability to express himself.*—Their instructor's failure to give a clear enough explanation before and after demonstration was also a frequent criticism. The students reported that they had difficulty because they did not know what they were supposed to observe or try to do, or why they were doing it.

3. *Analysis of errors.*—The students said that their instructors were not specific enough in telling them exactly what was wrong in their maneuvers.

4. *Riding controls.*—The students reported that before solo they often did not have the opportunity to practice stalling, landing, and getting out of difficulties by themselves because their instructors took over the controls on too slight a provocation.

5. *Interest in teaching flying.*—A frequent criticism was that the instructor lacked interest in the job of teaching flying.

6. *Self-confidence — Not afraid of plane or student.*—Students reported that their instructor sometimes feared a particular maneuver or had bad flying habits which he passed on to them.

From the examples of good and bad techniques of teaching flying a check list of 50 items was prepared by Lt. Wallace Nygard and Sgt. Daniel J. Grier of the Pilot Project. This check list was administered to 304 recent Advanced school graduates in class 44-7 at Central Instructors School, Randolph Field, Tex., and to 326 students in class 44-8. The first group of students was given the check list after they had attended a number of CIS courses; the second group was tested before they had received any instructor training. The students were asked to check the 10 items which they felt would most improve the instructional ability of their advanced school instructors.

The two groups gave relatively similar responses. Since the second and larger group was tested before any indoctrination at Central Instructors School, their results may be taken as most representative of the opinions of the average student. The responses were tabulated separately for graduates of single- and two-engine schools. Since only minor differences were found, these results were combined.

Table 14.1 lists in order of importance the 10 items most frequently selected by the students who had just entered Central Instructors School and not yet received any indoctrination there. The percentage of students who selected the item as being among the first 10 in which their advanced school instructors could have shown the most improvement is shown in the first column.

TABLE 14.1.—*Items on which advanced school graduates thought it most important for their instructor to improve*

Based on 326 recent graduates from Single- and Two-Engine Schools (Entering students in CIS Class 44-8)

Percentage selecting item among first 10 out of a list of 50	Rank order	I wish that my instructor would—
Percent		
65	1	Tell me immediately when I do something wrong.
56	2	Illustrate more frequently how the maneuvers (sequences) are used in combat.
55	3	Put greater teaching emphasis on performing in emergencies and using common sense.
53	4	More clearly outline every maneuver that students should practice on solo flights.
49	5	Be more thorough in analysis of errors.
42	6	Explain on the ground more in detail the what, why, and how of the maneuver, rather than teach by the "now you follow through" demonstration method so exclusively.
41	7	Give verbal instructions far enough ahead of a maneuver to allow plenty of time for executing the maneuver.
35	8	Give me more instruction in how a sequence should be done, instead of merely pointing out the errors.
30	9	Have more standardization of his instruction with other instructors.
29	10	Take more time to tie in each maneuver (sequence) with other maneuvers I have already learned.

The items in the check list are specific. Stating them in more general terms, it can be seen that the students felt that their instructors were deficient in giving explanations and in expressing themselves, in analysis of errors, and in adapting their methods to the individual student.

Qualities Considered to Be Most Important Characteristics of Good Flying Instructors

Further studies dealt more directly with the qualities which supervisors, instructors, and students considered the most important characteristics of the good flying instructor.

On the basis of examples of good and poor teaching techniques, interviews with supervisors, instructors and students, and experience gained in collecting examples of flying instruction from good and poor instructors, lists of qualities important for flying instruction were drawn up. One of these lists of 25 qualities was used in 3 studies by the Pilot Project. One hundred and forty-one students in 3 primary schools, 250 students in 2 basic schools, and 364 students in an advanced single-engine school, an advanced two-engine and a transition school were asked to check the 5 most important instructor qualities of the 25 qualities listed.

To the results of these studies may be added two other similar ones. In one study Capts. Richard P. Youtz and John T. Dailey of Psychological Research Unit No. 2 had 71 primary school supervisors place in rank order the qualities important in a flying instructor. In the other study, a check list, paralleling the items of the rating scale developed by Psychological Research Unit No. 2, was made up and presented to 309 instructors at 4 advanced flying schools by Capt. Stanford C. Ericksen and T/Sgt. Robert R. Blake. The instructors ranked the qualities in order of importance by giving the quality most important for a good instructor a rank of "1," the next most important quality a rank of "2," and so on down to the least important quality which was assigned a rank of "14."

TABLE 14.2.—*The six qualities most important for flying instructors*

[Arranged in the Order of Importance Assigned to Them]

As judged by supervisors (Primary school) N=71

1. Interest in teaching flying
2. Analysis of errors [2]
3. Adapting methods to individual [1]
4. Patience and self-control [1]
5. Ability to express himself [1]
6. Flying ability [1]

As judged by students (Primary school) N=141

1. Analysis of errors [2]
2. Patience and self-control [1]
3. Flying ability [1]
4. Ability to express himself [1]
5. Adapting methods to individual [1]
6. Interest in teaching flying

As judged by students (Basic school) N=250

1. Analysis of errors [2]
2. Patience and self-control [1]
3. Flying ability [1]
4. Adapting methods to individual [1]
5. Not using ridicule, sarcasm or abusive language
6. Ability to express himself [1]

As judged by instructors (Advanced school) N=309

1. Interest in teaching flying
2. Adapting methods to individual [1]
3. Analysis of errors [2]
4. Ability to express himself [1]
5. Flying ability [1]
6. Patience and self-control [1]

As judged by students (Advanced and Transition school) N=309

1. Patience and self-control [1]
2. Not riding controls
3. Adapting methods to individual [1]
4. Flying ability [1]
5. Ability to express himself [1]
6. Not using ridicule, sarcasm or abusive language

Qualities correlating most highly with anchor item of students' rating scale [3] N=1141

1. Flying ability [1]
2. Adapting methods to individual [1]
3. Analysis of errors [2]
4. Ability to express himself [1]
5. Use of flight time
6. Patience and self-control [1]

[1] Placed among the first six in importance by all six groups.
[2] Placed among the first six in importance by five out of six groups.
[3] There were two items of a general nature, namely, "Effectiveness of on-the-ground teaching" and "Conduct as an officer," which correlated more highly with the anchor item of the Students' Rating Scale than items 3, 4, 5, and 6. See table 14.6 for complete matrix of inter-correlations.

The six qualities considered most important by the groups involved in each of these studies are listed in table 14.2 in order of the importance assigned. Included in this table are also the items of the Students' Rating Scale developed by the Pilot Project which correlated most highly with the anchor item of the scale, "Effectiveness in putting flying training across." There is close agreement among supervisors, instructors, and students at all levels of training. The following four qualities were placed in the first six by all five groups: *adapting methods to the individual, patience and self-control, ability to express himself,* and *flying ability. Analysis of errors* was placed among the first six in importance by four of the five groups. These qualities also correlated highly with the anchor item in the Students' Rating Scale.

Study of Returned Combat Pilots [79]

In order to determine factors which are especially important in making the returned combat pilot a successful instructor, officers of the Pilot Project interviewed all returned combat pilots instructing at the advanced two-engine school at Brooks Field, Tex. The flight and squadron supervisors of these instructors, the commanding officer, and the director of training at Brooks Field were also interviewed to get their observations on the adjustment of these men to the instructor program. Pilot returnees attending various instructor schools such as the AAF Central Instructors School, Randolph Field, Tex., AAF Instructors School (Instrument Pilot), Bryan Army Air Field, Bryan, Tex., AAF Instructors School (Pilot 4-Engine, B-24), Smyrna Army Air Field, Smyrna, Tenn., and Instructors School (Pilot 4-Engine, B-17), Lockbourne Army Air Base, Columbus, Ohio, were also interviewed.

These interviews confirmed the general findings of the psychological research projects at the Bombardier and Navigator Central Instructors Schools that problems of morale, personality, and emotional adjustment were of especial importance in determining whether or not a combat returnee would make a good instructor. The problems of adjustment faced by these men were similar to those of returned bombardiers and navigators as described in the reports of the Bombardier and Navigator Projects.[80] They were on the whole, however, considerably less severe.

The more adequate adjustment of pilot returnees to instructor training was to be explained in part by the rapid advances in pilot technique which had occurred since these men were originally trained especially in two-engine and instrument flying. The men realized the value of the new techniques and were eager to learn. They were, however, more interested in the training itself than in becoming adept as instructors. The generally good facilities at the pilot instructors schools and the superior status of the pilot in the Army Air Forces seemed to be other factors contributing toward the generally better morale of these men than of returned bombardiers and navigators.

The chief difficulties of returned pilots were found to be emotional and to center about the following factors:

1. *Dislike of ground discipline and the details of administrative red*

---

[79] These studies were made by Maj. Neal E. Miller, Capt. Richard P. Youtz, Lt. William E. Galt, T/Sgt. Robert R. Blake and S/Sgt. Irving Robbins.

[80] Psychological Research in Bombardier Training, Psychological Research in Navigator Training.

tape to which they had not been accustomed in combat zones. Being sent to the instructors school was like being demoted to cadet status.

2. *Dislike for the routine schedule of continuous, monotonous, hard work.* In combat they had been accustomed to spurts of difficult, dangerous work followed by periods of relative inactivity during which they could do pretty much as they pleased. Some felt they had already done their part in the war and were entitled to a soft job.

3. *Difficulty of changing to a different kind of flying.* They resented the many checks which were a routine part of instructing but which were not adhered to in combat. Some of the men had lost certain aspects of precision necessary in instructing. Others objected to gliding the unarmored and lighter training planes so much more slowly than they had been accustomed to gliding their heavier planes in combat. There was the attitude that they had "been there and knew" and some expressed resentment about beginning all over again at the bottom and taking orders from men who had not been in combat and who were sometimes of lower rank.

4. *Reluctance to let students fly the plane.* Quite a few of these instructors felt that they were losing their flying skill and not getting any fun out of instructing because the students did all of the flying. Frequent narrow escapes in combat made others uneasy when a student, whom they did not trust at the controls, made various mistakes, such as coming in with one wing low, or in doing maneuvers, such as single-engine landings in a two-engine plane, which had been added to the training schedule since these instructors had been trained. Their uneasiness made it very hard for them to resist the impulse to take over the controls before the student had a chance to notice and correct the difficulty himself. Many expressed the opinion that instructing was more dangerous than combat and that they did not want to have survived combat only to be killed on a training mission.

Interviews disclosed that, for the married men, the most important source of satisfaction in the instructor assignment was the chance to have a relatively stable home life with their families. For them, the routine had its compensations in that they were able to plan ahead, counting on definite hours with their families. In a study made of the relationship between biographical factors and students' and supervisors' ratings of instructor proficiency, it was found that supervisors rated the instructors who were married and whose wives were living with them as better than the unmarried instructors.

Many other of the returned combat pilots did not relish an instructor assignment but preferred it to return to combat. The single men indicated that they would, for the most part, have preferred a roving assignment, such as in the Air Transport Command.

In order to secure some indication of the interest of returned combat pilots taking the instrument instructors course at Bryan in an instructor assignment, students in class 44-9 (the second class with a returnee population) were asked to indicate their interest in an instructor assignment on a 5-point scale. Forty-nine returnees, or 56 percent of the 87 men in the class, expressed high or moderate interest in this assignment; 24 men or 28 percent reported that they had no interest or were definitely opposed to such an assignment; and 14 men, or 16 percent, reported they were indifferent to a teaching assignment.

Summary of Basic Hypotheses about Qualities of Good Flying Instructors

As a result of the preceding studies and interviews, the following qualities are considered to be important to the success of a flying instructor. All of these factors are believed to be significant. No attempt is made to list them in order of importance. The attempt has been, rather, to place related qualities together.

1. *Interest in teaching flying.*—Experienced supervisors and instructors in both Primary and Advanced schools have ranked interest in the job of teaching flying as the one quality most important for a good instructor. Students also rate this quality as being important in an instructor. To the extent that an officer has a clear and correct idea of what the instructor assignment involves and is not choosing it merely to avoid some other job, such as combat, his preference for that assignment should predict his interest in it. In order to check on the men whose preferences may be based on inadequate knowledge or the desire to escape other assignments, the preference should be supplemented by other indirect measures of interest. In general, the men from certain backgrounds, with certain patterns of attitudes and related interests, are most likely to become and remain interested in teaching flying.

2. *Desire for stable family life.*—The basic assumption is that officers whose attitudes, interests, and circumstances are such that they will get satisfaction out of the type of life that instructors lead (a relatively routine, stable one for the wartime Air Forces) are more likely to become interested in teaching. This factor should be predictive of adjustment to the instructor situation regardless of whether the officer had a clear idea of the instructor's job or chose it to avoid some other assignment. Married officers, returned from combat and serving as instructors, stated that being able to live with their families and plan their time on a routine schedule was one of their sources of satisfaction with this assignment.

3. *Interest in people.*—Many instructors complained that their jobs were monotonous, flying over the same territory day after day. These men seemed to go stale after ten or twelve months of instructing. Other instructors reported that their job never became boring, that students were continually doing new things each day and that no two students were alike. These instructors reported that they were continuing to improve even after two or three years of instructing. The difference between these two is that the second type of instructor has

social insight and imagination and is interested in people. He is able to alter and adapt his technique of teaching to the individual needs of his student. One of the most important compensations for the effort of teaching is the satisfaction of seeing the student improve. Those interested in people are more likely to find this a source of keen satisfaction. Since interest, participation, enjoyment, and success are all likely to be related, having a history of participation in and enjoying activities involving people, having previously succeeded as a salesman, camp leader, physical director or teacher, should be favorable indications for success as a flying instructor.

4. *Ability to understand people and adapt techniques to the individual student.*—Supervisors, instructors, and students all ranked adapting teaching methods to the individual student among the very important qualities of a good instructor. This quality is especially important in a pilot instructor since learning to fly in the accelerated army program was likely to frighten, confuse, or discourage the student who was handled improperly. This ability is related to interest in people, the quality which has just been discussed.

5. *Patience and self-control.*—Students ranked this as being among the two most important attributes of an instructor. They made the point that it is very difficult to learn under conditions where the student feels that the instructor is critical of his mistakes, which are numerous, and may at any moment become angry and "blow up." Supervisors and instructors ranked this quality in the instructor in fourth and sixth place, respectively, in order of importance. Instructors mentioned that one of the more common sources of annoyance in their job was having student after student repeatedly make the same mistakes. As soon as one group of students improves, they graduate and a new group comes in so the process is repeated *ad infinitum.* Added to this is the fact that it is particularly exasperating to attempt to teach someone a motor skill where the instructor can see the mistakes easily but finds it difficult to explain because critical parts of the activity are not verbalized in our language. An instructor who does not possess a high degree of patience and self-control is likely to give his pupil a tongue lashing when he should be rewarding any slight signs of improvement which appear. This will often undermine the confidence of the student.

6. *Ability to express himself.*—Supervisors, instructors, and students all ranked this quality as being among the five most important attributes of a good flying instructor. Students stated that the lack of a detailed explanation before a maneuver is demonstrated and a concise statement as to just what was wrong with a maneuver were two of their most frequent criticisms of instruction. In order to give a good explanation, an instructor has to know his subject in words or diagrammatic symbols instead of just in terms of muscular movements. He may not need excellent grammar and vocabulary, but should be

able to express himself clearly. The problem of expression is made more difficult by the fact that so much of the skill is unverbalized.

7. *Ability to analyze errors.*—An instructor must be able to analyze a student's errors in order to tell him what he is doing wrong. Supervisors and instructors ranked this respectively as the second and third most important quality of a good instructor. Students ranked it somewhat lower but placed it well within the 10 most important qualities. A good pilot may not be able to tell the student specifically what he does incorrectly. It is necessary for the instructor to have his subject matter conceptualized in either verbal or visual symbols. Ability to do this is probably related to mechanical comprehension and knowledge of principles of flying.

8. *Ability as a pilot.*—The instructor must be a master of the art which he is trying to teach. He must not only be able to demonstrate the correct maneuver but also to mimic and exaggerate any type of error which the student may make. Supervisors placed flying ability sixth, instructors placed it fifth, and students rated it fourth in their list of the most important qualities of the good flying instructor.

9. *Self-confidence—Not afraid of plane or student.*—The instructor who is somewhat anxious is likely to take over the controls to relieve his own nervousness at times when he should allow the student to get practice in recognizing and correcting his own mistakes. Not riding controls was the quality selected by Advanced and Transition school students as being the second most important attribute of a good instructor. Furthermore, the instructor who is afraid either of the plane or the student is not likely to inspire confidence in his pupil. This is especially important in selecting instructors from the combat returnee population, since many of them still suffer from a generalized anxiety as a result of their combat experience.

10. *Conscientious habits of trying to do a good job.*—Some instructors, who have no particular interest in this assignment, try hard and succeed because they have formed the habit of doing the best work possible on any job that is assigned to them. Good work habits will help the instructor to cope with the more tiresome aspects of his job. To a certain extent conscientious work habits may be predicted from an individual's home background and other biographical facts. Conscientiousness may also be predicted from his behavior during cadet training and his attitude toward duty, discipline, and the army in general. Experience in the bombardier and navigator schools indicates that this type of factor is especially important with combat returnees.

11. *A good example of officer qualities.*—The instructor should be a good officer so that his students will respect him. His behavior on the ground and in the air should in all respects set a good example for the student, for his students will not only copy his flying but will also tend to pattern themselves after everything that he does. This is

especially true of Primary students who are just beginning their flying training.

STUDIES OF EXISTING METHODS OF INSTRUCTOR SELECTION

As another step in the research on flying instructors, studies were made of existing methods of instructor selection. These studies included investigations of the basis on which students are graduated from the AAF Central Instructors School at Randolph Field, of the factors which predict success at Central Instructors School, and of the way in which students are selected to be sent to that school. The task of the Central Instructors School was to train instructors for teaching contact flying in Basic and Advanced schools.

Basis for Graduation from Central Instructors School [8]

First a study was made of the basis for graduating or eliminating students from Central Instructors School. Interviews with supervisory personnel established the fact that with the relatively short period of time allotted for turning out large numbers of instructors, the main effort of the school was concentrated upon bringing them up to a satisfactory level of flying proficiency. They indicated that the graduation or elimination of a student-instructor was determined almost entirely by his ability to absorb the intensive course of flying instruction. In order to verify the actual operation of this policy at the Central Instructors School, a study was made of the relationship of flying grades and of academic ratings to graduation or elimination. This study involved 346 students in class 43-8. In each of the three groups at Central Instructors School (Basic, Advanced single-engine and Advanced two-engine), it was found that all of the graduates received a flying grade of D or better and all of the eliminees received a flying grade of F. No students received a grade of E. The fact that there was a *perfect* dichotomy between the graduates and eliminees on the basis of flying grades suggested that no other factors were involved in elimination and thus confirmed the results of interviews.

In order to confirm further this impression, the academic ratings of graduates and eliminees were compared. The academic rating was based on a student's proficiency in all ground school courses which included Analysis of Maneuvers and Psychology of Instruction. When this was done, it was found that there were no reliable differences between the academic ratings of graduates and eliminees for students in the Basic ($N=150$) or in the Advanced single-engine ($N=79$) groups. In the advanced two-engine group ($N=117$), the graduates received a significantly higher academic rating than the eliminees ($C.R=3.2$). The fact that there was on the whole so little

[8] Lt. Thomas N. Ewing and Sgt. John R. Rohrs were responsible for the statistical work involved in this study.

difference between the academic ratings of graduates and eliminees further confirmed the impression that students are graduated from this school on the basis of only one of the qualities necessary to the good flying instructor, namely, flying ability.

Prediction of Success at Central Instructors School [32]

After it was found that graduation from CIS (i. e., Central Instructors School) was determined solely by flying ability, an attempt was made to ascertain how well measures of flying proficiency available at the schools where the student-instructors are selected would predict th student's subsequent success at CIS. The Proficiency Card, AFTRC Form No. 2, was the record used to secure data on the flying proficiency of the students during their cadet training. It is a record which at that time followed the cadet through training and which contained estimates of his general pilot ability, ability on various aspects of flying, and on personal qualities at each stage of his training.

The proficiency card data showed that there was a low but positive relationship between ratings on the Proficiency Card at basic and advanced schools and elimination for flying deficiency at Central Instructors School. As would be expected, the ratings on flying ability were somewhat more predictive than the ratings on general personal qualities. Some of the specific flying ratings (such as the one on formation flying) were at least as predictive of success at CIS as the rating on General Pilot Ability. Combined scores, consisting of a simple sum of ratings on spins and stalls, acrobatics, and formation flying in basic school; judgment and common sense, night flying, and formation flying in advanced school, correlated 0.40 with graduation-elimination at Central Instructors School, while ratings on General Pilot Ability at advanced school correlated only 0.23. This study was based on the proficiency records of 746 Central Instructors School students in Classes 44–B, 44–C, and 44–D, Randolph Field, Tex.

Type of Student Sent to Central Instructors School [33]

Since the record of a student's flying proficiency in basic and advanced school was somewhat predictive of his success at Central Instructors School, a study was made to see if the students with better records were recommended and sent to this school for instructor training. Using a population of 2,523 students in Class 44–B, Eastern Flying Training Command, it was found that the pilots recommended for the instructors course at Central Instructors School were better on the basis of their proficiency card ratings than those not so recommended. It was further found that the recommendations

[32] Work on this study was accomplished by Capt. S. C. Ericksen, Sgt. John R. Rohrs, and Lt. Thomas N. Ewing.

[33] Work on this study was under the supervision of Capt. S. C. Ericksen. Lt. Thomas N. Ewing assisted in the study and is responsible for statistical analysis of the data.

for instructor assignment were followed quite well in assigning men to the CIS advanced single-engine and two-engine groups, but were not followed as well in assigning men to the CIS basic group. As a result, the students assigned to the advanced single-engine and two-engine groups at CIS were better than the rest of their class in ratings on general pilot ability, leadership, and the other flying and nonflying traits on the proficiency card, but those sent to the basic group were below the average of their class. This difference in the students which advanced schools sent to the Advanced and the basic groups at CIS may be explained by the fact that the students which were sent to the advanced groups were likely to come back to the same school as advanced instructors, while those sent to the basic group would not come back to that school. Therefore the advanced schools were motivated to send their best students for training as advanced school instructors but were not so motivated with respect to the students sent for training as Basic instructors.

To summarize, the current system of graduating (or eliminating) students in the Pilot Central Instructors School was found to be based almost entirely on flying proficiency; flying proficiency in earlier stages of training was somewhat predictive of successful completion of the training at Central Instructors School, and the men sent for instructor training in the advanced single-engine and two-engine groups were superior in flying proficiency to those not sent, while the men sent for training in the basic group were somewhat inferior because recommendations for assignments were not at all closely followed. The existing method of instructor selection thus took into account one of the qualities of a good instructor and, in general, channeled men into instructor training who were somewhat above average in this quality. It did not take much account of the many other qualities which are equally important in the good flying instructor.

Student Preferences [14]

Job analysis studies already described indicated that "interest in the job of teaching flying" was one of the most important qualities of a good instructor. Studies were therefore made of the attitude of students in advanced schools toward assignment as instructor. These studies were aimed at discovering whether enough students preferred the instructor assignment so that the selection of instructors could be made from among those who either wanted such an assignment or at least had no strenuous objections to it. Data on the attitude of the students were also relevant in deciding whether or not an attempt should be made to disguise instructor selection

[14] These studies were under the supervision of Capt. S. C. Ericksen. Lts. Thomas N. Ewing, Leo A. Hellmer and S/Sgt. Irving Robbins assisted him.

tests so that the students taking them would not be aware of their purpose. The results of these studies were as follows:

A substantial number of advanced school students in Classes 44-B and 44-C at Luke Field, Arizona (36 percent of the 886 students canvassed) expressed a preference for assignment as an instructor (first, second, or third choice out of 6 possible assignments). Students who had a preference for an instructor assignment were rated to be as good a group of flyers as those who had no preference for instructing.

In another study of 1,046 upper class students in three advanced schools (Class 44-E at Luke, Williams, and Douglas Fields), the instructor assignment was in fourth position out of 11 possible choices. Approximately 24 percent of the students indicated that they would be "highly satisfied" with an assignment as instructor. The cadets were asked to rate their own ability for each of the 11 listed assignments on which they were asked to indicate their preference. The results showed a rather high relationship between self-ratings of ability and preference.

Final results of this survey indicated that current quotas for instructor training at the AAF Central Instructors School (Pilot) could probably be met by selecting only those students who had some preference for this type of work and who, at the same time, met the other criteria for instructor selection.

SURVEY OF POSSIBLE CRITERIA FOR VALIDATION OF INSTRUCTOR SELECTION TESTS

Need for a Criterion

In order to refine and evaluate the instructor selection tests, it was necessary to have some criterion of how well the pilots who took the tests succeeded as teachers. The tests, or the individual items in these tests, that were found to predict success would then be kept and those which did not predict success would be discarded. Since there was no ready-made criterion of instructor proficiency, it was necessary to try to discover some good way of finding out how well potential flying instructors, who had been tested, eventually succeeded as teachers. The following subsections deal with studies of possible criteria for validating pilot instructor selection tests.

Judging Instructors by the Students They Produce

The ultimate measure of the success of a teacher is the quality of the students which he produces. An investigation undertaken at Psychological Research Unit No. 2 by Capts. Richard P. Youtz and John T. Dailey showed that in Army flying schools it was impracticable to use this measure as a criterion of teaching success for the following reasons: (1) The number of students assigned to each instructor was so small that it would have taken an unduly long time to build up statistically reliable groups; (2) in many schools, records on the assignment of students to instructors were available for the immediately preceding class but were incomplete for all preceding classes; (3) different administrative policies produced variations in the elimination rate from command to command, from school to school in each command, and from class to class within each school; (4) students were fairly often shifted from one instructor to another during any given phase of their pilot training. Weak students were likely to be shifted to those instructors who were believed to be the best. While this is a sound procedure from the point of view of training, it introduced serious distortions into the sampling.[85]

Since evaluating instructors by the quality of students which they produced seemed impractical, investigations were made of the feasibility of obtaining other types of criterion data.

Investigation of Data Available at Contact and Instrument Instructors Schools[86]

Since it was the policy of the Air Forces to send a considerable number of returned combat pilots to the various instructor schools from redistribution stations, personnel from the Pilot Project made surveys of the various types of instructors schools in order to determine what data were available against which a battery of instructor selection tests being administered to these returnees might be validated. The instructor schools investigated were: AAF Central Instructors School, Randolph Field, Tex.; Instructors School (Pilot 4-Engine, B-24), Smyrna Army Air Field, Smyrna, Tenn.; Instructors School (Pilot 4-Engine, B-17), Lockbourne Army Air Base, Columbus, Ohio; and AAF Instructors School (Instrument Pilot), Bryan, Tex. The conclusions with regard to the criterion data available at these schools are presented below.

No practice teaching of students by student-instructors nor any systematic program of simulated practice teaching of instructors by student-instructors was being carried out in any of the various CIS groups at Randolph Field, Smyrna, or Lockbourne. No separate grades were given of the instructing ability of the students independent of flying proficiency. Interviews with supervisory personnel indicated that it was necessary to use all the available time at these Central Instructors Schools to bring students up to the necessary level of flying proficiency and that no emphasis could be placed on teaching students to instruct in the air. All measures investigated at these schools were evaluations of the flying ability of the students rather than

[85] The Navy also found that they could not use the graduation-elimination rate of an instructor's students as an index of the instructor's proficiency for similar reasons.

[86] The surveys at the Instructors Schools were under the supervision of Lt. William E. Galt. Lt. Thomas N. Ewing, T/Sgt. Robert R. Blake, S/Sgt. Irving Robbins and S/Sgt. Walter Ismael assisted in the surveys. The statistical work involved was done by Sgt. John R. Rohrs.

measures of their proficiency as instructors. They would, therefore, be valid criteria for only this one of the many qualities important in the good flying instructor.

Three record forms were found available at the AAF Instructors School (Instrument Pilot), Bryan, Tex., which recorded in some measure a student's ability as an instructor. These were: (1) Recommendation for Association with the Instrument Instructor Program; (2) pass-fail; and (3) scores on an Instructor Check Ride. The first two were not suitable for use as criteria since the percentage of students not recommended and the percentage failing the course was relatively small.

In general, the desired criterion within the instructors schools is a measure of the instructional ability of the student in a practice-teaching situation. The instructor check ride was the only situation which gave any promise of fulfilling this condition, although it had the disadvantage of not being a natural teaching situation. In this check ride, which was administered once to each class of students, the student-instructor was required to simulate teaching the staff-instructor some previously assigned lesson. The student was graded on his *voice*, his *demonstration technique*, his *explanation technique*, and his *analysis of maneuvers*. He was graded as "Excellent," "Satisfactory," or "Unsatisfactory." An examination of the ratings of the first class of returned combat pilots showed a poor distribution of grades into the categories of "Excellent," "Satisfactory," and "Unsatisfactory," since too large a percentage (approximately 80 percent) fell into the middle or "Satisfactory" group.

Since it was necessary to improve the distribution of instructor check-ride scores if the check-rides were to be used as criterion data for validating instructor selection tests, a four-point scale of scoring was suggested to supervisory personnel. This was adopted. It had the effect of splitting the large "Satisfactory" group into two groups. It was also desired to enlarge the number of qualities rated on the check ride, to make them more objective and to apply a differential weighting to the various qualities. Such a revised check-ride card was drawn up by personnel from the Pilot Project in collaboration with picked Bryan instructors, but from an administrative standpoint it was not feasible to adopt many of the revisions recommended.

In order to evaluate further the adequacy of the Instructor Check Ride as a criterion against which to validate the instructor selection tests, studies were made to determine (1) the reliability of the check ride, (2) the intercorrelations of its various categories, and (3) the extent to which scores on the various categories of the check ride were influenced by previous experience as an instructor and by instrument flying proficiency.

The number and type of Bryan student-instructors involved in these three studies were as follows: 179 combat returnees and nonreturnees in class 44–9 in studies (1) and (2); 307 nonreturnees in classes 44–6, 44–7, and 44–9, and 374 nonreturnees in classes 44–6 and 44–7 in study (3). The results of these studies are presented in tabular form in table 14.3.

The reliability coefficients for the categories of the Instructor Check Ride when the retest was given by another check rider 7 to 10 days after the initial ride ranged from 0.25 for *Voice*, to 0.54 for *Explanation Technique*. The over-all reliability (sum of the categories) was 0.48; that of the sum of all categories except *Voice*, was 0.52.

The intercorrelations between the various categories of the check ride ranged from 0.62 to 0.90. In general, *Voice* showed the lowest intercorrelations with the other items. The generally high intercorrelations indicated a rather high relationship among the various categories with the exception of *Voice*. *Voice* probably correlated less highly with the other categories, because it was the least reliable. The high intercorrelations may have resulted in part from the fact that the various categories of the check ride measured similar things. The fact, however, that all of the intercorrelations were higher than the reliabilities of the individual items indicated that a considerable degree of "halo" effect was also present.

TABLE 14.3.—*Reliability, intercorrelations, and relationship of categories of the instrument instructor check ride to instructor experience and instrument flying proficiency*

Category	N=179						Previous experience as instructor N=307		Instrument flying proficiency N=374	
	1[1]	2	3	4	5	6	r_{bis}	SE.	r_{bis}	SE.
1. Voice	0.25	0.68	0.75	0.68	0.72	0.62	0.19	0.07	......	......
2. Demonstration technique	......	.40	.82	.89	.82	.77	.29	.07	......	......
3. Explanation technique	......	......	.54	.86	.84	.79	.20	.07	......	......
4. Knowledge of subject	......	......	......	.52	.90	.86	.30	.07	......	......
5. Analysis of maneuvers	......	......	......	......	.51	.86	.25	.08	......	......
6. Final Grade (over-all rating)	......	......	......	......	......	.50	.27	.08	......	......
7. Total score on six categories	......	......	......	......	......	......	.30	.07	.13	.03

[1] Mean-square contingency coefficients corrected for class index.

Previous experience as an instructor was found to improve instructor proficiency as measured by the ratings on the Instructor Check Ride Correlations were relatively low but all positive. They ranged from 0.19 for *Voice* to 0.30 for *Knowledge of Subject*.

In determining the relationship between scores on the Instructor Check Ride and instrument flying proficiency, the Basic Proficiency Check Ride was used as a measure of instrument flying skill. An unweighted combination of the different scores on this check ride of instrument flying proficiency was correlated with the unweighted combination of scores on the Instructor Check Ride. It was found that the combined score on the Instructor Check Ride was relatively independent of instrument flying skill as measured by the Basic Proficiency Check Ride. The correlation between these two check rides was 0.13 with a S. E. of 0.05. This showed that the Instructor Check Ride does not merely duplicate other measures of flying proficiency.

The results of the above studies indicated that the Bryan Instructor Check Ride is a possible criterion for the validation of the instructor selection tests. The fact that scores on the Instructor Check Ride were relatively independent of the rating on instrument flying ability and were somewhat related to previous experience as an instructor suggests that this was a desirable criterion of performance as a teacher. A disadvantage of the Instructor Check Ride as a measure of teaching proficiency was that the teaching situation involved was a simulated one (student-instructor teaching a staff instructor) rather than a real one (student-instructor teaching a student). There was the further disadvantage that it would not be possible to validate the instructor selection tests on the population of returned combat pilots flowing through the AAF Instructors School (Instrument Pilot) at Bryan, Tex. This was due to the fact that the number of returned combat pilots taking this course, on whom scores on the battery of tests administered at Redistribution Stations were also available, was found to be small.

Early Work on Ratings of Instructors by Their Students

Since the prospects of securing data at instructors schools against which to validate the pilot instructor selection tests were not encouraging, the development of some form of measurement of the individual instructor's teaching ability in the real teaching situation at the flying schools became important. A man's teaching ability after he is actually functioning in an instructor assignment, his on-the-job performance, should be a more valid criterion of his instructor proficiency than his manner of performance while receiving training at an instructors school. The administrative and statistical difficulties of judging instructors by the students they produce have already been discussed in this chapter. It became apparent that perhaps the only practical method of securing an evaluation of the flying instructor's teaching ability was through some type of rating scale.

Prior to the establishment of the Psychological Research Project (Pilot), Lt. Col. R. T. Sollenberger, from the Psychological Section, Headquarters AAF Training Command, had students at a primary school rate their instructors. He used a scale adapted from one which Lt. Commander Lowell Kelly developed and used in the Navy's flying training program.[87] Lt. Col. Sollenberger found that the primary students had a strong tendency to rate all of their instructors at the favorable extreme of the scale and concluded that the distri-

[87] U. S. Navy. A Diagnostic Scale for Rating Flight Instructors. N. A. P. T. C. Form No. 21L.

bution of responses was not adequate to justify further work at that level of training.

Members of the Pilot Project judged that better responses might be secured from students who knew more about flying and had had experience with a number of different instructors. Therefore, the nine items in the scale which could be adapted for use in advanced schools were administered to 1,046 students in class 44-E at Luke, Williams, and Douglas Advanced Flying Schools. A statistical analysis of the data showed the distribution of instructor ratings on the nine-item scale provided a fairly adequate spread from low to high ratings. This distribution was considerably better than that secured by Lt. Col. Sollenberger at the Primary school level, where the majority of instructors was rated at the favorable extreme on most items.

The reliability of the Advanced school students' ratings of their instructors was also satisfactorily high. When the average of ratings by four of each instructor's students was compared with the average of another randomly selected four, the correlation was 0.78. In interpreting this correlation, two facts should be kept in mind: each instructor was rated by different students, in contrast with the ordinary stituation in which all individuals are rated by the same pair of judges. There may be some tendency for the different students assigned to the same instructor in the same class to influence each other's opinions.

The relatively few favorable ratings by the Avanced-school students should *not* be taken to indicate inferior instruction at this level. It is more likely that the difference was related to the fact that the advanced students knew much more about flying and, hence, were not so inferior to their instructors in this respect; also, they had had experience with more instructors and, hence, had a better standard of comparison against which to evaluate their instructors. This interpretation was corroborated by a later study in which a longer students' rating scale of flying instructors was administered to students in the Transition school at Fort Worth Army Air Field. It was found that the students with instructor experience tended to rate their instructors somewhat lower than did the men who had not been instructors.

This study indicated the desirability of a more complete analysis of instructor ratings by students as one of several potential criteria to be used in instructor selection and for the validation of the instructor selection tests.

Exploratory Work on Supervisor's, Student's, and Self-Ratings of Instructors

Scale for rating of civilian instructors in primary schools by their supervisors.[33]—As one means of validating a battery of instructor

[33] Capt. John T. Dailey was the chief contributor in the development of this scale.

selection tests developed at that unit, personnel of Psychological Research Unit No. 2 constructed a descriptive rating scale on which supervisors rated their instructors. The items included were the 14 items of a 22-item checklist of characteristics considered by supervisors to be most important in the flying instructor. The items of the checklist were selected after interviews with supervisors at three primary schools; it covered the points judged by supervisors to be important characteristics of the good flying instructor. Two over-all items were also included in the rating scale. This descriptive rating scale was administered at five primary schools. Each instructor was rated by two civilian supervisors who were in charge of flights of instructors, usually the flight supervisor and assistant flight supervisor.

An analysis of the results of the administration of the scale showed that, for the most part, the percent of the instructors assigned each descriptive choice on each item was fairly stable from school to school. The entire range of choices on each item was used by the raters and, in general, the distributions of responses were fairly symmetrical. The agreement among the distributions of the various schools was taken as evidence that the descriptive statements tended to have consistent meanings from school to school. It was possible to obtain ratings by two different supervisors of 198 of the instructors at 5 of the schools. The correlation between total scores on the rating scale for two raters was 0.65 thus indicating that the scale had fairly high reliability. The degree of independence of the two ratings of an instructor is not known. Although the ratings were made independently, the supervisors had undoubtedly previously discussed their instructors' proficiency, and the correlation between the two ratings possibly was raised by this pooling of information.

The intercorrelations of the items of the rating scale ranged from 0.26 to 0.90 with the majority of them relatively high. This indicated the presence of a "halo" effect in the ratings as well as a possibly close relationship among the factors measured by the items.

Early forms of rating scales for evaluating basic, advanced, and transition military flying instructors.—The first step in exploring the possibilities of using rating scales to secure criterion data on military flying instructors was taken by Psychological Research Project (Pilot) and consisted in comparing ratings by the instructor's supervisor, his students, and by the instructor himself. It was thought that each of these types of ratings might have its own advantages in dealing with certain areas of the instructor's proficiency and adjustment to his job.

On the basis of the examples of good and bad teaching techniques, the checklists of qualities most desirable in an instructor, and other materials described in the job analysis section of this chapter, Sgt. Daniel J. Grier and Lt. Wallace Nygard constructed two 47-item rating scales. One of these was for students to rate their instructors; the other for instructors to rate themselves. Though the specific form of statement in each scale was adapted to its use, the subject material and alternatives were parallel. A shorter scale of 16 items was constructed for the supervisors. In constructing this scale, an attempt was made to select the 16 items which were most important and on which the supervisors would be most likely to be able to give relevant ratings. The statement of the item was in most cases identical with that used in the student's scale.

The three scales were administered to student officers taking transition training at Fort Worth Army Air Field. A total of 74 supervisors' ratings, 72 self-ratings, and 230 students' ratings was secured. Results on a sample item are presented below.

Distribution of responses			Item
Students	Supervisors	Self	Ability to express himself
Percent	*Percent*	*Percent*	This instructor has:
35	28	19	4. An excellent command of language and never has difficulty in making things clear.
55	62	63	3. A good command of language.
10	10	15	2. A mediocre command of language.
00	00	03	1. A poor command of language resulting in explanations which are usually rather confused.

Statistical analysis of the data derived from this administration indicated that of the three scales, only the student's rating scale showed promise of giving sufficiently reliable results. The reliability of the student's scale when the ratings of an instructor by one student were correlated with the ratings of that instructor by another student was 0.33. In a sample of 44 instructors, the reliability increased to 0.44 when the ratings of two students were correlated with the ratings of two other students of the same instructor. This is approximately the increase in reliability that is predicted by the Spearman-Brown formula.

Difficulty was experienced in obtaining enough ratings of instructors by supervisors to do an adequate statistical analysis, but the reliability of the whole scale (−0.09) and of the individual items proved to be so low that this form of supervisor's scale was abandoned as impractical. The low reliabilities of the items of the scale, half of which were negative and only one of which was substantially positive, may be interpreted in the light of the fact that most supervisors are unable to ride with the instructor and student in the teaching situation, and thus do not have an opportunity to make good observations of the instructor's teaching ability. This consideration and the low reliabilities obtained led to the decision to secure ratings of instructors by supervisors on only two items: (1) Over-all flying proficiency, and (2) over-all value as an instructor.

Since it was impossible to correlate one group of ratings on the self-rating scale against another, total scores on the self-rating scale were correlated with total scores of the students' ratings of the same instructor. The average total score of 4 students' ratings of an instructor on a sample of 44 instructors was correlated with the instructors self-ratings and gave a coefficient of 0.12. It was judged that revision of the scale would not substantially raise the correlation so no further effort was expended in developing this scale. Further work along this line was limited to securing the one rating on which it was thought the instructor's own opinion would be most significant, namely, a rating of his liking for the job of flying instructor.

Since students have the best opportunity to observe the instructor's teaching and since the results on the student's rating scale were most promising, further work was concentrated on developing and refining this scale.

DEVELOPMENT OF STUDENT'S RATING SCALE [a]

Refining and Shortening of Student's Rating Scale

Work on the development of the student scale was carried on at the Advanced and Transition school level because it was believed that the students with more experience would be better able to rate their instructors and since it was known that such students give a better distribution of responses than those in Primary schools. Furthermore, it was known that returned combat pilots would first be instructing at the advanced and transition levels, and the scale was to be used primarily for collecting data on these men for validating the battery of instructor selection tests being administered at redistribution stations.

The reliability of each of the items in the 47-item student's rating scale was determined. For the purpose of determining item reliabilities, all available pairs of students' ratings secured at Fort Worth Army Air Field were used, a total of 103 randomly matched pairs. The reliability of the individual items ranged from 0.08 to 0.43. Thirty-four of the items had a reliability above 0.10.

The 34 items with a reliability above 0.10 were selected for further screening and refinement. Where it seemed desirable, the wording was changed to improve the clarity of the alternatives and the distribution of responses. In order to refine this scale further, it was desired to eliminate those items which either (1) had the lowest reliability, or (2) had the most overlapping with other items in the scale. Therefore, the scale was administered to 364 students in the four-engine Transition school at Fort Worth Army Air Field and the reliability of each item and its intercorrelation with all of the other items was calculated. The reliabilities were calculated from 158 randomly matched pairs of students who were taught by the same instructors;

[a] The major work in the development and use of the rating scale was done by Sgt. Daniel J. Orier. Additional contributors were: S/Sgt. C. P. Gershenson and Sgt. James Stratton (IBM tabulations and statistics); Lt. J. W. Nygard (preliminary work); Cpls. George Fassnacht, Palmer Tibbetts, Alvin Roitblatt, Robert Smith, William Wiesenberg, Leonard Nieder, and Ralph Wurst (machine calculators). The project was under the supervision of 1st Lt. William E. Galt.

the intercorrelations were calculated from the entire sample of 364 students' ratings.

It was found that the changes in wording had generally improved the distribution of the ratings. Discarding the most unreliable items of the 47-item scale had not only shortened it but also increased the reliability of the total scale from 0.33 to 0.36. The intercorrelations of the items in this scale ranged from 0.04 to 0.62. They are not presented in detail at this point, however, since the intercorrelation among the 18 items of the final scale were computed on a much larger sample and are presented later. The reliability of each item of the scale is presented in table 14.4.

TABLE 14.4.—*Reliabilities of each item of 34-item form of student's rating scale of flying instructors N=168*

Item No.		First rating		Second rating		*r*
		Mean	SD	Mean	SD	
1	Analysis of errors	2.22	0.85	2.11	0.83	0.13
2	Interest in job of instructing	2.21	.63	2.17	.67	.32
3	Ability to demonstrate maneuvers	2.48	.68	2.46	.76	.10
4	Effectiveness of on-the-ground teaching	2.11	.69	2.00	.64	.19
5	Use of "chewing," swearing, etc., in teaching	2.42	.72	2.41	.76	.39
6	Encourages students to ask questions	2.62	.67	2.52	.63	.27
7	Riding the controls	2.10	.77	1.98	.78	.08
8	Ability to express himself	2.12	.65	2.08	.73	.16
9	Reaction to improvement in flying	2.25	.77	2.30	.79	.26
10	Personal interest in the individual student	1.85	.93	1.87	.92	.33
11	Use of flight time	2.40	.76	2.32	.78	.24
12	Emphasizes understanding of airplane, etc	2.09	.72	2.00	.82	.11
13	Ability to adapt teaching methods to students	1.96	.82	1.87	.95	.16
14	Encouragement to study flying outside plane	2.25	.90	2.23	.87	.25
15	Patience and self-control	2.40	.77	2.42	.77	.28
16	Effectiveness in putting flying training over	1.73	.83	1.70	.81	.25
17	Willingness to repeat instructions	2.44	.72	2.40	.73	.26
18	Emphasizes teaching through flair for dramatic	1.37	.92	1.30	.85	.07
19	Stimulation of desire to fly	2.01	.81	1.92	.84	.22
20	Teaching how to perform in emergencies	2.06	.83	2.15	.78	.32
21	Correcting an error during flight	2.37	.70	2.30	.72	.25
22	Relating maneuvers to combat	1.55	.78	1.61	.77	.20
23	Does not tell student the "what," "why" and "how"	2.33	.71	2.31	.71	.18
24	Coordinating of flight line, ground instructing	1.68	.72	1.68	.81	.20
25	Does not tell students how to correct errors	2.46	.71	2.37	.76	.29
26	Precision of maneuvers	1.97	.57	1.92	.58	.60
27	Maintenance of progress on schedule	1.60	.78	1.63	.78	.24
28	Voluntary extra help	1.89	.96	1.73	.97	.21
29	Personal appearance and promptness	2.23	.73	2.28	.71	.07
30	Maintaining proper teacher-student discipline	2.06	.71	2.04	.79	.01
31	Use of technical terms and expressions	2.44	.77	2.42	.78	.17
32	Interest in teaching instrument flying	2.24	.65	2.20	.52	.13
33	Emotional upset of student due to way of teaching	2.35	.89	2.32	.92	.15
34	Flying skill (independent of instructor ability)	1.92	.65	1.91	.71	.33

Items which intercorrelated most highly with the other items, had the lowest reliability, the poorest distribution, and which were judged to be least important were eliminated, and two new items were added. This revised and shortened scale of 25 items was administered to upper class students at an Advanced single-engine school, an Advanced two-engine school, and a Transition school (Class 45–C at Aloe, Enid, and Liberal respectively) to a total of 364 students.

Reliabilities were calculated separately for the three schools and then combined by the Fisher z-transformation. They are based on 182 randomly matched pairs of students. The reliability of the items ranged from −0.05 to 0.45. The two new items, "confidence in this airplane" and "conduct as an officer," had reliabilities of 0.10 and 0.35 respectively. The reliability of the total scale was increased to 0.49. In the use of this scale as a criterion for validating instructor selection tests and other research, ratings of about four students were averaged. From the Spearman-Brown formula it can be estimated that averaging ratings of four students should increase the reliability of the scale to 0.79.

At this administration, students were requested to indicate on their score sheets the five items they considered most important in the scale, i. e., the items dealing with the five qualities most important for flying instructors to possess. They were also asked to indicate any items they felt should be omitted from the scale either because they were unimportant or because the student did not have the proper basis to rate his instructor on this quality. The six items that were most often chosen as being among the five most important qualities were presented in table 14.2 in the section on job analysis. An example of a suggested omission was that the item on "stimulation of the student's desire to fly" should be omitted, as this quality in an instructor was considered to be unimportant at the Advanced and Transition school level.

On the basis of these comments and suggestions, together with distributions of items, item reliabilities and intercorrelations, the scale was further cut to the 18 most important and reliable items. It was now considered suitable as a tool to be used in the validation of the instructor selection tests.

Reliability of Final Form of Student's Rating Scale

The final 18-item form of the student's rating scale was designed to secure ratings on instructors by advanced and transition school students. Exploratory results had shown that such students with more experience would be better able to rate their instructors and exploratory results had confirmed this hypothesis.

The value of the scale was tested out at all levels of training and the scale was administered to all students in Advanced and Transition schools for four successive classes beginning with class 45-D, and to all students in basic schools for two successive classes beginning with class 45-H, for the purpose of obtaining criterion data against which to validate the battery of instructor selection tests administered to returned combat pilots at redistribution stations.

Distributions of responses on each item of the scale were tallied for 210 primary students, 355 basic students, 387 advanced students, and 284 transition students. The percentage of students selecting each alternative on each item is presented in table 14.5. It will be seen that in general the distributions for primary students are not as satisfactory as for students at the more advanced levels of training. They are considerably better, however, than the distributions reported by

Lt. Col. R. T. Sollenberger on a 9-item scale which he administered to primary students at Stamford Flying School, Arledge Field, Stamford, Tex., in September 1943.

TABLE 14.5.—*Reliability of total score and distribution of responses of Elementary, Basic, Advanced, and Transition students on 18-item student's rating scale*

Reliability[1]		Elementary	Basic	Advanced	Transition	Item and cue statements
		r=.27 N=68	r=.60 N=233	r=.36 N=782	r=.55 N=613	
Distribution						
Item No.	Statement No.	N=210	N=355	N=387	N=284	
		Percent	*Percent*	*Percent*	*Percent*	1. *Analysis of errors.*—This instructor—
1	1	1	2	1	1	Merely says the maneuver is wrong but doesn't say why.
	2	3	8	5	7	Is somewhat indefinite about what is wrong and how to correct it.
	3	34	43	43	53	Gives a fairly good idea of what is wrong and how to correct it.
	4	61	46	51	39	Locates the specific cause of the difficulty and tells exactly how to correct it.
						2. *Interest in job of instructing.*—This instructor—
2	4	25	13	21	18	Is extremely enthusiastic about his job.
	3	53	44	53	46	Seems very interested in his job.
	2	20	33	21	33	Is interested in his job but not enthusiastic.
	1	2	10	5	3	Shows slight interest in instructing.
						3. *Ability to demonstrate maneuvers in the air.*—This instructor's demonstrations of flying are—
3	4	65	36	54	40	An excellent example.
	3	28	40	35	38	A very good example.
	2	6	16	9	16	A good example.
	1	1	8	2	6	A satisfactory example.
						4. *Effectiveness of on-the-ground teaching.*—In general, the discussion which this instructor gives before and after the flight—
4	1	4	5	1	6	Could be greatly improved upon.
	2	12	9	7	11	Is somewhat inadequate.
	3	42	50	34	43	Is fairly good.
	4	42	36	58	39	Is thorough and excellent.
						5. *Use of "chewing," swearing, ridicule, or sarcasm in teaching.*—This instructor uses such methods—
5	4	74	79	83	82	Only when absolutely necessary.
	3	13	15	10	13	Occasionally when unnecessary.
	2	11	5	5	3	A great deal of the time.
	1	2	1	2	2	Almost all of the time.
						6. *Encourages students to ask questions.*—This instructor—
6	1	0	1	1	1	Made me feel I should not ask questions.
	2	12	10	5	16	Neither encourages nor discourages questions.
	3	38	30	36	37	Seemed to want me to ask questions if I did not understand.
	4	50	59	58	46	Definitely encourages me to ask questions.
						7. *Riding the controls.*—This instructor takes over—
7	4	61	55	70	56	Only when lives or planes are endangered.
	3	28	33	22	23	Occasionally when it seems unnecessary.
	2	8	9	7	15	Considerably more often than necessary.
	1	3	3	1	5	On the slightest provocation.
						8. *Ability to express himself.*—The explanations which this instructor gives—
8	1	1	2	1	1	Tend to be rather confused.
	2	4	6	5	8	Are sometimes hard to understand.
	3	50	52	51	56	Are usually fairly clear.
	4	45	40	43	35	Are always perfectly clear.
						9. *Reaction to improvement in flying.*—When I show improvement this instructor—
9	4	40	37	45	45	Always lets me know I have done well.
	3	45	43	40	42	Sometimes lets me know I have done well.
	2	9	14	11	10	Rarely lets me know I have done well.
	1	6	6	4	4	Almost never lets me know I have done well.

See footnote at end of table.

TABLE 14.5.—*Reliability of total score and distribution of responses of Elementary, Basic, Advanced, and Transition students on 18-item student's rating scale*—Con.

Reliability [1]		Elementary	Basic	Advanced	Transition	Item and cue statements
		$r=.27$ $N=68$	$r=.60$ $N=233$	$r=.36$ $N=782$	$r=.55$ $N=613$	
Distribution		$N=210$	$N=355$	$N=387$	$N=284$	
Item No.	Statement No.					
		Percent	*Percent*	*Percent*	*Percent*	10. *Personal interest in the individual student.*—This instructor takes—
10	1	7	10	10	7	Little personal interest in the individual student.
	2	22	22	21	22	Some personal interest.
	3	35	44	45	40	Quite a bit of personal interest.
	4	36	24	24	31	A great personal interest.
						11. *Use of flight time.*—This instructor—
11	4	70	48	55	46	Makes exceptionally good, well-planned use of time.
	3	28	44	39	40	Makes good use of the time.
	2	2	8	5	11	Makes fair use of the time.
	1	0	1	1	3	Makes poor use of time, tends to waste it.
						12. *Emphasizes an understanding of the airplane, flying safety, and flying regulations.*—This instructor teaches this—
12	1	5	4	3	5	With incidental emphasis while teaching flying procedure.
	2	17	21	26	24	With considerable emphasis.
	3	33	32	30	38	With great emphasis.
	4	45	43	41	33	As one of the most important features of flying.
						13. *Conduct as an officer.*—The behavior of this instructor on the ground and in the air in all respects—
13	4	45	52	56	51	Sets an excellent example for the student.
	3	40	36	33	35	Sets a good example for the student.
	2	13	12	8	12	Sets a fairly good example for the student.
	1	2	1	3	2	Sets a somewhat poor example for the student.
						14. *Ability to understand the problems of the individual student and adapt teaching methods to him.*—This instructor—
14	4	47	36	43	36	Always seems to adapt his methods to my problems.
	3	37	45	38	39	Often seems to adapt to my problems.
	2	10	13	14	16	Sometimes seems to adapt to my problems.
	1	7	6	5	8	Rarely seems to change his methods to fit my problems.
						15. *Encouragement to study flying outside the airplane (use of flight manuals, technical orders, etc.)*—In getting students to use these materials this instructor—
15	1	20	5	5	6	Pays little attention to it.
	2	10	9	13	14	Is relatively ineffective.
	3	44	47	52	55	Is fairly effective.
	4	27	39	30	25	Is very effective.
						16. *Confidence in this airplane.*—This instructor—
16	4	91	67	76	67	Displays absolute confidence in his ability to meet all emergencies in this airplane.
	3	8	27	20	27	Seems quite confident in his ability to meet all emergencies in this airplane.
	2	0	4	4	5	Seems fairly confident in his ability to meet all emergencies in this airplane.
	1	1	2	0	1	Seems doubtful about his ability to meet all emergencies in this airplane.
						17. *Patience and self-control.*—This instructor is—
17	1	2	4	3	4	Extremely impatient.
	2	9	13	8	11	Somewhat impatient.
	3	31	29	25	26	Rather patient.
	4	58	54	64	60	Extremely patient.
						18. *Effectiveness in putting flying training across.*—This instructor's over-all effectiveness is—
18	4	43	22	40	29	Exceptionally high—he approaches the ideal as a teacher.
	3	46	49	45	43	Very high—a very good teacher.
	2	10	22	12	22	Good—an adequate teacher.
	1	1	8	3	5	Limited—a fair teacher.

[1] The reliability figure is based on the correlation of the rating of an instructor by 1 student with the rating of the same instructor by another student in the same class.

The reliability of the student's rating scale was determined in two different ways—intraclass and interclass reliability. The latter of these two is the more rigorous test of reliability since there is little possibility for students in successive classes to influence each other in the evaluation of their instructor.

Agreement among students in same class.—The reliability of the total scale when two students in the same class rating the same instructor are randomly paired and the rating of one student is correlated with the rating of the other student is presented in table 14.4. The reliability of the scale was computed from data gathered at three primary schools, four basic schools, seven advanced schools (single-engine and two-engine), and nine transition schools. The reliability of the scale for each school was computed separately and the reliabilities for all schools at a given level of training were combined by Fisher's z-transformation. The reliabilities were as follows for the different levels of training: Primary ($N=68$) 0.27; Basic ($N=233$) 0.60; Advanced ($N=782$) 0.36; Transition ($N=613$) 0.55. Use of the Spearman-Brown formula for predicting the reliability of the scale when 4 students rate the same instructor raises the reliability to 0.60 for primary schools, 0.86 for basic schools, 0.69 for advanced schools (single-engine and two-engine combined), and 0.83 for transition schools. In general, the reliability of the scale at all levels of training was satisfactory. Although the reliability for primary students was not as high as it was for students at the more advanced levels of training, probably because of the inexperience of the students, the scale is still useful as a measure of the instructor's ability, particularly if at least four independent ratings of an instructor are combined.

Agreement among students in different classes.—Administration of the student's rating scale to four successive classes in advanced and transition schools made it possible to make a final check on the reliability of the scale by comparing scores on a single instructor rated by students from different classes. This decreased the possibility of a lack of independence in ratings, since students of one class would not be likely to influence the judgments of students in another class in rating the instructors.

A total of 281 instructors were found in the data who had been rated by students from two classes. An average of 3.25 students in each class rated each instructor. There were 159 Advanced school instructors and 122 transition school instructors in the group on which the class-to-class reliability of the scale was computed. The total scores on the scale were first normalized and converted into standard scores ranging from 1–9. Using these standard scores, the class-to-class reliability of the scale was then computed. The reliability of the scale computed in this manner was 0.36.

Comparison of two types of reliability.—In order to determine the amount of shrinkage in class-to-class reliability, the percentage of the

277 instructors coming from each school was computed and the intraclass reliability of that school was converted to a z-transformation. It was necessary to drop 4 of the original 281 instructors since they were from schools which had not provided sufficient material upon which to calculate intraclass reliabilities. The intraclass reliabilities of all schools were then combined on the basis of having the percentage of instructors from each school the same as in the sample of 277 on which the class-to-class reliability was computed. This gave an intraclass reliability of 0.49 when the rating of an instructor was made by two students and the rating of one student was correlated with that of the other. Use of the Spearman-Brown formula for predicting the reliability of the scale when an average of 3.25 students rate the same instructor (as in the class-to-class study) raises the intraclass reliability to 0.75. The comparative shrinkage from intraclass to interclass reliability is, then, from 0.75 to 0.36. Two factors are probably responsible for this shrinkage in reliability when the more rigorous test of class-to-class reliability is applied: (1) The students do not have a chance to influence each other in the rating of an instructor and (2) the instructor may himself change in his teaching from class to class.

The class-to-class reliability of the over-all item of the student's rating scale, "Effectiveness in putting flying training across," was 0.23. Since the class-to-class reliability of total scores on the scale was 0.36, the total score has a significantly higher reliability (1 percent level or better) than the over-all item alone. Furthermore, the over-all item probably had a higher reliability in this study where the instructor was rated on it after he had been rated on the other 17 items of the scale than it would have had if it had been the only item on which the instructors were evaluated.

Intercorrelation Among Items on Student's Scale

Using students' ratings secured on 1,141 advanced and transition school instructors, the intercorrelations of the items of the scale were computed. An average of 3.12 students rated each instructor, and the average rating of each instructor on each item was used in calculating the item intercorrelations. The total scores were normalized and converted into standard scores ranging from 1 to 9 before the intercorrelations were run. The intercorrelation of each item with every other item and with total score, supervisor's rating of flying ability, and supervisor's rating of instructor proficiency are presented in table 14.6. It will be seen that the intercorrelations among items range from 0.12 to 0.76 which indicates that although the different qualities being rated in the instructor are related to each other, many of them make separate contributions to the total evaluation of an instructor.

TABLE 14.6.—*Intercorrelations among items of 18-item student's rating scale and supervisor's over-all ratings*

N=1141; SE of a zero correlation=0.03

Item No.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total Score	Supervisor Flying	Supervisor Teaching	Mean	S. D.	Item No.
1	Analysis of errors		39	47	54	32	46	30	49	34	41	45	42	41	54	38	35	39	59	62	02	04	6.84	1.15	1
2	Interest in instructing			28	41	19	34	13	31	28	42	33	35	38	33	[illegible]	22	25	48	53	03	09	5.46	1.57	2
3	Ability to demonstrate maneuvers				43	32	34	37	41	26	33	42	40	44	45	33	45	36	61	61	07	09	6.39	1.56	3
4	On-the-ground teaching effectiveness					27	55	29	46	31	45	45	44	38	44	46	26	34	57	61	01	04	6.59	1.47	4
5	Use of "chewing," swearing, sarcasm, etc.						37	34	28	31	30	20	15	44	46	12	15	67	44	46	−04	−02	7.47	1.22	5
6	Encourages students to ask questions							25	35	34	42	36	38	38	41	38	23	39	44	53	−03	01	6.87	1.24	6
7	Riding the controls								25	29	25	28	15	30	40	18	34	41	37	44	03	03	7.01	1.50	7
8	Ability to express himself									25	38	35	40	39	49	33	30	34	52	57	04	07	6.60	1.05	8
9	Reaction to improvement in flying										45	19	26	30	44	27	15	32	35	53	−03	−01	6.37	1.55	9
10	Personal interest in the individual student.											33	40	43	52	34	25	41	50	68	−04	−03	5.44	1.65	10
11	Use of flight time												41	40	34	39	31	23	51	50	04	09	6.89	1.22	11
12	Emphasizes an understanding of the airplane, etc.													37	35	45	21	24	44	54	−03	02	6.12	1.61	12
13	Conduct as an officer														46	33	27	51	53	61	−01	05	6.89	1.30	13
14	Ability to understand the problems of the individual.															33	29	54	60	70	00	00	6.13	1.50	14
15	Encouragement to study flying outside the airplane.																24	17	42	50	−02	07	6.20	1.52	15
16	Confidence in this airplane																	21	37	42	07	06	7.39	0.99	16
17	Patience and self-control																		50	50	−03	−02	6.82	1.50	17
18	Effectiveness in putting flying training across.																			76	08	11	6.01	1.52	18
	Total Score																				06	08	5.02	1.95	Tot. Sc.
	Supervisor's rating of flying ability																					70	2.05	0.66	Sup. Fly.
	Supervisor's rating of teaching proficiency.																						2.05	0.68	Sup. Tch.

The scores used in computing the intercorrelations were the average of ratings by one or more students. Each of the 1141 instructors was rated by an average of 3.1 students.

In general, the items which one would logically expect to be related had the highest intercorrelations and those which one would expect to be unrelated gave low intercorrelations. For example, "use of 'chewing,' swearing, ridicule, etc., in teaching" correlated 0.67 with "patience and self-control" and 0.46 with "ability to understand the problems of the individual student" but only 0.15 with "confidence in this airplane." Since the signs of all coefficients have been adjusted so that a positive one means a positive relationship between good qualities, this means that the instructors who did not swear at their students were considered to be the most patient. Fortunately, a strong "halo" effect, causing the students to give indiscriminately higher ratings on all items to the instructors they liked, did not seem to be present. This is shown by the fact that the instructors who were rated highest on not swearing were not necessarily given a higher rating on their confidence in the airplane.

In order to secure a rough estimate of which items of the rating scale were related to each other and what more generalized factors the rating scale measured, profiles presenting graphically the correlation of each item with every other item of the rating scale were constructed. An inspection of these graphs showed that the items could be sorted into the following groups on the basis of similarity of their profiles.

Group I

Patience and self-control.

Ability to understand the problems of the individual student and adapt teaching methods to him.

Use of "chewing," swearing, ridicule, or sarcasm in teaching.

Riding the controls.

Group II

Interest in job of instructing.

Encouragement to study flying outside the airplane (use of flight manuals, tech orders, etc.).

Emphasizes an understanding of the airplane, flying safety, and flying regulations.

Use of flight time.

Group III

Ability to express himself.

Analysis of errors.

Encourages students to ask questions.

Ability to demonstrate maneuvers in the air.

The general factor underlying the items in group I might be considered as a personality factor involving patience and social adaptability. The general factor underlying the items in group II might be described as enthusiasm for and a systematic approach to the job of instructing. The general factor which best describes the items in group III is analytic and verbal ability. The other items of the scale did not definitely fit into any of the above groupings. The profiles of the items in the three groups are presented in figures 14.1 through 14.3.

Correlation Between Student's and Supervisor's Ratings

A study was made of the relationship of the three measures later used in validating the battery of instructor selection tests administered to returned combat pilots at redistribution stations. These three measures were: (1) Student's average total rating of the instructor on the 18-item student's rating scale. There was an average of 3.12 students rating each instructor; (2) supervisor's over-all rating of the flying ability of the instructor; (3) supervisor's over-all rating of the instructing proficiency of the instructor. The two latter ratings were on a 3-point scale (in upper 25 percent of instructors, in middle 50 percent of instructors, or in lower 25 percent of instructors in the school).

Intercorrelations computed between these three types of measures showed that there was little relationship between supervisors' and students' ratings of instructor proficiency ($r=0.08$). There was also little relationship between supervisors' ratings of flying ability and students' ratings of instructor proficiency ($r=0.06$). There was, on the other hand, a close correspondence between supervisors' ratings of flying ability and their ratings of instructor proficiency ($r=0.70$). These correlations are all based on 1,141 cases. Similar high correlations between supervisors' ratings of an instructor's flying ability and instructing proficiency have been reported at Primary flying schools by Psychological Research Unit No. 2. It is evident that supervisors, who often have little opportunity to observe the instructor while he is instructing in the air, base their estimate of his instructional ability almost solely upon what they can observe, namely, his flying proficiency. Students and supervisors differed markedly in their appraisal of an instructor.

ANALYSIS OF BACKGROUND FACTORS RELATED TO INSTRUCTOR SUCCESS

In connection with the collection of criterion data for the validation of instructor selection tests, personal data items were secured on instructors who were subsequently rated by their students for instructor proficiency on the 18-item scale and by supervisory personnel on their over-all flying ability and over-all instructor proficiency. This was done for two reasons: (1) It was necessary to study the effect of these factors on the ratings in order to know whether or not they must be controlled in validating the tests. For example, if the men who had had only a small amount of instructing experience were

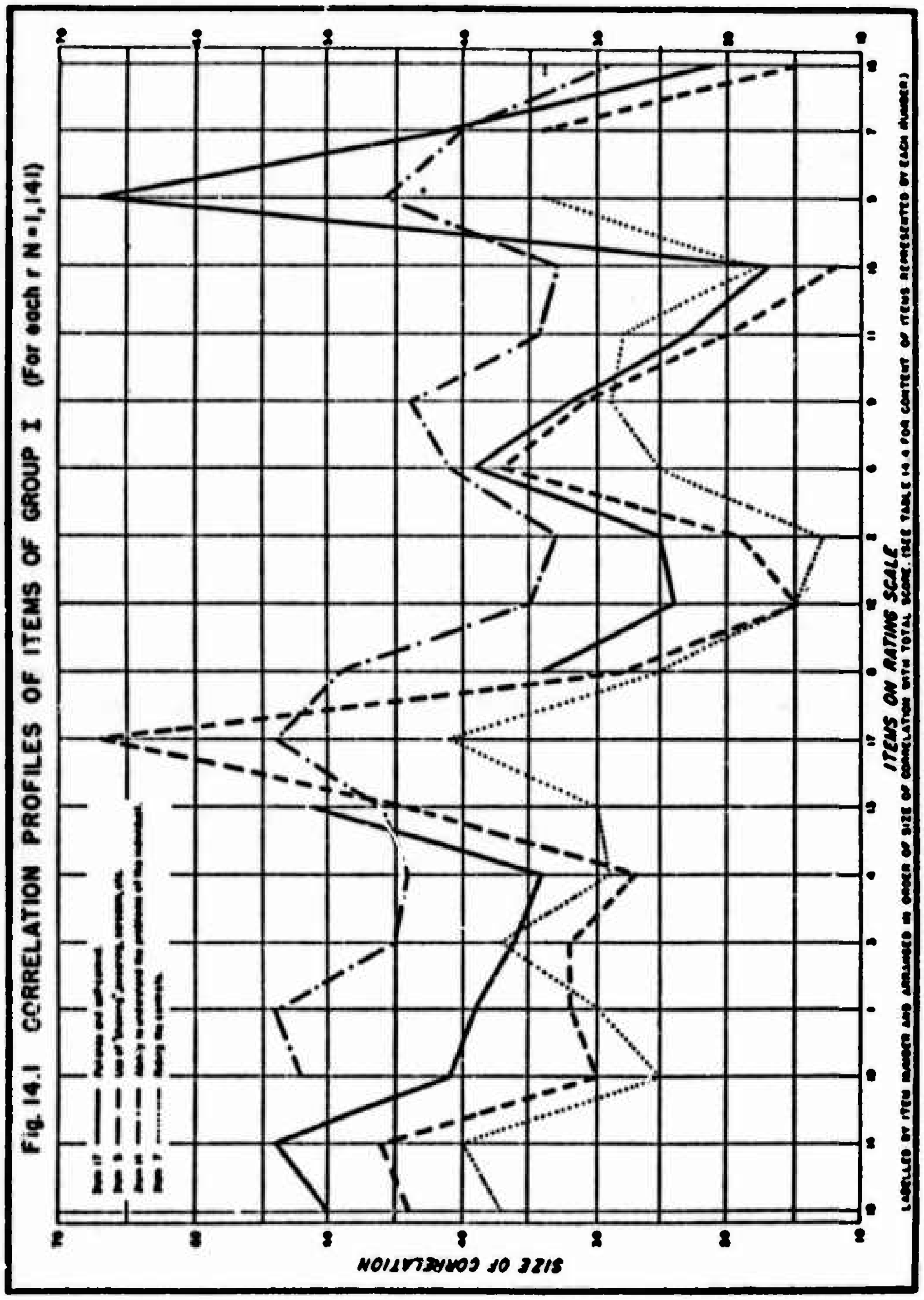
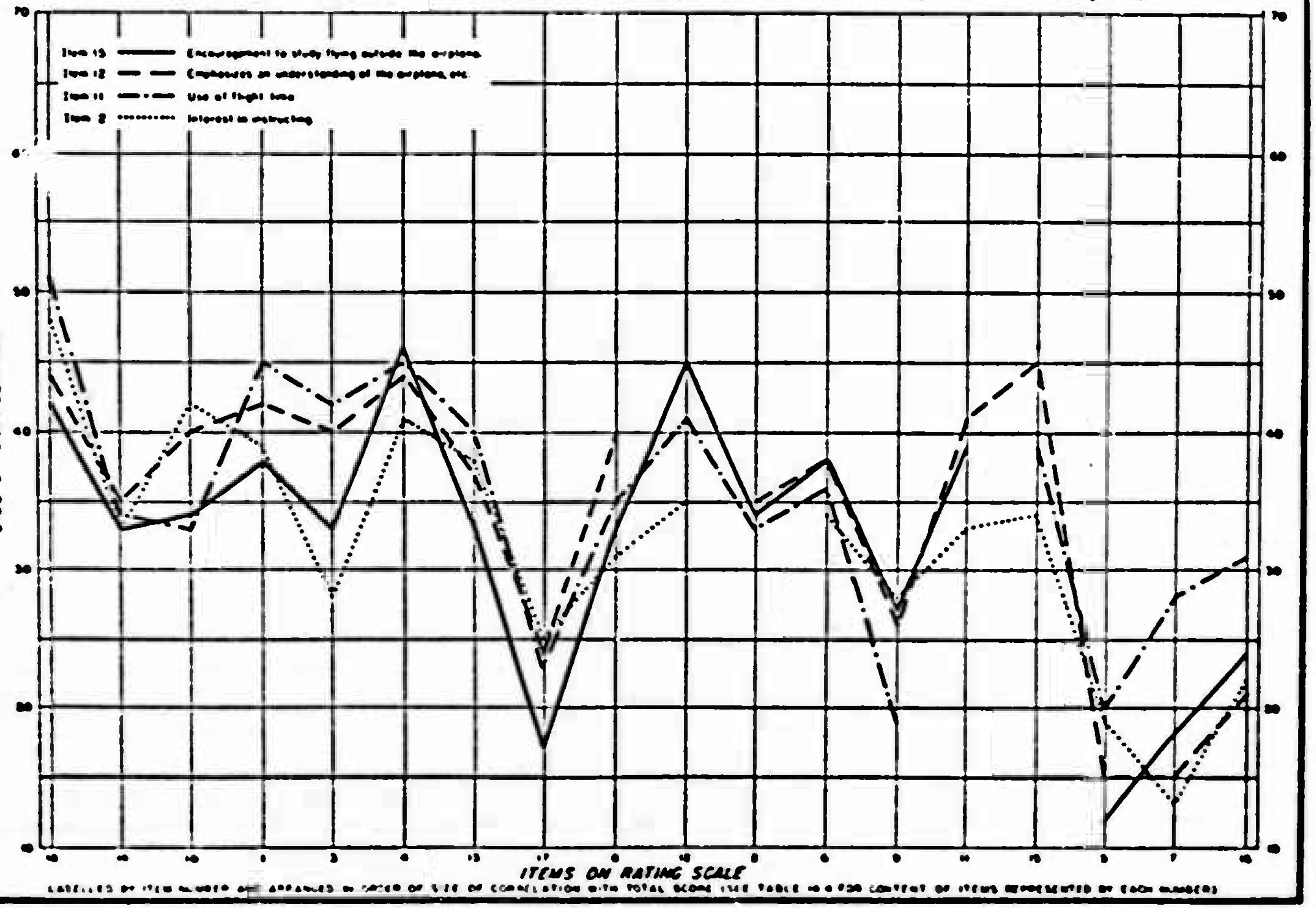


Fig. 14.1 CORRELATION PROFILES OF ITEMS OF GROUP I (For each r N=1,141)

Fig. 14.2 CORRELATION PROFILES OF ITEMS OF GROUP II (For each r N=1,141)



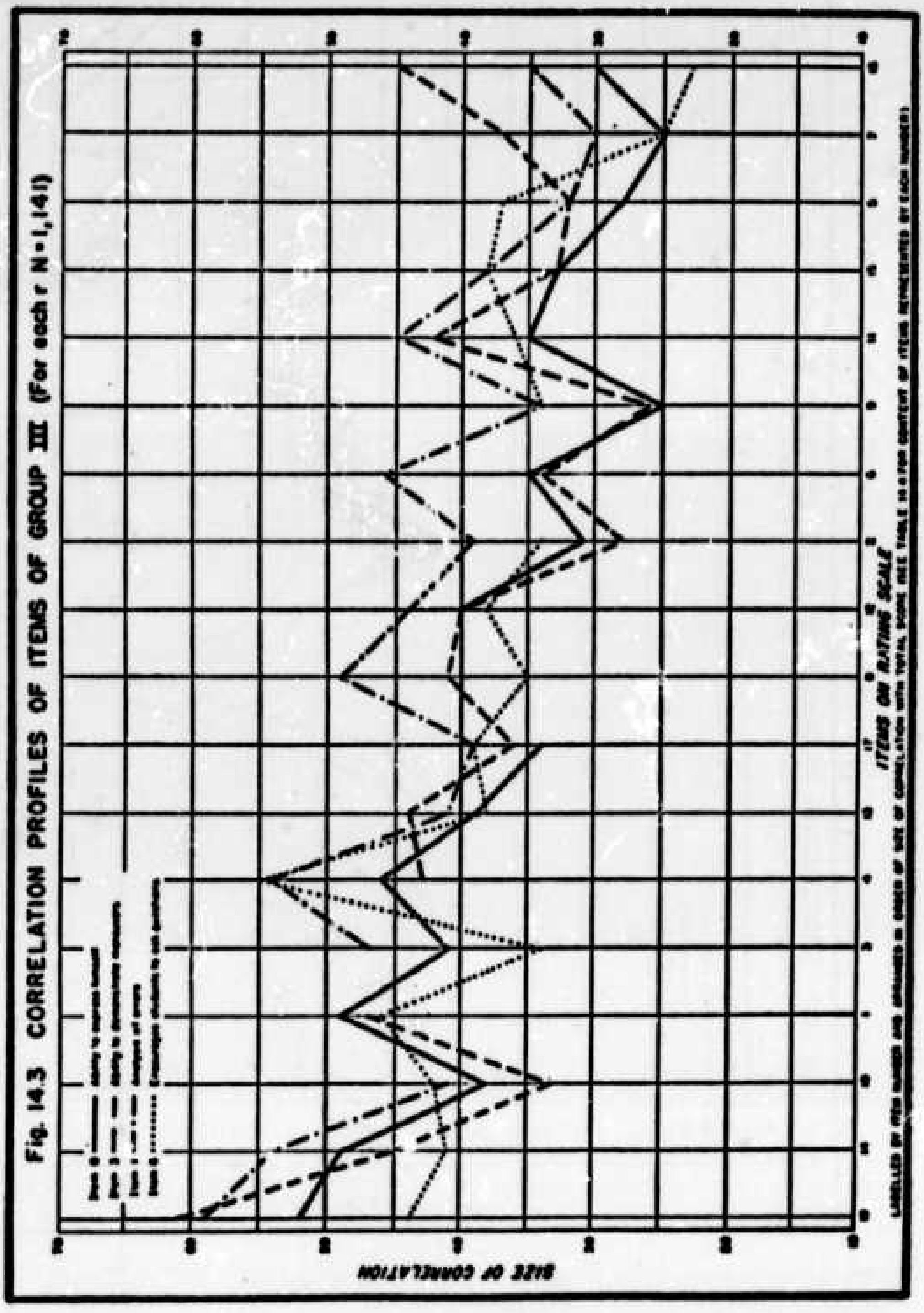


Fig. 14.3 CORRELATION PROFILES OF ITEMS OF GROUP III (For each r N = 1,141)

found to get markedly lower ratings, this group would have to be treated separately in validating the tests; (2) the data in themselves were valuable information. In addition to being necessary for deciding which groups may be considered homogeneous enough to be combined in the validation study, it was believed that the determination of the relationship of these factors to the proficiency of an instructor would be of considerable value to the Training Command. For example, it should reveal whether or not the common supposition is true that instructors improve with experience at first, and then go stale and get poorer the longer the period of instructing.

In addition to identifying data such as name, army serial number, etc., pertinent information of a biographical, or personal, nature were secured on a total of 1,284 Advanced single-engine, Advanced two-engine and Transition school instructors. Of this number, 404 were returned combat pilots and the remaining 880 had had no combat experience.

Method Used in Analysis of Data

The three measures described in page 14–36, i. e., students' ratings of teaching ability, supervisors' rating of teaching ability and supervisors' rating of flying proficiency, were used as criteria in evaluating the personal data items. The distributions of responses on the personal data items were studied in relation to the scores made on these measures.

The average total scores on the 18-item student rating scale were first normalized and converted into standard scores ranging from 1 to 9. The instructors were grouped according to the various categories of each personal data item, for example, according to the number of months instructing experience they had had at the school. For each of the groups the mean of the converted standard scores on the 18-item student's rating scale was computed. The means of the supervisors' ratings of teaching and of flying ability were also computed for each group. Graphs were then constructed showing the relationship between these mean scores and the various divisions of each biographical item.

A study of the graphs and of the means for the various divisions of the personal data items showed that there was no significant trend in the relationship of students' or supervisors' ratings to the following ones:

1. Previous experience as flying instructor:
   a. Type of school.
   b. Type of plane.
2. Logged hours of flying—nonmilitary.
3. Type of instructing being done at present.
4. Number of combat hours.
5. Type of plane flown in combat.

6. Redistribution station to which returned from combat.
7. Date of arrival at redistribution station.

These items were therefore not further analyzed. The data on the remaining biographical items gave promise of showing significant differences between the means of the various subgroups and therefore were tested by the analysis of variance technique. The instructors were divided on the basis of their scores on each biographical item into two or three groups with approximately an equal number of cases in each group. These groupings were then treated by the analysis of variance technique to determine whether there was any relationship between the biographical item and the rating. The F-ratio test was used to determine the probability of obtaining, by chance, differences among the groups as large as those observed. Two of the items thus tested failed to differentiate between the subgroups at the 5 percent level of significance on either students' ratings, supervisors' ratings of teaching ability or supervisors' ratings of flying proficiency. These items were: training at a central instructors school and experience as a nonflying instructor. All other items tested showed a reliable difference on one or more of the three criteria.

After the relationship between the personal data items and the total score on the student's rating scale had been determined, the relationship between certain selected personal data items and each of the 18 items of the scale was computed. This more detailed analysis was done on those personal data items which had been found to be significantly related to the total score on the student's rating scale. To these were added a few personal data items which seemed likely to yield significant relationships with the individual items on the student's scale even though they were not related to the total score.

Results of Analysis of Biographical Factors

The results on those personal data items which were selected for the more detailed analysis are presented in table 14.7. The results for the other personal data items which were analyzed in less detail are presented in table 14.8. The relationships of the most significant personal data items to the total score on the student's scale and to the two supervisors' ratings are presented graphically in figures 14.4 to 14.12.

The main findings on each of the personal data items which yielded significant results were as follows:

Length of time as flying instructor.—All of the items of the student's rating scale on which the instructors with more months of teaching at the same station were rated as significantly better than those with less months, were items which dealt with the technical aspects of instructing. They were: confidence in this airplane; ability to demonstrate maneuvers; analysis of errors; ability to express himself; on-the-ground teaching effectiveness; use of flight time; emphasis on

TABLE 14.7.—*Relationship of selected items of personal data to each item in student's rating scale of teaching proficiency and to supervisor's ratings*

	Total	Month's instruction at this station			Combat service			Rank			Age			Most education completed			Liking for job of flying instructor			
		Less than 5 months	5 months or more	Level of significance	Nonreturnee	Combat returnee	Level of significance	Flight officer—2d lieutenant	1st lieutenant or higher	Level of significance	Less than 24	24 or older	Level of significance	Through high school	1 year college or more	Level of significance	Indifferent—dislike	Very much—fairly much	Level of significance	
Number of cases	1,284	543	741	......	880	404	......	551	733	......	506	778	......	617	667	......	378	906	......	....
Rating scale items	*SX* [1]	*M*	*M*	*Percent*	*M*	*M*	*Percent*	*M*	*M*	*Percent*	*M*	*M*	*Percent*	*M*	*M*	*Percent*	*M*	*M*	*Percent*	
1. Analysis of errors	61,635	6.74	6.89	5	6.84	6.81	......	6.89	6.78	......	6.81	6.84	......	6.79	6.86	......	6.77	6.85	......	1
2. Interest in instructing	41,316	5.36	5.51	......	5.50	5.32	......	5.50	5.41	......	5.39	5.48	......	5.37	5.52	......	5.06	5.61	1	2
3. Ability to demonstrate maneuvers	55,457	6.17	6.53	1	6.39	6.35	......	6.40	6.37	......	6.42	6.35	......	6.36	6.40	......	6.41	6.37	......	3
4. On-the-ground teaching effectiveness	58,450	6.48	6.66	5	6.61	6.46	5	6.66	6.53	......	6.54	6.61	......	6.56	6.60	......	6.46	6.64	5	4
5. Use of "chewing," sarcasm, swearing, etc	73,217	7.57	7.36	[1] 1	7.36	7.63	1	7.36	7.52	5	7.34	7.52	5	7.44	7.45	......	7.42	7.46	......	5
6. Encourages students to ask questions	62,477	6.89	6.85	......	6.85	6.91	......	6.95	6.81	[1] 5	6.91	6.84	......	6.86	6.87	......	6.86	6.87	......	6
7. Riding the controls	65,616	7.00	6.97	......	6.96	7.03	......	7.03	6.95	......	6.95	7.00	......	6.96	7.00	......	7.05	6.95	......	7
8. Ability to express himself	57,051	6.48	6.65	1	6.62	6.48	5	6.60	6.56	......	6.56	6.59	......	6.49	6.66	1	6.54	6.59	......	8
9. Reaction to improvement in flying	55,015	6.52	6.25	1	6.21	6.60	1	6.21	6.47	1	6.29	6.41	......	6.30	6.42	......	6.43	6.33	......	9
10. Personal interest in student	40,965	5.59	5.25	[1] 1	5.26	5.69	1	5.26	5.50	5	5.33	5.44	......	5.32	5.47	......	5.37	5.41	......	10
11. Use of flight time	63,026	6.80	6.97	5	6.97	6.73	1	7.02	6.80	[1] 1	6.94	6.87	......	6.88	6.92	......	6.81	6.93	......	11
12. Emphasizes an understanding of the airplane, etc	50,853	5.96	6.17	1	6.07	6.10	......	6.07	6.09	......	5.99	6.14	......	6.03	6.13	......	6.02	6.11	......	12
13. Conduct as an officer	62,911	6.90	6.85	......	6.85	6.92	......	6.79	6.94	5	6.82	6.91	......	6.79	6.95	5	6.79	6.91	......	13
14. Ability to understand the problems of the student	51,139	6.12	6.06	......	6.05	6.21	......	6.09	6.11	......	6.10	6.10	......	6.08	6.12	......	6.08	6.11	......	14
15. Encouragement to study flying outside of the airplane	52,789	6.02	6.39	1	6.26	6.16	......	6.21	6.24	......	6.11	6.31	5	6.18	6.28	......	6.16	6.26	......	15
16. Confidence in this airplane	71,175	7.26	7.46	1	7.37	7.39	......	7.36	7.39	......	7.38	7.38	......	7.34	7.42	......	7.35	7.39	......	16
17. Patience and self-control	62,402	6.92	6.68	[1] 1	6.69	6.98	1	6.74	6.82	......	6.71	6.83	......	6.79	6.78	......	6.81	6.77	......	17
18. Effectiveness in putting flying training across	49,322	5.85	6.11	1	6.01	5.98	......	5.99	6.00	......	5.89	6.07	5	5.93	6.06	......	5.87	6.05	......	18
Total score	37,003	4.95	5.02	......	4.89	5.21	1	4.76	5.16	1	4.80	5.11	1	4.80	5.17	1	4.84	5.05	......	Total
Supervisor's-flying	5,959	1.97	2.18	1	2.14	1.85	[1] 1	2.06	2.04	......	2.01	2.07	......	2.03	2.07	......	1.96	2.09	1	Sup. Fly.
Supervisor's-instructing	6,029	1.88	2.19	1	2.15	1.85	[1] 1	2.08	2.04	......	1.99	2.10	1	2.04	2.08	......	1.94	2.11	1	Sup. Inst.

[1] Reversals, i. e., poorer ratings were given instructors who had more months of instruction or who were first lieutenants or higher.

an understanding of the airplane, etc.; encouragement to study flying outside the airplane; and the over-all item, effectiveness in putting flying training across. All of the items on which the instructors with more teaching experience were rated as significantly poorer than those with less experience, were items dealing with the personal aspects of instructing. They were: use of "chewing," sarcasm, swearing, etc.; patience and self-control; reaction to improvement in flying; and personal interest in the student.

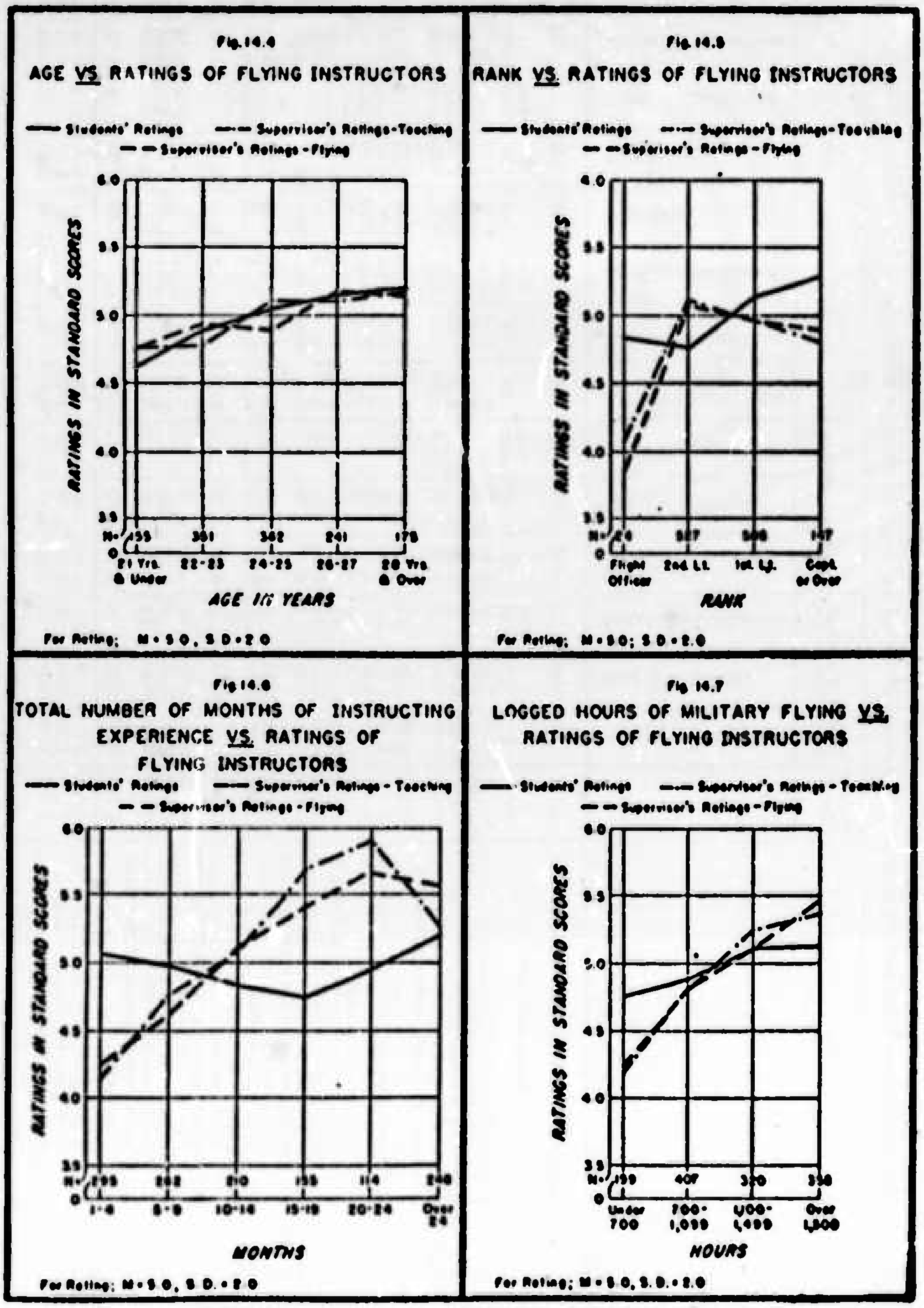
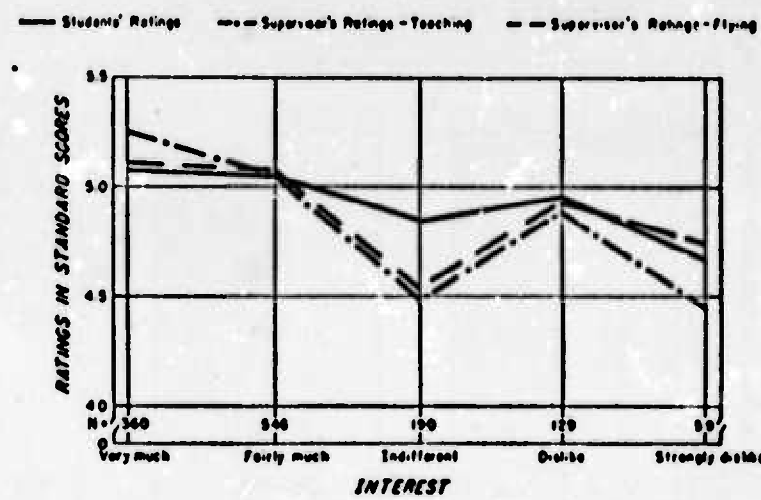


Fig. 14.8

LIKING FOR JOB OF FLYING INSTRUCTOR VS. RATINGS OF FLYING INSTRUCTORS



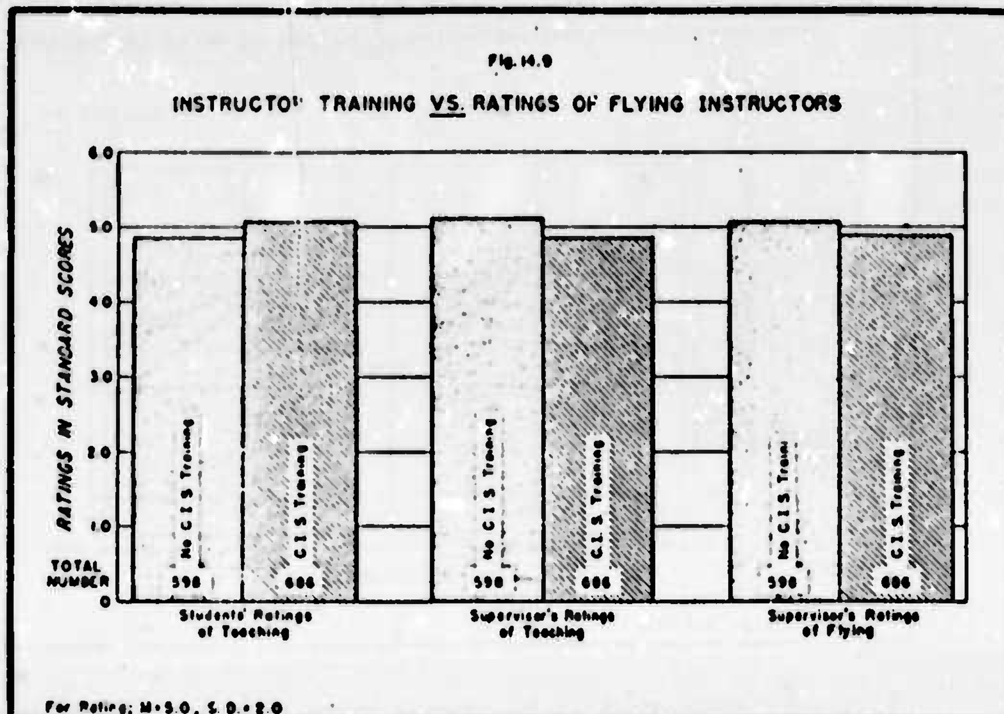
For Rating: M = 5.0, S D = 2.0

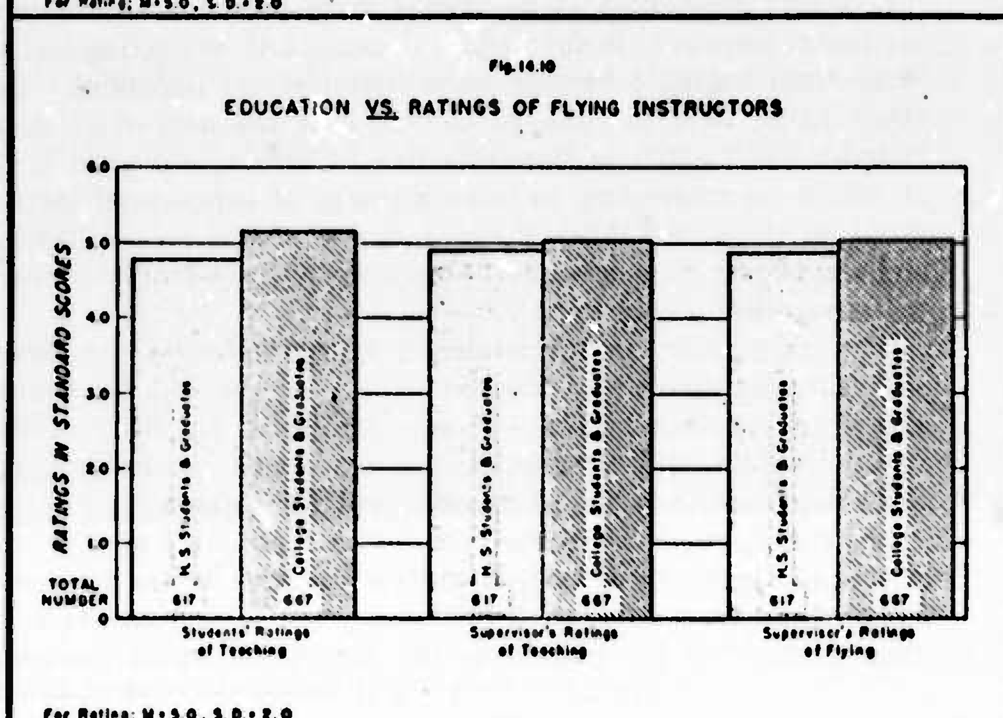
It would seem that as he spends more time teaching flying, the instructor improves in the technical aspects of instructing but at the same time tends to become more irritable and impatient. On the other hand, it is possible, of course, that the instructors find this "hard-boiled" attitude more effective in turning out good students. It would be interesting to have a group of experienced instructors rated on these two types of items, technical and personality, before and after being given a rest from teaching flying by temporary rotation to some other assignment.

The total score on the student's rating scale did not show any difference between the instructors with more and with less months of experience at that station. This is accounted for by the fact that the instructors with more experience were rated as reliably better on some items of the scale and reliably poorer on others.

In general, supervisors rated the instructors with a greater number of months instructing at that station as both better teachers and better flyers than those with fewer months. This seems reasonable, since in making his judgment, the supervisor would naturally be guided more by factors of technique and less by the student-instructor relationship which he would have little opportunity to observe.

The preceding analysis has been based on number of months instructing at the same station. Similar results on the student's total score and on the supervisors' two ratings were also obtained for the total number of months instructing experience at all stations. These results are presented graphically in figure 14.6. In this figure it can be seen that although, in general, the instructors with more





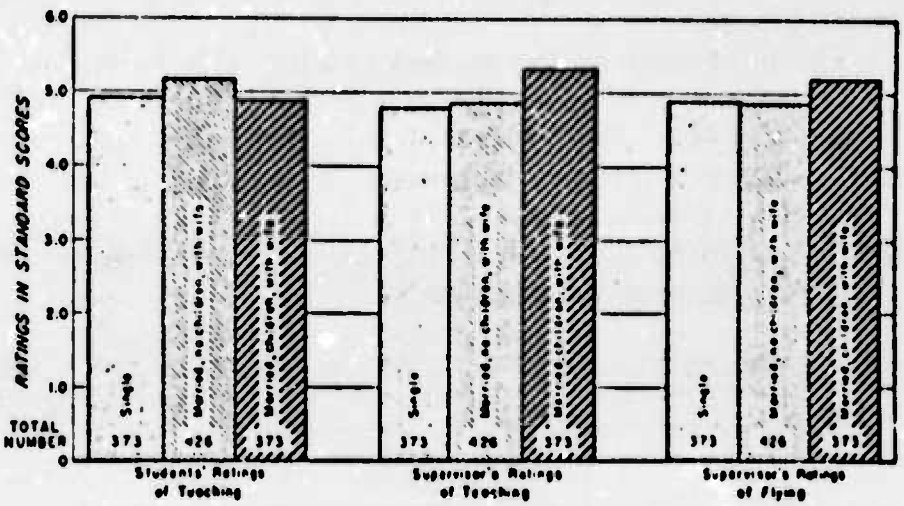
experience are given the higher ratings by the supervisors, there is some tendency for those who have been instructing more than two years to be given lower ratings on teaching ability than those who have been instructing from 20 months to 2 years. This tendency for a reversal at the far end of the curve was found to be reliable at the 1 percent level. This appears to confirm the widespread opinion that flying instructors tend to go "stale" after teaching flying too long.

Convincing proof would require a contrast of possible selective factors differentiating the groups with differing experience.

Combat service.—Eleven of the items of the student's rating scale on which the instructors with combat service were rated as significantly better than those without combat service were ones dealing with patience and self-control and the student-instructor relationship. They were: use of "chewing," sarcasm, swearing, etc.; patience and self-control; reaction to improvement in flying; and personal interest

Fig. 14.11

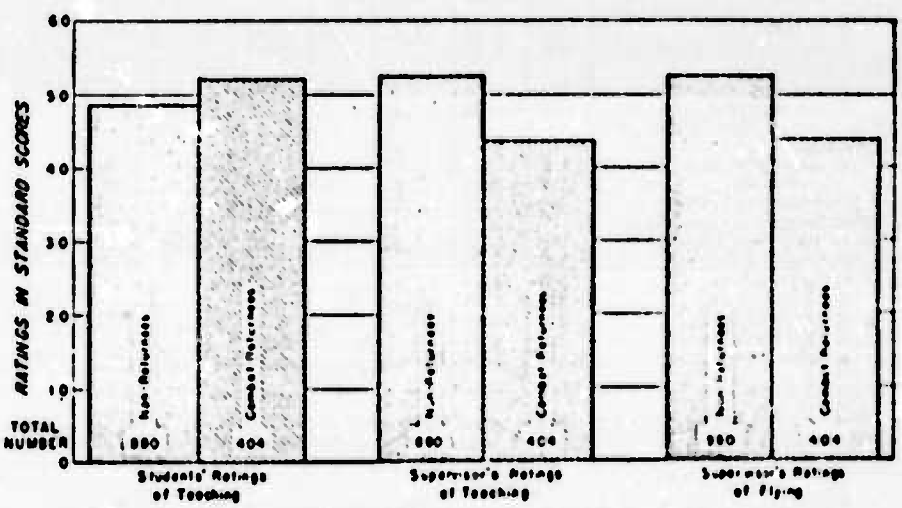
MARITAL STATUS VS. RATINGS OF FLYING INSTRUCTORS



For Rating, M = 5.0, S.D. = 2.0

Fig. 14.12

COMBAT SERVICE VS. RATINGS OF FLYING INSTRUCTORS



For Rating; M = 5.0, S.D. = 2.0

in the student. All of the items on which the instructors with combat experience were rated as significantly lower than those without combat experience were items dealing with the technical aspects of instructing. They were: ability to express himself; on-the-ground teaching effectiveness; and use of flight time.

It will be noticed that the difference between combat returnees and nonreturnees is similar to the one between the less and more experienced instructors, except that in the latter case, statistically reliable differences were secured for more of the items in the student's scale. The similarity of the results for these variables is not surprising since, in general, the combat returnees had less teaching experience than the nonreturnees.

The total score on the student's rating scale gave the instructors with combat service a reliably higher rating than those without combat service. Supervisors, on the other hand, rated nonreturnees as both better teachers and better flyers. This seems logical since the nonreturnees appeared to be better in aspects of instructing such as use of flight time and ability to express themselves, which are easier for the supervisor to observe.

TABLE 14.8.—*Relationship of additional items of personal data to total score on student's rating scale of teaching proficiency and to supervisor's ratings*

Biographical items	Students' ratings—total score				Supervisors' rating—flying			Supervisors' rating—instructing		
	N	Mean	F-ratio	P	Mean	F-ratio	P	Mean	F-ratio	P
Marital status:										
a. Single	373	4.91	2.54	0.20	2.02	3.66	0.05	2.00	7.07	0.001
b. Married, no children, wife with you	436	5.18			2.01			2.02		
c. Married, with children, wife with you	373	4.90			2.12			2.17		
Logged hours military flying:										
a. 0-1099 hours	606	4.81	6.46	.05	1.93	37.40	.001	1.93	40.60	.001
b. 1100 to 1000 + hours	673	5.12			2.15			2.17		
C. I. S. training:										
a. None or local school	598	4.89	3.11	.20	2.08	2.73	.20	2.10	3.61	.20
b. All C. I. S.	686	5.08			2.02			2.02		
Experience nonflying instructor:										
a. None to less than 6 months	1163	4.98	.33	.50	2.05	1.25	.50	2.06	.07	.50
b. 6 months to 5 years	121	5.09			1.98			2.04		
Total instructor experience:										
a. Less than 1 month	30	5.63	1.57	.20	1.83	16.51	.001	1.73	19.07	.001
b. 1-4 months	265	5.01			1.80			1.73		
c. 5-9 months	262	4.98			1.93			1.98		
d. 10-14 months	210	4.85			2.09			2.09		
e. 15-19 months	155	4.73			2.19			2.30		
f. 20-24 months	114	4.95			2.27			2.37		
g. Over 2 years	218	5.20			2.24			2.16		

P = Probability of obtaining by chance alone a difference of this magnitude. A "P" of 0.01 means that the probability that the observed difference is due to chance factors is less than 1 in 100.

Military rank of instructor.—The items of the student's rating scale on which the instructors with higher military rank were rated as reliably better than those with lower rank were ones dealing with patience and self-control and the student-instructor relationship.

They were: use of "chewing," sarcasm, swearing, etc.; reaction to improvement in flying; personal interest in the student; and conduct as an officer. There were only two items on which the higher ranking officers were rated as reliably poorer than the lower ranking ones. These were: encouragement of student to ask questions; and use of flight time.

The total score on the student's rating scale gave the instructors with higher military rank a reliably better rating than the lower ranking ones. Supervisors, on the other hand, tended to rate the higher ranking instructors as inferior to the lower ranking ones in both teaching and flying proficiency. The differences were, however, not statistically reliable.

Age of instructor.—Of the three items on the student's rating scale on which the older instructors were reliably rated better than the younger ones, two dealt with patience and self-control and the student-instructor relationship while the third was the over-all item. They were: use of "chewing," sarcasm, swearing, etc.; encouragement to study flying outside of the airplane; and effectiveness in putting flying training across. One of the items on which the older instructors were rated reliably better coincided with one on which the higher ranking instructors were also rated as reliably better. The younger instructors were not rated reliably better than the older ones on any of the items.

The total score on the student's rating scale gave the older instructors a reliably better rating than the younger ones. Supervisors also reliably rated the older instructors as better than the younger ones in their instructing ability, but did not rate the older group reliably as better flyers. This would indicate that maturity is a decided asset to a flying instructor but does not increase his flying proficiency.

Educational level of instructor.—The instructors with some college education were rated reliably higher than those who had not gone beyond high school on two items of the student's rating scale. One of these items, "conduct as an officer," coincided with one on which the higher ranking officers were also rated reliably as better. The other item, "ability to express himself," is one which would logically be expected to be highly correlated with educational level. There were no items on which the instructors who had not gone beyond high school were rated reliably as better than those who had.

The total score on the student's rating scale gave the instructors with more than a high school education a reliably better rating than those who had not gone beyond high school. There was no reliable relationship, however, between the education of the instructor and supervisors' ratings of his teaching or flying proficiency.

Liking of instructor for his job.—The instructors who liked their assignment as instructors were rated reliably as better than those who disliked their assignment on two items on the student's rating scale.

seem that the type of behavior which leads students to believe that the instructor is interested in his job, irrespective of whether he really is interested or not, causes the students to give him a higher rating in teaching proficiency.

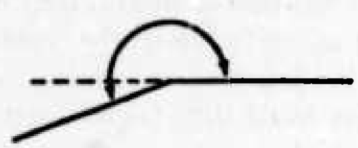
CONSTRUCTION AND VALIDATION OF BATTERIES OF PILOT INSTRUCTOR SELECTION TESTS

The Battery Developed for Selecting Primary Civilian Instructors [90]

The Civilian Flying Instructor Test (CFI Test) developed by Psychological Research Unit No. 2 was designed as a teaching aptitude test for civilian flying instructors. The final form of this test (Form B) consisted of nine subtests. These are described briefly below.

Test 1: *CFI Test, Part I, Angular Judgment* (*CP217A*).—Since field trips to primary flying schools had disclosed that both instructors and cadets were frequently called upon to make judgments of the magnitudes of angles, it was decided to include a test of angular judgment in the Civilian Flying Instructor Battery. It was a speed test involving perception and judgment of the magnitude in degree of a visually presented angle. The test consisted of 45 items, requiring 8 minutes for administration. It was scored $R-\frac{W}{4}$. A sample item is given below:

Item 5



(*a*) 170° (*b*) 180° (*c*) 190° (*d*) 200° (*e*) 210°

Test 2: *CFI Test, Part II, Aviation Information.*—This is a power test measuring knowledge of specific aviation information, similar to the pilot items of General Information, CE505E, used in the aircrew classification battery. The test consisted of 30 items, requiring 15 minutes for administration. It was scored $R-\frac{W}{4}$. A sample item is:

Lt. Col. Buzz Wagner was a famous American
- *a.* fighter pilot.
- *b.* bomber pilot.
- *c.* speed and stunt pilot.
- *d.* test pilot.
- *e.* bombardier.

[90] Capts. John T. Dailey, Glenn L. Finch, and Richard P. Youts are primarily responsible for the selection and development of this battery of tests.

Test 3: *CFI Test, Part III, Complex Mechanical Comprehension.*—This was a power test consisting of items selected from several forms of the Aviation Cadet Qualifying Examination, in which a mechanical drawing of some mechanism was presented and questions were asked about the functional relation of parts. The test consisted of 15 items, requiring 20 minutes for administration. It was scored $R-\frac{W}{4}$.

Test 4: *CFI Test, Part IV, Aviation Vocabulary.*—This was a power test consisting of original items on vocabulary terms found to be frequently used in describing, explaining, or teaching elementary flight maneuvers. The test consisted of 30 items, requiring 20 minutes for administration. It was scored $R-\frac{W}{4}$. A sample item is given below:

"Motion of an aircraft relative to the air, in which the lateral axis is inclined and the airplane has a velocity component along the lateral axis" is known as

a. stalling.
b. rolling.
c. sideslipping.
d. pitching.
e. yawing.

Test 5: *CFI Test, Part V, Reading Comprehension on Aviation Topics.*—This was a power test consisting of items selected from various forms of the Aviation Cadet Qualifying Examination on the basis of apparent face validity. A paragraph of technical material was presented, followed by a series of questions which could be answered from information presented in the paragraph. The test consisted of 15 items, requiring 15 minutes for administration. It was scored $R-\frac{W}{4}$.

Test 6: *CFI Test, Part VI, Pedagogical Judgment.*—This was a power test constructed especially for this battery. The items are based on an outline of the course on the Psychology and Technique of Instruction which had been given at the Central Instructors School (Primary Pilot) at Randolph Field. Each item involved an act of judgment in selecting one of five alternate responses to a situation commonly encountered in primary flight instruction. The test consisted of 15 items, requiring 15 minutes for administration. It was scored $R-\frac{W}{4}$. A sample item is given below:

Informing a student how well he is doing

a. is practically always a useful technique for keeping up motivation.

b. is useful only with the poorest students, if ever.

c. is poor because it makes satisfactory students overconfident.

d. is bad because it is not in accord with Air Corps tradition.

e. will cause poor students to become nervous and good students to let down.

Test 7: *CFI Test, Part VII, Biographical Inventory.*—This consisted of biographical items in a format somewhat similar to Biographical Data, CE602D, used in the aircrew classification battery. The test consisted of 45 items, requiring 20 minutes for administration. It was scored $R-W$. A sample item is given below:

What is the extent of your nonflying teaching experience before your present position?

a. Public or private academic teaching.

b. Teaching in a trade or vocational school.

c. Teaching or coaching athletics, sports, gym, etc.

d. Teaching use of machines, etc., in industry or teaching salesmen, clerks, etc., in business.

e. No teaching experience.

Test 8: *Aircraft Control Analysis, Form A, (CI622AX1) and Form B (CI622AX2).*—This test was developed by T/Sgt. Paul McReynolds. It measured the ability to analyze rapidly and accurately the effect of aircraft control movements on the actions of the plane. The task was to decide whether each control position is correct for a given maneuver. Both forms were scored $R-\frac{W}{4}$.

Test 9: *Otis Gamma Test, Form AM (CI623A).*—A standard intelligence test which was added to the battery as a possible source of additional validity and for purposes of intercorrelation with the other sub-tests. It was scored Rights only.

Reliabilities of tests of the battery.—The Otis Test, the Mechanical Comprehension section of the CFI Test, and the Reading Comprehension section of the CFI Test were fully developed tests adopted without change and it was, therefore, deemed unnecessary to compute reliabilities for these tests. The investigators believed that the various items in the Biographical Inventory were aimed at such different things that a conventional statement of the reliability of this inventory would not be relevant. Reliability data for the other five tests are given in table 14.10. It can be seen that all of the tests, with the exception of Pedagogical Judgment, have satisfactorily high reliabilities.

Criteria used in validating tests.—After an investigation of all available types of criterion data at Primary schools, the three following measures were used in validating the battery of tests.

1. Over-all ratings of instructor's teaching ability by flight supervisors and squadron supervisors. Each instructor was rated by his flight supervisor and his squadron supervisor. In validating the tests the ratings of the two supervisors on each instructor were averaged.

2. Over-all ratings of instructor's flying ability by flight supervisors and squadron supervisors. Each instructor was rated by his two supervisors and the two ratings were averaged for validation purposes.

TABLE 14.10.—*Reliabilities of tests administered to civilian flying instructors*

Tests

	Angular judgment	Aviation information	Aviation vocabulary	Pedagogical judgment	Aircraft control	
					Form A	Form B
N	734	518	518	518	420	98
Reliability	[2] 0.87	[1] 0.80	[1] 0.79	[1] 0.42	[1] 0.92	[1] 0.93

[1] Computed for instructors by Hoyt method.
[2] Odd-even reliability corrected by Spearman-Brown prophecy formula for unclassified aviation students.

3. Ratings of instructors by supervisors on a 16-item descriptive rating scale developed by Psychological Research Unit No. 2. This rating scale has already been described on pages 307 to 309 of this chapter. For validation purposes each scale item was weighted 1, 2, or 3 according to its median rank of importance assigned by supervisors at the schools where the rating scale was administered.

Correlation of tests with criteria.—The product-moment correlations of the 9 tests and over-all teaching rating, over-all flying rating, and the 16 items of the descriptive rating scale are given in table 14.11.

In addition, tetrachoric correlations were calculated between each of the 16 items on the descriptive rating scale and each of the 9 tests in the instructor selection battery. Some items were found to correlate relatively highly with the tests. The multiple correlation of the 9 tests was 0.46 with pilot ability, 0.44 with ability to analyze errors, and 0.43 with ability to express oneself. The nine tests had very low multiple correlations with other items, such as interest and conscientiousness. The multiple correlation of the aggregate weighted score of the descriptive rating scale with all 9 tests was 0.35. Since these multiple correlations were based on weightings derived from the same sample, they would be expected to shrink upon repetition and, in general, the higher ones would be expected to shrink the most.

Flying ability was slightly better predicted by the battery of instructor selection tests used than was teaching proficiency. While certain elements in teaching proficiency seemed to be measured by the nine tests used in the instructor selection battery, other important elements were left almost completely untouched by these tests.

TABLE 14.11.—*Summary of validities of the tests for civilian flying instructors against items in the descriptive rating scale*

N = 433 [1]

	Validities (r_{bis})									
	1 A. J.[2]	2 A. I.	3 M. C.	4 A. V.	5 R. C.	6 P. J.	7 B. I.	8 ACA	9 Otis	Multiple R [3]
1. Pilot ability	0.18	0.22	0.26	0.27	0.18	0.31	0.25	0.22	0.09	0.46
2. Confidence	.06	.04	.03	.08	.03	.19	.24	.04	.22	.34
3. Anchor	.12	.11	.14	.10	.07	.09	.24	.10	.22	.31
4. Motivate cadets	.11	.10	.04	−.02	.02	.12	.16	.05	−.07	.21
5. Patience	.04	.19	.13	−.05	−.02	.03	.06	.01	.18	.21
6. Interest	.01	−.02	.03	−.07	.00	.08	.11	−.08	−.05	......
7. Up to schedule	.02	.12	.04	.03	−.03	.12	.09	.12	.19	.20
8. Extra help	.11	.05	.10	.10	.09	.16	.05	.01	.18	.21
9. Adapting methods	.12	.11	.09	.04	.01	.18	.20	.14	.21	.31
10. Conscientiousness	.00	.00	.15	.09	.02	.09	.09	.06	.04	......
11. Expression	.17	.23	.21	.18	.20	.15	.25	.30	.16	.43
12. Grade slips	.09	.12	.11	.17	.12	.08	.21	.26	.14	.26
13. Analyzing errors	.17	.13	.16	.14	.16	.21	.32	.23	.08	.44
14. Students' attitude	.03	.10	.06	.07	.09	.16	.19	.10	.09	.26
15. Personality	.14	.21	.16	.08	.11	.21	.20	.12	.08	.34
16. Anchor	.16	.13	.19	.13	.08	.14	.31	.07	.12	.38
Over-all teach. rating	.00	.05	.24	.12	.05	.04	.18	.01	.03	.28
Over-all fly. rating	.15	.21	.28	.20	.26	.05	.26	.24	.09	.41
Average weight DRS score	.12	.07	.18	.06	.05	.09	.28	.10	.06	.35

[1] With the exception of ACA based on 313 and Otis on 239 students.
[2] Abbreviations refer to tests in same sequence as presented in text pp. 334 through 336.
[3] Using only tests with validities of 0.12 or greater. Since the weights were based on this same sample, the multiple correlations will be subject to shrinkage upon repetition of the validation, and in general, the higher correlations will be expected to shrink more than the lower ones.

The Battery of Instructor Selection Tests Administered at Redistribution Stations to Returned Combat Personnel

Six of the battery of instructor selection tests administered at redistribution stations to returned combat personnel were weighted for pilot instructing and were known as the Pilot Instructor Selection Battery. These tests were developed by the Psychological Research Projects and other psychological agencies in the Air Forces. They were all tried out experimentally and refined before the final battery was constructed. The Pilot Instructor Selection Battery was given to all returned combat pilots passing through redistribution stations during the period 16 November 1944 through 21 June 1945. The individual tests in this battery are described briefly below.

1. *Mechanical Comprehension Test, DS201A.*—This was a test of the officer's knowledge of mechanical relationships. It was similar to the mechanical comprehension test which had been one of the most promising of the tests administered to civilian flying instructors. Mechanical drawings were presented and questions were asked, the correct answer to which presupposed a knowledge of the functional relations of the parts. The test consisted of 40 items. The administration time was 15 minutes. The scoring formula was $R - \frac{W}{3}$.

2. *Aircrew Officer Blank, DE405A.*—This was a survey of experience and attitudes which available evidence indicated should affect success as an instructor. Some questions called for statements of attitudes;

others required factual information about past experience. The test consisted of 90 items. The administration time was 30 minutes. The scoring formula was R. Two sample items are given below.

My preference regarding return to combat duty is

a. I want to return soon.
b. I hope to return sometime.
c. I don't care whether I return or not.
d. I'd somewhat prefer not to return.
e. I'd strongly prefer not to return.

How easy do you find it to meet and get acquainted with strangers when there is no one present to introduce you and get the conversation going?

a. Very easy to get acquainted; nothing to it.
b. Usually no difficulty making acquaintance.
c. About the average hesitancy and slowness.
d. Slower and less successful than the average.
e. Very much slower and harder than for the average.

3. *Reasoning Test, DG501B.*—This was a test of the ability to solve arithmetical problems involving reasoning. The test consisted of 30 items, requiring 20 minutes for administration. The scoring formula was $R-\frac{W}{4}$. A sample item is given below.

A pilot has flown 120 hours in 70 days. How many hours will he have to fly in the next ten days in order to average 2 hours of flying time per day for the entire period?

a. 20 hours.
b. 40 hours.
c. 70 hours.
d. 90 hours.
e. 120 hours.

4. *Verbal Comprehension Test, DG101B.*—This was a test of the ability to comprehend written material of a rather complex nature. A technically written paragraph on some subject was presented and a series of questions were asked on it. Answering the questions correctly presupposed a complete understanding of the verbally presented material. The test consisted of a total of 30 questions on 6 paragraphs of material, requiring 20 minutes for administration. The scoring formula was $R-\frac{W}{4}$.

5. *Personal Inventory, DE201C.*—This was an inventory of information concerning factors such as the officer's interests, fears, and social adjustment. It was aimed at predicting a number of qualities considered to be desirable in an instructor and believed to be related to such factors. The inventory consisted of 45 items, requiring 10 min-

utes for administration. The scoring formula was *R*. Two sample items are given below. In each question a forced choice was required between two answers—the one on the left side of the page and the one on the right.

a. I like working by myself. *b*. I like working with others.

a. I wish I wouldn't feel so tired. *b*. I wish I could have a more responsible job.

6. *Analysis of Maneuvers, DS401B.*—This was a test of an understanding of maneuvers which an instructor must be able to explain to his students. It was based on the assumption that pilots who understand the fundamental principles of flying and are able to apply them to the analysis of maneuvers will be more likely to be able to explain flying to the student and to analyze his errors. The test consisted of 60 true-false items, requiring 20 minutes for administration. The scoring formula was *R–W*. Two sample items are given below.

(False) If the nose of the airplane moves to the right of the intended flight path in executing a loop, there was probably insufficient correction of torque.

(True) With altitude held constant, the angle of attack would be greater in a steep turn at 120 m. p. h than in level flight at 120 m. p. h.

Reliability of tests of battery.—The reliabilities of the six tests of the Pilot Instructor Selection Battery are presented in table 14.12. It can be seen that the reliabilities of all of the tests were satisfactorily high, with the exception of Analysis of Maneuvers which was considerably lower than the others. The reliabilities ranged from 0.42 to 0.83.

TABLE 14.12.—*Reliabilities and intercorrelations of tests and inventories in the pilot instructor selection battery administered at redistribution stations*

N=400 [1]

Name of test	1	2	3	4	5	6	7	r_{tt} [2]	Means	S. D.
1. Reasoning, DO501B	......	0.50	0.29	0.00	0.08	0.38	0.00	0.85	15.17	5.68
2. Verbal comprehension, DO101B	0.50	......	.34	–.05	.07	.45	–.09	.78	15.21	5.00
3. Analysis of maneuvers, DS401B	.29	.34	......	.01	.05	.43	–.01	.42	30.78	7.94
4. Air crew officer blank, DE405A	.00	–.05	.01	......	.09	.06	.61	.71	35.41	5.93
5. Personal inventory, DE301C	.08	.07	.05	.09	......	.05	.08	.81	15.57	3.37
6. Mechanical comprehension, DS301A	.38	.45	.43	.06	.05	......	.02	.83	18.10	6.14
7. Instructor preference, DE501C	.00	–.09	–.01	.61	.08	.02	......	......	4.88	2.50
8. Instructor stanine	......	......	......	......	......	......	.18	......	5.13	1.96

[1] These cases were tested at AAF Redistribution Station No. 2, Miami Beach, Fla., from 15 January to 1 May 1945.

[2] Reliability computed by Hoyt method.

Intercorrelations of tests of battery.—Intercorrelations between the six tests used in the Pilot Instructor Selection Battery were computed for a population of 400 pilots tested between 15 January and 1 May 1945 at Redistribution Station No. 2, Miami Beach, Fla. These are also presented in table 14.12. It will be seen that the tests were satisfactorily independent with intercorrelations ranging from −0.05 to 0.50.

Criteria used in validating the tests.—As already discussed on pages 302 through 306 of this chapter, no suitable "in training" criteria could be found for validating the Pilot Instructor Selection Battery of tests. The three following on-the-job criteria were employed in the validating study:

1. Students' ratings of instructors secured on the 18-item Students' Rating Scale developed by the Pilot Project. The average total score of all students rating the same instructor was used in the validation study. An average of 3.12 students rated each instructor.
2. Supervisor's over-all rating of the instructing proficiency of the instructor.
3. Supervisor's over-all rating of the flying ability of the instructor.

The two latter ratings were on a 3-point scale (in upper 25 percent of instructors, in middle 50 percent of instructors, or in lower 25 percent of instructors in each school.)

As already reported on pages 317 and 319 of this chapter, correlations between these three types of criterion data showed that there was little relationship between supervisors' and students' ratings of instructor proficiency ($r=0.08$), or between supervisors' ratings of flying ability and students' ratings of instructor proficiency ($r=0.06$). There was, on the other hand, a close correspondence between supervisors' ratings of flying ability and their ratings of instructor proficiency ($r=0.70$).

Validation results.—The validation results were based on a population of 215 returned combat pilots on whom validation data had been secured and on whom test scores were available.[11] Each test of the Pilot Instructor Selection Battery, the instructor stanine, an aggregate score based on the weighted averages of all the tests, and the instructor preference were validated against each of the three criteria mentioned above. Before computing the correlations, the students' rating scale scores and the scores on each of the instructor selection tests were normalized and converted into standard scores ranging from 1 to 9. The correlations obtained are presented in table 14.13.

It will be seen that the correlations range from −0.19 to 0.04 and that the majority of them are negative. The conclusion must be drawn that there is no positive relationship between the scores made on the Pilot Instructor Selection Tests and the ratings which a returned combat pilot will receive from his students and supervisors on his instructing proficiency and flying ability while actually teaching students in the field. This lack of relationship may be due to either

[11] Ratings were secured on a total of 1,281 instructors of whom 405 were combat returnees, but only 218 of these were in the group of pilots who took the instructor selection tests at the redistribution Stations.

TABLE 14.13.—*Validities of pilot instructor selection tests administered at redistribution stations*

	Student's rating N=215	Supervisor's rating			
		Teaching N=179	Flying N=179	Mean	S. D.
Mechanical comp., DS201A	0.02	−0.17	−0.10	5.12	1.93
Aircrew officer blank, DE405A	.00	−.10	−.04	5.04	1.95
Reasoning test, DO501B	−.05	−.12	−.01	4.95	2.02
Verbal comprehension, DO101B	−.19	−.16	−.10	5.11	1.87
Personal inventory, DE201C	−.07	−.02	.07	5.05	2.01
Analysis of Maneuvers, DS401B	.04	−.03	.02	5.04	1.90
Instructor stanine	−.04	−.18	−.07	5.42	1.89
Instructor preference	−.04	−.02	−.08	5.09	2.49
Student's rating	....	[1].08	[1].06	5.19	1.91
Supervisor's rating of teaching	[1].08	....	[1].70	1.83	.62
Supervisor's rating of flying	[1].06	[1].70	....	1.83	.66

[1] Based upon population of 1,141 nonreturnee and combat returnee instructors.

(1) an inability of the tests to separate the men who will make good instructors from the men who will not, or (2) the inadequacy of ratings of students and supervisors to determine the ability of an instructor. It cannot be said with any degree of certainty which one of these alternatives may have produced these low validities. The reliability of each of the tests of the instructor selection battery is relatively high. Five of the 6 tests have a reliability above 0.70, the other one a reliability of 0.42. The more rigorous test of the reliability of the student's rating scale (i. e., correlation between ratings made by two different classes) yields a coefficient of 0.36. The reliability of both the tests and the student's rating scale is therefore high enough to allow a significant correlation to be obtained if a relationship existed. Furthermore, the scores on both the student's and the supervisor's rating scales do correlate significantly with such factors as age and educational background of the instructor which would logically be expected to be associated with his ability as a teacher.

Attempt to explain negative correlations.—In general, the correlations between the Pilot Instructor Selection Tests and both types of ratings were negative; the correlation between instructor stanine and supervisors' ratings was −0.18, a correlation of a size which would be expected to occur by chance only 2 times in 100. Since it seemed extremely unlikely that the true relationship between the supervisors' ratings and general intelligence and the other factors measured by the instructor selection tests would actually be negative, an attempt was made to locate sources of constant error which might have biased the correlation in a negative direction and obscured a positive relationship, if one existed.

The most reasonable hypotheses which could be made concerning such factors were as follows: (a) That the first returnees selected as instructors at the redistribution stations were not selected on the basis of their test scores but that the later ones were somewhat selected for instructor jobs on this basis. This would mean that the average instructor stanine of the later group would be somewhat higher than that of the earlier one; (b) that the first returnees sent out by redistribution stations were given teaching assignments sooner than the later ones. This would mean that the later group would have been instructing for a shorter period than the earlier one; (c) that the returnees who had been instructing for a shorter period were given lower ratings by their supervisors. Since they would also be the ones who were selected on the basis of their instructor stanines, such a factor would tend to produce a negative correlation between supervisors' ratings and instructor stanines.

In order to test these hypotheses, correlations were computed between the following: (1) Instructor stanine and date of arrival at redistribution station; (2) instructor stanine and total months of instructing; (3) supervisors' ratings of teaching proficiency and date of arrival at redistribution stations; (4) supervisors' rating of teaching proficiency and total months instructing; (5) date of arrival at redistribution stations and total months instructing. These correlations are presented in table 14.14.

TABLE 14.14.—*Intercorrelations among instructor stanine, supervisors' rating, date of arrival at redistribution station, and total number of months instructing*

	1	2	3[1]	4
1. Instructor stanine	....	−0.18 (179)	0.03 (215)	0.02 (215)
2. Supervisors' rating—teaching	−0.18 (179)	....	0.14 (179)	0.03 (179)
3. Date of arrival at redistribution station	0.03 (215)	0.14 (179)	....	−0.21 (215)
4. Total number of months instructing	0.02 (215)	0.03 (179)	−0.21	....

[1] Later arrival at redistribution station given high weight.
The figures in parentheses indicate the number of cases involved in each correlation.

Partial correlation based on above matrix

$r_{12.3} = -0.19$

$r_{12.4} = -0.18$

Semipartial correlation based on above matrix

$r_{(1.3).(2.4)} = -0.18$

On the basis of these correlations it was possible to partial out the effect of date of arrival of the returnees at the redistribution stations from the correlation between supervisors' ratings and instructor stanine. It was also possible to partial out the effect of the total number of months of instructor experience. Finally a semipartial correlation was calculated to eliminate the effect of date of arrival at the redistribution stations from the instructor stanine, and number of months of instructor experience from the supervisors' ratings of instructor proficiency. None of these procedures affected the correlation between instructor stanine and supervisors' ratings. Therefore, the factors which were outlined above could not have produced this negative correlation or have obscured a positive relationship, if one existed. All of the other variables on which data were available were considered, but it did not seem likely that any of them could have produced a spurious negative relationship.

Possible Reasons for Different Validities of Selection Batteries Administered to Civilian and Pilot Returnee Instructors

Better validity was secured on the tests administered to civilian flying instructors than those administered to pilot returnees at redistribution stations. Since some of the tests in the two batteries (particularly the ones involving mechanical and verbal comprehension) were highly similar and since the batteries were valid for one population and not for the other, it seemed desirable to look for factors which might account for these different results. Some of the more likely possibilities are listed below:

1. *Much wider range of ability in civilian instructor population.*— The civilian flying instructors were tested immediately after the rapid expansion of the flying training program. At the beginning of the expansion, it was possible for the contractors to secure a group of civilian instructors with very high professional standards. Toward the end of the expansion, the need was so great and the supply so limited that the contractors were forced to employ a considerable number of instructors who barely met the lowest standards of flying ability, educational background, and general intelligence. At the time the civilian instructors were tested, there was a very wide range in their ability.[93]

The pilot returnee population, on the other hand, represented a highly selected, homogeneous group. Those with the lower intelligence and poorer educational background had been weeded out by the Army Air Forces qualifying examination. A further selection, emphasizing factors related to pilot ability, was effected by the classification tests and by elimination in training. Those who survived these hurdles were highly selected and made up a relatively homogeneous group. It would, therefore, be easier for the tests to differentiate among the civilian instructors with a wide range of ability than among the more highly selected returnee group.

2. *Conditions in the schools at the time validating data were collected.*— The date for validating the battery of civilian flying instructor tests were collected during late 1943 and 1944 when the schools were running at full capacity and when the instructor and supervisory personnel at the schools were relatively stable. Data for validating the pilot instructor selection battery administered at redistribution stations was collected in the fall of 1945 when many of the schools were being closed and when there was a constant turnover in instructor and supervisory personnel. One would expect that more valid ratings

[93] With the training of more inspectors it was subsequently possible to eliminate the poor ones and regain the originally higher standards.

would be obtained in the former case where supervisors observed the performance of their instructors for longer periods of time under more stable conditions.

3. *Motivational factors.*—The civilian flying instructors to whom the battery of tests was given were on the whole highly motivated. The jobs for which they were held were interesting and well paid and, in addition, guaranteed temporary protection against their induction into the armed forces. Since they were all probably highly motivated, any differences in performance were largely dependent upon differences in ability. Interviews indicated that many of the returned combat pilots were, on the other hand, not interested in an instructor assignment. Since the returnee pilots varied considerably in their motivation and were all of relatively good aptitude, it seems likely that differences in motivation were an important factor in producing differences in performance. But it is difficult to measure differences in motivation by a battery of instructor selection tests. Furthermore, the returnees were tested at the redistribution station and, as has already been shown, their interest in the job of instructing changed quite a bit by the time they had arrived at the flying schools and found out more about their assignment. Since the civilian instructors were given the tests after they had been teaching at the schools, these tests were more likely to reflect their motivation at that time. Thus, differences in motivation would be more of a disturbing factor in the validation of tests on returnees than in the validation of tests on civilian instructors.

Discussion of Validation Results Secured on Bombardier, Navigator, and Pilot Populations

Psychological Research Project (Navigator) has reported that none of the Navigator Instructor Selection Tests investigated had more than a chance relationship with on-the-job criteria of instrutor proficiency.[93] The tests investigated were Survey of Attitudes and Scientific Background, and the weighted combined-score on these two tests. The on-the-job criteria used were ratings by students, fellow instructors, flight commanders, supervisors, and by the instructor himself. The correlation between the tests and the criterion data ranged from −0.18 to 0.11, with the majority of them being positive but very low.

Psychological Research Project (Bombardier) used supervisors' and students' ratings of instructor proficiency as on-the-job criteria

[93] In addition to securing on-the-job criteria of instructor proficiency, the Navigator and Bombardier Projects also collected considerable criterion data on returnee instructors while they were receiving their training at the instructor training schools. In the bombardier schools the instructor selection tests were validated against criterion data which were fairly heavily weighted with ground-school grades based on written tests. Rather high correlations were found in the bombardier schools between such criteria and the instructor stanine and some of the individual instructor selection tests. The Pilot Project did not attempt to validate the tests against criteria secured at the Pilot Central Instructors Schools since both elimination policy and proficiency records at these schools were for the most part based solely on the student's flying ability and did not reflect his instructing proficiency.

for validating the Bombardier Instructor Selection Tests. The correlations between the individual tests and the criterion data ranged from 0.11 to 0.43. In general, the correlations were fairly substantial, the students' ratings tending to correlate more highly with the tests than the supervisors' ratings.

Some of the factors which are thought to be responsible for the definitely better validation results secured on the Bombardier and the possibly better results secured on the Navigator battery of instructor selection tests in comparison with those secured on the Pilot battery are discussed briefly below:

1. *Instructor tasks may be different for the different types of instructors.*—Since four of the tests (Arithmetic Reasoning, Mechanical Comprehension, Personal Inventory, and Verbal Comprehension) were identical for the Bombardier and Pilot batteries and since statistically significant validities were obtained for these tests on a bombardier returnee population and not on a pilot returnee population, the differences in validation results could not have been produced by differences in the tests. They must have resulted either from the fact that the abilities measured by the tests were important in the instructor performance of bombardiers but not important in the instructor performance of pilots, or from other factors.

2. *Criterion ratings may have been better for bombardiers and navigators than for pilots.*—The same two types of ratings (ratings by students and ratings by supervisory personnel) were used in validating the tests on all three types of instructors. These two types of ratings correlated highest for bombardier instructors ($r=0.57$), next for navigator instructors ($r=0.27$) and lowest for pilot instructors ($r=0.08$). This may be partly due to the fact that supervisors have more opportunity to observe the performance of bombardier and navigator instructors since their instruction always takes place in multiengine planes where there is space for an observer. Furthermore, the unit commander in navigator schools was much better able to observe the instructors under him because so much instructing was done on the ground. It may also be partly due to the fact that it is easier for supervisors to observe the type of teaching procedure which characterizes good and poor bombardier and navigator instructors. Due to the fact that supervisors do not have much opportunity to observe pilot instructors in a teaching situation, and the difficulty of making valid observations as to what constitutes good and poor techniques of instructing, supervisors base their estimate of the teaching proficiency of the pilot instructor almost entirely on what they can easily observe, namely, his flying ability. A correlation of 0.70 was found between these two ratings by the Pilot Project. No single factor seems to have such influence in determining the ratings of teaching proficiency given to bombardier and navigator instructors by their supervisors.

3. *Range of ability probably more restricted in pilot and navigator returnee populations.*—During the first part of the war, the effect of the classification procedures was to send the student with lower general aptitudes to bombardier training. Two factors tended to lower the quality of the bombardier students: the elimination policies, and the preference, of most cadets, for pilot training. Therefore, the pilots and navigators would be expected to be the most select groups, while the bombardiers should contain numbers of men with considerably lower ability. This possibly wider range of ability among bombardiers, particularly at the low end of the distribution, may have been one of the reasons why the bombardier instructor selection tests proved to be the most effective.

SUMMARY OF RESULTS AND RECOMMENDATIONS FOR FUTURE WORK

Main Areas of Work and Chief Results

Since the flying instructor occupied a key position in the pilot-training program, a considerable amount of work was devoted to the development of procedures for evaluating and selecting flying instructors. The areas which this research covered and the chief results may be summarized as follows:

Job analysis.—Studies in this area delineated five attributes which supervisors, instructors, and students at all levels of training considered most important for the good flying instructor. These were: patience and self-control, ability to adapt teaching methods to individual students, flying ability, analysis of errors, and ability to express himself.

Development of supervisor's and student's rating scales.—Two rating scales, based on the qualities which the job analysis studies had shown were important, were constructed and refined. One of these was a 16-item scale for the rating of flying instructors by supervisors. The correlation between ratings by supervisors and assistant supervisors was 0.65. This is an estimate of maximum reliability since these two raters had probably influenced each other's opinion of the instructor. The intercorrelations of the items of the scale ranged from 0.26 to 0.90, with the majority of them rather high. This indicated the presence of a "halo" effect in the ratings as well as a possibly close relationship among the factors measured by the items.

The other scale consisted of 18 items and was for the rating of flying instructors by their students. When each instructor's students in a given class were divided into two halves and the ratings of these two groups were correlated with each other, the estimate of split-group reliability corrected for full number was 0.75. Correlating the average of the ratings by students in one class with that of students in a different class at least 5 weeks later, yielded a test-retest reliability of 0.36. Both of these figures are based on an average of 3.25 students per class for each instructor. The marked reduction in the size of the correlation when the students came from two different classes was probably caused by removing the opportunity for the students in the different groups to influence each other. It is also possible that the instructors changed somewhat during the period of five or more weeks between the two ratings.

One of the 13 items on the scale was an over-all rating. On different classes the test-retest reliability of this item was 0.23. The fact that the reliability of the total score was definitely higher ($r=0.36$, difference significant above the 1 percent level) than the single over-all item, indicated the value of rating different aspects of performance separately and then totaling these ratings.

These results were secured at Advanced and Transition schools. The split-group reliability at the basic level of training was approximately the same, but the reliability of ratings at the primary level, where the students were relatively inexperienced, was lower.

The intercorrelations among the items of the scale ranged from 0.12 to 0.76, with the majority of them below 0.40. This indicates that, in spite of "halo" effects and similarities among the different attributes rated, the various items in the scale are independent enough to be useful.

Development and validation of instructor selection tests.—On the basis of interviews, check lists, students' and supervisors' opinions, and other information gained in the job analysis studies of the factors related to instructor success, experimental instructor selection tests were constructed. One battery of these tests was administered to civilian primary flying instructors while instructing at a primary flying school subsequent to their instructor training. This battery was validated against the 16-item supervisor's rating scale. The multiple correlation of the aggregate weighted score of the 16-item scale with the battery of 9 tests was 0.35. Flying ability was slightly better predicted by the tests than was instructing proficiency. Validation against the separate items of the rating scale showed that some of them were measured by this battery of instructor selection tests while others were not.

The other battery of tests was administered at redistribution stations to returned combat pilots, some of whom later became instructors. This battery was validated against the 18-item student's rating scale and against supervisor's over-all rating of instructing proficiency and flying ability. There was no significant relationship between performance on any of the instructor selection tests in this battery and subsequent ratings which the flying instructor received from his students or his supervisors while teaching at Advanced and Transition schools.

Relationship of biographical factors to successful teaching.—Items of biographical and personal data were secured on 1,284 flying instructors and related to the average rating which these instructors received from students and supervisors.

It was found that preference expressed at a redistribution station did not predict how well an instructor would like instructing after he arrived at a pilot school and started to teach. Once he was on the job, however, the instructor's like or dislike for his assignment was quite stable from class to class.

The older flying instructors received reliably higher ratings as teachers than the younger ones from both their students and their supervisors. The instructors with more education and those with a higher military rank were also rated as reliably better by their students, but were not rated as better instructors by their supervisors. Students gave combat returnees better ratings as instructors than nonreturnees, but supervisors gave the nonreturnee the better rating in both teaching and flying. Instructors who liked their assignment were rated as better instructors by their supervisors but not by their students.

The relationship of length of experience as a flying instructor to supervisors' ratings of teaching was curvilinear; the ratings became higher with more experience up to two years but were somewhat lower beyond that point. The total score on the student's rating scale showed no statistically reliable difference for instructors with different amounts of experience. However, an analysis for each of the items of this scale showed that the instructors who had taught flying for longer periods of time received reliably higher ratings on technical aspects of teaching ability, such as analysis of errors and ability to express themselves, and reliably lower ratings on personality items, such as interest in the student and patience and self-control. Apparently, longer periods of teaching flying improved the instructor's technical ability but spoiled his patience and personality.

Recommendations for Future Work

Introduction of practice teaching in central instructors schools.—Teaching cannot be learned by lectures and demonstrations alone, without supervised practice, any more than flying can. If Central Instructors Schools are going to emphasize teacher training in addition to standardized flying, it will be necessary for them to have an experimental school in which the student-instructor can practice teaching cadets under expert supervision. Such a procedure will be less unfair to the cadets involved than it would be for them to be in the first class taught by the same instructor in an ordinary school without special supervision. In two-engine planes the supervisor can ride from time to time with the instructor and his student; in single-engine planes magnetic wire sound-recorders and air-to-ground radios may be used to sample the performance. Lectures and discussions will be much more meaningful to student-instructors after they have had some experience in practice teaching. The practice-teaching school will also tend to keep the staff instructors in closer touch with the problems of cadet training.

Research should be devoted to developing special techniques for grading the student on his practice teaching so that eliminations from central instructors schools may be based on teaching as well as flying ability. Grades on teaching ability at the Central Instructors School should be validated against subsequent performance in the pilot training schools to which the instructors are assigned.

Measurement of teaching proficiency at pilot schools.—The 18-item student rating scale should be refined and shortened somewhat. On the basis of data collected as part of the validation of instructor selection tests, the test-retest reliability of each item, when the same instructors are rated by different classes, should be determined. In the light of these reliabilities and the intercorrelations presented in this chapter, the items which are least reliable and least independent should be eliminated.

The 16-item descriptive scale developed for supervisors to rate civilian instructors should also be refined and adapted for use at all levels of military flying instruction. In order to make any fundamental improvements in the supervisors' ratings of instructor proficiency, however, it will probably be necessary to make systematic attempts to see that he gets a better opportunity to observe his instructors in the actual teaching situation. In the reduced peacetime training program, it should be possible for two-engine supervisors to ride with their instructors and students more frequently and for single-engine supervisors to sample instruction by techniques such as air-to-ground radio or magnetic wire recorders.

Both the student and the supervisor rating scales should be used to allow the instructors to see themselves as others see them. It is believed that it would be highly desirable to try using the present form of the student rating scale for this purpose.

Since the ultimate test of the instructor is the quality of the students which he produces, it would be highly desirable to validate and item-analyze both the student and the supervisor rating scales against the quality of students produced by the instructor under controlled conditions. The flying aptitude of students assigned to different instructors should be equalized on the basis of pilot stanines in primary schools and on the basis of stanines and preceding flying grades in other schools. Transfer of students from one instructor to another should be minimized, and where transfer is absolutely necessary, the student should be dropped from the study. Special attention should be paid to securing the best possible measure of the student's proficiency. In addition to providing a basis for improving the scales, determining the items which best discriminate between instructors producing good and poor students would be a fundamental contribution to the job analysis of the qualities most important for the good flying instructor.

Improved student and supervisor rating scales would be useful: (1) In improving instructors on the job by letting them know their weak points; (2) in validating instructor selection tests, and Central Instructors School grades; (3) in studying the way in which factors, such as length of assignment as an instructor and other biographical data, are related to teaching proficiency; and (4) in measuring the results of training experiments conducted to evaluate the effect of procedures, such as adding variety to the life of the flying instructor by temporarily rotating him to other assignments.

Attempt to predict teaching success.—Further effort should be devoted to the difficult problem of trying to develop and validate a battery of tests and inventories for selecting those pilots who are most likely to become successful flying instructors. Such tests should be developed, not only for their possible usefulness in the peacetime program, but also as an important technique to have in reserve to deal with any future need for rapid emergency expansion.

CHAPTER FIFTEEN

Summary of Main Results and Recommendations for Future Work

Maj. Neal E. Miller

This chapter summarizes the main results of psychological research on pilot training during the second World War and presents the chief recommendations for further work. More detailed summaries and recommendations have already been presented at the end of each of the preceding chapters.

The opportunity for psychological research on pilot training developed gradually out of the original work on pilot aptitude tests by aviation psychologists. On 1 February 1944 an organization responsible for research on pilot training, the Psychological Research Project (Pilot), was established at Randolph Field, Tex. The work started slowly, but as the psychologists learned more about flying and the pilots responsible for training discovered the usefulness of the scientific techniques employed by the psychologists, research continued to develop at an accelerated pace.

The areas of research were determined by directives from higher headquarters. The two main areas were: To develop objective measures of flying skill, and to develop techniques for the evaluation and selection of flying instructors. Certain other types of work were also carried forward. An analysis of the job of the student pilot was made in order to provide a basis for developing objective measures of flying skill and for improving pilot aptitude tests. Subjective measures of flying skill and fixed gunnery scores were evaluated as possible criteria for validating the pilot aptitude tests and as supplements to the objective measures of flying skill. Since the combat pilot had to possess factual information as well as flying skill, printed tests of flying information were developed to test types of knowledge which could be measured more efficiently on the ground than in the air. The pressure of work which was directed by higher headquarters and was more closely related to the classification program, prevented the amount of attention being devoted to training experiments which this type of work is believed to merit. Two minor studies were conducted which illustrate the types of issues that can profitably be decided by the experimental method.

The following sections present the main results and the suggestions for future work in each of these areas.

ANALYSIS OF THE PILOT'S TASK

Military flying training is different from civilian in that it is conducted under greater pressure and requires more precision, but the basic principles are much the same.

Certain characteristics which make the task of flying more difficult than that of driving a car are determined by ways in which the air differs from the ground. A car is steered to the left or right on the surface of the ground. An airplane is steered in two additional ways: besides steering to the left or right in the air, the pilot must raise and lower the nose of his airplane and control the tilt of its wings. Because of the various ways in which these three types of movement interact, they more than triple the complexity of the task. For example, in making turns the stick and rudder must be used together properly. If too great pressure is applied to the stick, the airplane will tend to slip; if too much is applied to the rudder it will tend to skid. Because the lift of the wings is reduced when they are banked in a turn, the back pressure on the stick must also be increased. To cite another example, whenever the nose of a single-engine airplane is pointed up, a tendency for it to swerve to the left must be corrected by pressure on the right rudder, and if air speed is to be maintained, the throttle must also be advanced to increase the power.

The control movements for these three types of steering are relatively simple, much like those of driving a car. A large part of the increased complexity is perceptual, keeping track of the course of the airplane over the ground, of the attitude of the nose and wings, and sensing a lack of balance, or tendency for the body to sway to one side, if stick and rudder are not properly coordinated when rolling into turns.

The fact that the air is transparent and is without any reference points makes difficult the planning of a three-dimensional course through it and places a premium on visualization of spatial relationships. For example, in a forced landing the student must allow for the amount of altitude which will be lost on each turn and leg, planning a course down through space which will bring him to the ground headed upwind in the middle of a suitable field. At high altitudes, the earth appears to be moving very slowly and the pilot must respond to relatively slight changes in the movement of a distant terrain which is usually not clearly marked off for him like the sides and center of a highway. On the other hand, when the pilot is close to the ground in landing, speed blurs nearby reference points and he must respond to the position and movement of objects farther away.

A number of additional consequences flow from the fact that the air is a fluid, elastic medium. Crosswinds may produce drift. In landing, it is especially important for the pilot to correct for drift by pointing the nose slightly upwind and then straightening his airplane out just before it touches the ground. If he does not do this, there will be a sideward component to his landing which will strain the landing gear and may throw him into a ground loop. Discriminating the differences between the type of relative movement of the foreground produced by slight tendencies to drift and that produced by slight tendencies to turn presents the student with a new and difficult problem, especially since this discrimination must be made against an unsteady background of wavering movement produced by slight turbulence. Furthermore, the student must be doing a number of other things at the same time and constantly be alert to compensate for the effects of sudden gusts.

At high air speeds a slight movement of the controls will produce a considerable effect; when the air speed is lower, as in landing, much more extensive movements are required to produce the same results. Though the extent of movement differs greatly, the amount of pressure exerted on the controls in producing these movements remains approximately the same. The student, therefore, must learn to control his aircraft in terms of pressures instead of movements.

Since the airplane will stall and fall unless its speed is maintained, flying is paced. The student pilot cannot slow down or stop as an inexperienced driver can, whenever the task gets beyond him. Before he has had sufficient practice to enable him to make quickly and automatically the various types of perception and response, the new student is likely to be continually falling behind the pace of the task. As his skill increases, the student becomes better able to keep up, but before one maneuver becomes thoroughly learned, the fast-moving military curriculum is likely to confront him with a new and more complex task.

Closely allied to the paced nature of the task, is the problem of division of attention. Much of the difficulty of flying is derived not so much from the demands of each of the separate tasks as from the fact that the pilot has to keep track of a number of different things simultaneously. For example, in landing he must watch for other planes, plan to land in the first third of the field, correct for drift, keep his wings level, maintain the proper gliding air speed, judge his height above the ground, and finally, achieve a three-point attitude just before the airplane touches the ground. His attention must be shifted rapidly among all of these different factors. Furthermore, correcting any given error, for example, too much air speed, may require a little time. During this time the pilot must shift his attention to check other things, finally coming back to see, for example, whether he has slowed down to the proper air speed. Before the proper reactions to various stimuli have become more or less automatic, the need for keeping track of a large number of different factors places a severe strain on the student pilot.

In addition to the special demands placed upon a pilot by the characteristics of the new medium through which he travels, it is necessary for him to be well motivated in order to survive the rigorous course of training and to be adjusted emotionally to the unfamiliar and somewhat dangerous conditions of flying. He must also understand the design and function of his equipment in order to get maximum performance under the emergency conditions of combat and have a knowledge of weather and navigation in order to plan flights properly.

Flying is a complex task requiring many different abilities. The fact that the best printed tests in the classification battery used during the war were found to predict flying ability at least as well as the best apparatus tests, confirms the observation that perception, visualization of spatial relationships, knowledge of mechanical principles, and motivation are important factors in learning to fly. The fact that the addition of apparatus tests to this battery improved its ability to predict, suggests that motor skill and ability to perform a complex paced task are also important factors.

The flying experience of members of the Pilot Project has emphasized the difficulty of making exact statements about the relative importance of different abilities. The task of flying is complex and requires a large number of abilities. In many instances a given error or deficiency in performance may be produced by any one of a large number of causes. For example, the student may become tense and emotionally upset because he has difficulty in dividing his attention or he may have difficulty in dividing his attention because he is tense or emotionally upset. Furthermore, difficulty in dividing attention may arise because of poor native ability to discriminate some of the cues, because the student has not yet learned to respond automatically to the cues and hence must pay special attention to them, or simply because the student has not been told how most efficiently to shift his attention from one cue to another. Finally, one may put forward the hypothesis that some students cannot keep track of everything because they have narrow spans of attention.

It may be that attempts to make a more penetrating analysis of flying skill will not be very profitable until knowledge of simpler psycho-motor skills, in situations which are easier to control, has been increased and a clearer idea is developed of the general structure of human perceptual, motor and intellectual abilities. Until knowledge is greatly increased so that it is easier to put an analysis of the pilot's task into words, it will be desirable for psychologists working on pilot aptitude tests or training research to secure the necessary information about flying by taking a certain amount of systematic pilot training.

One of the most important uses of measures of flying proficiency was to separate the students into two groups: Those who should continue and those who should be eliminated from further flying training. Because of its great practical importance, more care was devoted to deciding who should be failed than to assigning other flying grades. Usually the check rider examined the student's previous grades and discussed them with the instructor before the flight in order to know what weaknesses to look for. After the check ride, he almost invariably took into account the opinion of the instructor which was based on much more extensive contact with the student. Students were given an average of four check rides during primary training. The clear-cut cases (best and poorest students) received fewer than four while doubtful cases were given as many as a dozen check rides. The system of grading these rides was entirely subjective in terms of categories such as A, B, C, D, E, and F, with a great concentration of grades in the category of C.

As with all systems of subjective grading, there were considerable differences in the standards of different check pilots in the same school and of different groups of the check pilots at various schools. These were reflected in the wartime elimination rates of different classes from different schools which ranged from a low of 10 percent to a high of 60 percent. The differences between schools and classes were much greater than would be expected by chance. They could not be accounted for by differences in the average level of aptitude of the students as measured by the pilot stanine. Such differences are of course inefficient because they mean that whether or not the student is eliminated may depend as much on whether he happens to be sent to an easy or a difficult school as upon his flying ability.

When flying grades are used as a basis for elimination, they are really predictions of what the student will do in subsequent training. A grade of "B" should mean that the student will have a good chance to succeed in subsequent training while a grade of "F" should mean that the student will be dangerous to himself and others if he is allowed to continue.

The correlation between different grades given at the same school was not a very useful index of their true validity in predicting flying ability because the two grades were never independent judgments. As has already been pointed out, the check riders and instructors always discussed the student in order to secure the best possible picture of his total performance. When grades at one school are compared with those at the next, for example Primary compared with Basic, this error does not occur because the instructors in the more Advanced school do not pay much attention to the grades which have been given at the more elementary level.

Whenever the grades at one level of training were compared with those at the next, the correlations were invariably quite low, ranging between 0.20 and 0.35. In a typical study on 2,905 students, the correlations between grades in Primary and Basic was 0.27. Even when allowances are made for the fact that the range has been restricted by elimination of the poor students at the Primary level, these correlations indicate poor prediction. In fact, the pilot stanine, based on objective tests administered on the ground to the students before they entered preflight school, seemed to predict the flying grades in Basic school as well as did Primary grades which were based on subjective observation of ten weeks of flying training in the air.

Like most subjective judgments, the flying grades appear to involve a strong halo effect. The intercorrelations among grades for different maneuvers and different aspects of flying were quite high, suggesting that the instructor was influenced by his general opinion of the student and tended to minimize the weak points of good students or the strong points of poor ones. These high intercorrelations reduced the possibility of using grades on different aspects of flying performance to diagnose specific weaknesses for corrective training, to validate specific aptitude tests, or to select students for different kinds of specialized training.

Scores derived from the comments which the instructors wrote on the students' errors had lower intercorrelations and showed more promise of possible usefulness as differential criteria. A factor analysis of these scores yielded several independent factors some of which were predicted by the pilot aptitude tests and others of which were not. Thus it suggested areas in which new tests should be developed. A check list derived from an analysis of the instructor's comments would probably be a useful addition to the present grade slip.

OBJECTIVE MEASURES OF FLYING SKILL

In objective measurement absolute (in addition to relative) agreement among observers is achieved by using a permanent independent standard. Instead of using subjective standards such as "poor" or "too slow" which are dependent upon the individual making the judgment, the performance is defined in terms of a permanent standard, such as a reference point on the plane or an instrument reading, which is relatively independent of the specific observer. To the extent that such standards can be clearly defined, different observers will give the same absolute grade to the performance. It is obvious that if this type of standard can be achieved, it will tend to reduce the differences among the elimination rates at different schools.

In addition to achieving absolute agreement among observers by the use of a permanent independent standard, the task must be clearly defined to the student and administered under controlled conditions. In the first place, the student must know exactly what he is expected to do; he must be told how the various measures of a given maneuver will be scored and weighted. This is especially necessary in flying because it is a complex task in which the student can improve one part of his performance at the expense of another by shifting his attention. If the task is not clearly defined, two students with the same ability may get different scores because they have different opinions concerning what they are supposed to do. Similarly, unless conditions are standard, the scores will reflect variations in these conditions as well as in the skill of the student. The airplane is a complex and variable piece of equipment and the air is an unstable laboratory with changes in wind, turbulence and density. For these reasons control of the conditions of testing is an especially difficult problem in securing standardized, objective measures of flying skill.

Most of the work of the Psychological Research Project (Pilot) in this new area was concentrated on developing methods for enabling the check pilot or instructor to secure objective scores using the equipment available on the average army airplane. This approach made it possible to secure a broader view of the relatively unexplored area of research than would be possible if work had been concentrated on the much slower process of developing recording instruments and analyzing their records. Furthermore, since no new equipment is required, it is much easier to put the measures developed into immediate practical use. After the area is thoroughly explored, a better basis will be available for the logical development of recording devices for situations in which more accurate measurements are needed. The general soundness of this approach was indicated by the fact that in many instances it has been possible to make the measures objective enough so that observers agree, only to find that the chief source of difficulty was not in errors of measurement but in erratic day-to-day fluctuations of performance.

Sources of Unreliability

In order to be useful, a measure must be reliable. If one administration of a measure does not predict a second administration of the same measure, it cannot be expected to predict anything else.

Lower reliability when test and retest are administered on different days.—One general finding, in the research of the Pilot Project at all levels of training, is that the agreement between repetitions of the same objective measure on different days is much lower than that between repetitions on the same day. At the Primary level of training, the reliability of objective measures on 340 pairs of accuracy landings with test and retest in immediate succession was 0.12 for zone of landing, 0.32 for landing attitude, and 0.32 for whether the airplane was dropped in or bounced more than 3 feet. When a test and retest were on different days, in different airplanes with different check riders, these reliabilities were reduced to 0.02, 0.04, and 0.00 respectively. In a similar comparison on a representative high altitude maneuver, the reliability of altitude deviation in a steep turn was found to be 0.38 when the turns were given in immediate succession and 0.17 when they were on different days and in different airplanes by different check riders. Similarly, for instrument flying at the Basic level of training, the reliability of altitude deviation in straight-and-level flight was 0.59 when test and retest were given in immediate succession and 0.21 when they were administered on different days. For altitude deviation in instrument steep turns, the correlation was 0.55 for immediate succession and 0.24 for different days. Finally, the same type of difference was observed for the total score on the check ride to measure instrument flying in the AT-6. The correlation between the odd and even items of this ride (corrected for double length) was 0.91 on the first day and 0.89 on the second day, but the correlation between the total scores on these two days was only 0.46.

The large drop in reliability when test and retest are administered on different days does not seem to be limited to objective measures of flying skill; it is also found in other measures of complex types of performance in the air. In measuring "distance off," the ultimate criterion of successful navigation, the Psychological Research Project (Navigator) found that the reliability of different legs of the same mission was 0.44. When the results of missions on different days were compared, however, the reliability dropped to −0.07. Similar results were found for the computations of air speed drift, and other specific aspects of the task which contributed to the final result of "distance off."[94]

Flexible gunnery scores present a similar picture. In one of the best controlled studies using the gun camera and B-29 sighting equipment, the reliability of attacks within a mission was 0.72 while the correlation between missions flown on different days in different airplanes dropped to 0.27.[95] Finally, in a study of bombing accuracy the correlation between measures of circular error was found to be 0.56 for odd-even releases and 0.03 for odd-even missions.[96]

Implications of drop in correlation when test and retest are given on different days.—The drop in the reliability of objective measures of flying skill when test and retest were given on different days in different airplanes by different check riders has three important implications: (1) Any estimates of reliability which are based on the comparison of odd and even items or test and retest within the same

[94] See chapter 6 of Report No. 10 of this series, *Psychological Research on Navigator Training.*

[95] An unpublished study by the Department of Psychology, School of Aviation Medicine. See also chapter 9 of No. 11, *Psychological Research on Flexible Gunnery Training*, prepared by Psychological Research Unit, No. 11.

[96] See the discussion of table 2.11 in Chapt. 2 of Report No. 9, *Psychological Research on Bombardier Training.*

flight will greatly overestimate the type of reliability which has significance in the practical situation, namely the ability of the measure to give consistent results over a period of time; (2) in any check ride used to measure the results of training experiments, it is important to control the effects of conditions of testing by measuring matched pairs of experimental and control subjects as nearly in succession as possible in the same airplane and by the same check rider; (3) if a reliable measure of the individual student is desired, check rides must be repeated on a sufficient number of days in different airplanes by different check riders so that it will be possible to average out the effects of variations in conditions of testing and of fluctuations in the individual's ability to perform.

Relatively good agreement between independent observers.—The fact that the measures are considerably more reliable when test and retest are given in the same flight suggests that a considerable part of the low reliability when test and retest are given on different days is caused by variations in the performance measured rather than by errors in measuring that performance. This conclusion is further confirmed by the fact that there is fairly good agreement between measurements of different observers scoring the same thing at the same time. For accuracy landings at the primary level of training, the scores by observers in the airplane and on the ground correlated with each other 0.83 for zone of landing, 0.79 for landing attitude, and 0.88 for dropped or bounced more than 3 feet. On a preliminary form of an advanced two-engine check ride, 16 out of the 24 scores had observer reliabilities of 0.80 or above.[97] While these observer reliabilities are not perfect, they are considerably higher than the test-retest reliabilities. This further confirms the conclusion that the low reliability of test and retest on different days is to be attributed to erratic changes in performance rather than to errors in measuring that performance.

Suggestions for further work.—The results which have just been presented suggest that future research should be concentrated on trying to standardize conditions of testing and to minimize the effects of day-to-day variability in performance. There is a need for a more exact and detailed analysis of the causes of the day-to-day variability. A study should be directed toward determining the relative amounts of variance introduced by the observer, by the airplane, by weather, and by changes in the level of the performance of the individual. A more exact knowledge of the causes of variability should provide a better rational basis for devising means to cope with the problem.

The problem of variability in performance is, of course, a general problem which is not limited to performance in the air. There is

[97] Each of these studies involved a number of different pairs of observers. Since no corrections were made for any differences between the different pairs of observers, the correlation coefficients represent absolute as well as relative agreement.

some suggestion in the work of the Department of Psychology at the School of Aviation Medicine that as psychomotor tests are made more complex, the reliability of measurements based on a given length of performance tends to be reduced. It seems possible that the more readily controlled apparatus test situation could be used to discover certain fundamental principles which, in turn, might help to guide research on the development of objective measures of flying skill. It would be insteresting to use this kind of test to study problems such as (1) The effect of the complexity of a task upon its reliability, (2) the effect of the interval between test and retest upon their correlation, (3) how this effect varies with the complexity of the task, (4) the effect of different methods of scoring upon reliability, and (5) the effect of having the subject know exactly how he is scored upon the reliability of the score.

One of the ways to deal with variability of performance is to average it out by repeating the measurements on a sufficient number of flights. This suggests the desirability of further work on objective measures for the daily grade slip.

Relationships Among Different Indices Used to Evaluate Measures

Each of the objective measures of flying skill developed by the Pilot Project was evaluated in a number of different ways. It was possible to gather a certain amount of data on the relationships among various criteria for the goodness of each measure.

Relationship between reliability and validity.—For the 81 items on a scale of instrument flying at the Basic level of training, the validity of each measure was determined by its correlation with instructors' ratings. The correlation between this index of validity and the test-retest reliability of each of the 81 measures was 0.34. At the Primary level of training comparable data were available for 63 measures. For these measures, the correlation between test-retest reliability and validity as measured by ability to predict pass-fail was substantially zero. The correlation between the reliability of a measure and its ability to discriminate between students with different amounts of training was also zero. Another study on 26 measures at this same level of training also showed no relationship between reliability and validity. The lack of relationship between reliability and validity for the Primary measures was produced in part by a number of measures in which the time for the pilot to execute various maneuvers such as turns was recorded. These measures of time had very high reliabilities, indicating that the pilot's performance was consistent in this respect from day to day. They had, however, a very low validity, indicating that this aspect of his performance, though consistent, was not relevant to his skill.

Relationship between ability to discriminate between different amounts of training and ability to predict pass-fail.—At the Primary level of training two different criteria were used to evaluate measures of flying skill: ability to discriminate among students with different amounts of training, and ability to predict whether the student would pass or fail. Each student was tested on two successive days. The first criterion was determined separately on each of the days; the second was determined for the average of the two days' performance. The correlation between the ability of a measure to discriminate between students with different amounts of training on the first day and this same ability on the second day was 0.41 for the 63 measures on which complete data were available. This correlation is significant at the 0.1 percent level and indicates that the groups used in this study were large enough to determine with some reliability the rank order of the measures in respect to this criterion.

The correlation between the ability of each measure to discriminate between students with different amounts of training and its ability to predict pass-fail was, however, substantially zero. If this lack of relationship is found in other studies, it will mean that these two types of criteria are quite different and that the errors characterizing the students who are deficient in experience are different from those characterizing students deficient in aptitude. If the relationship between these two criteria is low, two different scales will have to be developed: One for diagnosing weaknesses in training, graduating students on a proficiency basis and measuring the results of training experiments, and another for deciding which students should be eliminated and validating aptitude tests.

Types of Measures Developed and Evaluated

The largest part of the work of the Pilot Project on developing objective measures of flying skill was devoted to devising, trying out, and evaluating different measures. A considerable portion of this work was concentrated on developing measures relevant to the earliest stages of training in primary flying schools. Since this was the period when the greatest number of eliminations occurred, it was believed that more accurate methods of measurement would be of greatest practical value during this stage of training. Most of the remainder of the work was concentrated on objective measures of instrument flying skill, because of the extreme importance of instrument flying in combat and transport, and because this type of flying seemed to lend itself to objective measurement.

The number of measures on which empirical data have been gathered at each level of training is summarized in table 15.1. In making this table minor variations of a measure were not counted. Thus, altitude range in steep turns to the right and steep turns to the left was considered as one measure. However, altitude range in turns, altitude range in straight-and-level flight, and altitude with instruments covered and instruments visible were considered separate

TABLE 15.1.—*Number of objective measures of flying skill experimentally tested at each level of training*

	Number of measures investigated	Number of correlations calculated
Primary phase of contact flying	131	1,178
Basic phase of instrument flying	78	176
Advanced twin-engine (TB-25) phase of instrument flying	112	181
Advanced single- and two-engine contact, basic contact, et al	202	509
Total	523	2,044

measures. In order to give a bird's-eye view of the amount of data gathered, the table also reports the number of correlations which have been calculated for the measures in each category. When the measures for each level of training are summed, it can be seen that a total of 523 different measures were tried out and that in evaluating these a total of 2,044 correlation coefficients were calculated.

Different Methods of Scoring

Three different methods of scoring instrument flying were compared: (1) The time-sample method in which the check rider read the deviations from the correct altitude, heading, etc. at specified intervals throughout the maneuver; (2) the range method in which he scored the difference between the highest and the lowest reading on an instrument during a maneuver; and (3) the limits method in which he scored the single largest deviation (in either direction) from the correct reading. In all three methods the observations were recorded by marking on diagrammatic representations of the instrument, i. e., altimeter or directional gyro. It was found that, although the time-sample method had a higher observer reliability, it did not have a higher test-retest reliability or higher correlation with instructors' grades than the range method. Furthermore, the time-sample method was harder to administer. In one study the range method seemed to have higher test-retest reliabilities and higher correlations with instructors' grades than the limits method; in another, the differences were quite inconsistent though tending to be in the same direction.

Combining Measures into a Total Score

In connection with the measures of instrument flying skill for the basic level of training, a little work was done on investigating the most effective ways of combining separate measures into total scores. In representative maneuvers, different measures (for example, altitude, air speed, and heading) were combined in three different ways. In one of these methods of combination the different measures were added up with equal weights. In another method the amount of weight given to each measure was determined by regression equations to compute beta-weights maximizing the validity of the combined score. In the third method the weights were determined by the expert judgment of pilots. The weights which the pilots recommended amounted to a modified type of cut-off score in which the student was penalized more for a single bad error in any one measure than for moderate errors in a number of measures. These three methods were tried out in getting total scores for measures on each of two representative maneuvers: Level out, and instrument turns. When allowance was made for the moderate amount of shrinkage which would be expected when the beta weights were cross-validated, these three different methods yielded approximately equal results.

In selecting and combining objective measures of flying skill, a special problem has been encountered. Sometimes it is necessary to include a measure which in itself is not reliable or valid in order to make sure that the student adheres to the conditions which make another, more important measure valid. For example, the amount of altitude lost or gained in a steep turn is a relatively good measure. In order to make sure that the student performs a steep turn, however, it is necessary to have some measure of his angle of bank. If the student is not penalized in some way for failing to maintain a proper angle of bank, there is nothing to prevent him from making a shallower turn in which it is easier to hold altitude. If he is simply required to repeat turns on which his bank is not satisfactory, there is nothing to prevent him from shallowing out the bank and hence getting another chance at the turn as soon as he sees that his altitude control is not going to be very good. Thus, this measure which is not very good on its own merits must be included and given a weight in order to establish a condition for administering a better measure, altitude deviation. This illustrates one of the difficulties inherent in trying to measure a complex skill like piloting an airplane.

Scales Composed of Selected Measures

The best of the measures devised and evaluated at each level of training were selected for further study. The tentative scales composed of these measures are described in the chapters dealing with the different aspects of this work. Only one of these scales had been tried out, before the end of the war disrupted training to the extent that further research had to be suspended temporarily. This scale was composed of 81 measures of instrument flying skill for the basic level of training. The odd-even reliability of this scale (corrected for double length) was 0.91 for the first day of testing and 0.89 for the second. The reliability of test and retest on different days in different airplanes was 0.46. From this it can be estimated that the reliability of the sum of the scores for two check rides on different days would be 0.63. The sum of the scores for two check rides on different days correlated 0.51 with instructors' ratings. The check rides were given at the end of training, the instructors' ratings were made before these check rides on the basis of observations throughout training, and the check pilots who administered the check rides were unaware of these ratings.

Suggestions for further work.—One of the next steps in evaluating such a check ride is to determine its ability to predict performance in the following, or advanced, stage of training. Another step is to determine the reliability and validity of a check ride composed of the best items selected from this scale.

The ability of the items in this scale to discriminate between students with different amounts of training should also be determined. In addition to supplying an index of the usefulness of these measures in evaluating the results of training experiments, such a study would yield further evidence on the fundamental question of whether or not the measures which are best able to discriminate between students with different amounts of training are also best able to discriminate between students with the same training but different aptitude.

At the same time that further data were being gathered from a larger number of cases on the reliability and validity of the items and of the total score, a study should be made of the intercorrelations of the different items. This might throw light on the structure of flying skill and would indicate any duplication which should be eliminated.

Similar studies should, of course, be carried out on the scales developed for different types of flying at other levels of training. A suggested experiment is described in more detail at the end of Chapter 6.

USE OF OBJECTIVE MEASURES IN A LARGE-SCALE STUDY OF THE EFFECT OF ADDITIONAL TRAINING ON FLYING SKILL

While it was still in the early stages of developing objective measures of flying skill, the Pilot Project was asked by the Chief of Staff of the Training Command to use the opportunity afforded by some temporary training freezes to develop objective measures for all levels of training and to measure the effects of additional training at each level.

The objective check rides developed for each level of training were designed to sample differences between groups; they were not designed to be reliable enough to discriminate between individuals. These objective measures were administered at 36 different schools during a single week to test the flying skill of over 8,000 students in four different types of training.

In the primary, basic, and advanced single-engine schools, students who had received the normal 10 weeks of training and those who had received an additional 5 weeks during the freeze (making a total of 15 weeks) were tested at the same time under comparable conditions.

The results showed that at each of these three levels of single-engine training, adding an additional 5 weeks produced an improvement in the performance of the students. At none of these levels had the students reached their limit of improvement within the normal period of training.

The fact that the students improved with additional training at a given level does not mean, however, that they would necessarily benefit more from that training than from being moved on to a more advanced type of airplane. The question is, to what extent are the effects of additional training on a lighter plane transferred to performance in flying a heavier one? It was possible to secure some evidence on this question in the measures at advanced two-engine schools because the training freeze happened to coincide with a change to a different kind of airplane. One group of students was available which had received 10 weeks of normal advanced two-engine training on the lower powered AT-17 or AT-10 followed by 5 weeks on the faster and heavier TB-25 (Mitchell bomber), making a total of 15 weeks. Another group of students had received 10 weeks of advanced two-engine training, all on the TB-25. The results showed that, in general, 10 weeks of training, all on the TB-25 was better than 15 weeks of training, 10 of which was on a lower-powered two-engine airplane and only 5 on the TB-25. Apparently, the transfer of training was imperfect enough so that 5 weeks on the larger airplane was of more value in learning to fly it than 10 on the smaller one.

Another group of students had 15 weeks of training, all on the TB-25. In general, these students were the best of all, but the additional training did not help all aspects of their performance. During the additional training, there was a tendency to forget those aspects of procedure which were emphasized during the first weeks and relatively neglected later on.

Finally, single-engine students who had had enough civilian flying to solo before entering primary were compared with those who had not. It was found that during the earlier phases of military training, the students who had soloed before primary were definitely superior to those who had not. The difference between these two groups, however, did not seem to persist beyond the basic level of flying training. These results are in line with what is known about the effects of previous civilian flying upon the likelihood that the student would be eliminated at the different levels of training. The results of investigations on fixed gunnery also indicated that by the time the students had reached that advanced stage of training, there was no difference between those who had soloed before entering primary and those who had not.

The fact that it was possible to achieve these results when the measures were administered to over 8,000 students in a single week indicates that it is practicable to employ objective measures of flying skill in evaluating the results of a large-scale pilot-training program.

GENERAL RECOMMENDATIONS FOR DEVELOPING OBJECTIVE AND SUBJECTIVE MEASURES OF FLYING

A scientifically constructed grade slip.—Day-to-day fluctuations in performance appear to be one of the greatest sources of difficulty in securing good measures of flying proficiency. One way of dealing with these is to average them out by measuring performance on a sufficient number of days. The difficulty in administering a large number of special check rides suggests that it would be desirable to investigate the possibility of improving the daily grade slip which is filled out by the instructor on each dual ride. Various experimental forms of grade slips should be tried out and scientifically evaluated. Objective measures, relevant to the particular stage of training, should be included in the grade slips. Such measures will serve as useful points of reference and it is believed that they will tend to exert a standardizing influence upon the entire grade slip.

A check list should be devised for recording the most frequent and important types of errors as determined by analyses of instructors' comments similar to the ones reported in chapter 5. Such a check list will tend to standardize the comments and make it much easier to determine, for example, which kinds of errors in Primary are most predictive of poor performance at Basic. A blank space should be provided for writing in comments not adequately covered by the check list.

Finally, this grade slip should contain such subjective grades as may be necessary to deal with aspects of performance which cannot easily be covered by the objective measures or the check list.

Some simple system should be worked out for summarizing the facts on the grade slip into total scores on over-all performance, and possibly on a few different relatively independent aspects of performance. There should also be a simple system for summarizing the daily grades on each subphase of training at a school. It is believed that if sufficient ingenuity is devoted to the problem, a grade slip can be designed which will contain more valuable information in a more usable form, and yet be easier to fill out than it is to write in comments conscientiously on the present type of grade slip.

Each item on this grade slip should be scientifically evaluated; the predictive powers of the objective items, the check list and subjective items should be compared. It would also be desirable to compare the predictive power of the instructor's over-all subjective evaluation with that of a total of the scores on each of the items. In making his over-all subjective evaluation of the check ride, the instructor should be allowed to know the results of studies showing the correlation of each type of error with failure in subsequent training.

The results of a study, reported in chapter 8, in which students were graded on each dual instrument ride, appears to indicate that this more rigorous type of grading does not interfere with training, but if anything, produces students who can fly with greater precision.

Mechanical recording.—As relevant aspects of performance are located by thorough exploratory work, it may be worth while to try to develop mechanical aids to simplify the check pilot's task. In some cases it may be necessary to make a detailed record of performance in order to decide which aspect should be measured. In general, however, it is believed that the recording devices should be designed to yield total scores which can be immediately read from a dial for each aspect of behavior sampled. For example, a work adder or a movement-counter type of device might be attached to the throttle, stick or rudder. It might be desirable to have the device record changes in pressure on the stick and rudder rather than movements. A simple photo-electric recorder might be designed to integrate the total amount of movement of the needles on the altimeter, directional gyro, etc. All of these devices should be designed so that the check pilot could start them at the beginning of a maneuver and stop them at the end, or so that they could be set to run for a given period of time. Then the check pilot could read the dials and zeroize them at leisure after each exercise was completed. It might be possible to design for each instrument a pair of exceedingly light auxiliary hands which could be zeroized at the beginning of a maneuver and would record the maximum deviation in either direction on that instrument during the maneuver.

FIXED GUNNERY

Flying the airplane in such a way that the bullets from the guns fixed to it would hit a target was a relevant part of the fighter pilot's task, required considerable skill, and yielded an objective score. A number of studies of fixed gunnery were made. First, the differences between classes and schools were studied and found to be important. It was found difficult to control these differences or to develop corrections for the underlying factors such as geographical and seasonal variations in turbulence and visibility. The best way to secure the most relevant scores, therefore, seemed to be to normalize the data, or analyze them separately, for each class and school.

Although the results of a single day's firing were comparatively unreliable, the results for the entire period of training yielded relatively reliable scores. For 1,064 students the odd-even reliability (corrected for double length) of 1,200 rounds of air-to-air firing was 0.63. Similarly, for 932 students the reliability of 400 rounds of air-to-ground firing was 0.59. Although the students were learning throughout this performance, an empirical check demonstrated that the reliability varied with the number of rounds sampled in approximately the way which would be predicted from the Spearman-Brown formula. That somewhat different skills were involved in air-to-air and air-to-ground firing was indicated by the fact that for 1,175 students the correlation between these two types of scores was only 0.22.

The equally-weighted sum of air-to-air and air-to-ground fixed gunnery scores in the Training Command correlated 0.38 with a similar score in the First Fighter Command. The fact that the correlation is no higher than this suggests that the task of fixed gunnery in the tactical airplanes in the First Fighter Command is considerably different from that in the lighter airplanes in the Training Command. In view of this difference and the fact that the learning curves were still going up steeply at the end of the amount of practice given in the Training Command, it would seem highly desirable to conduct an experiment to see how much different amounts of training in fixed gunnery in the Training Command would transfer to performance in the First Fighter Command.

Finally, on a sample of 1,000 students, the pilot stanine was found to correlate 0.19 with an equally weighted combination of air-to-air and air-to-ground fixed gunnery scores. Those tests in the classification battery which had the highest correlations with the criterion of pass-fail in primary training also tended to have the highest correlations with fixed gunnery scores in transition training.

PRINTED TESTS OF FLYING INFORMATION

Need and Method

Reports from combat theaters stressed the fact that, in addition to flying skill, the combat pilot had to have good judgment, based on a thorough knowledge of practical aerodynamics, engines and equipment, navigation, weather, and principles of instrument flying. In order to get maximum performance, the pilot had to have a knowledge of the flying characteristics of his airplane. Since enemy fire created sudden emergencies in the air, he had to have a comprehensive understanding of what his plane and its equipment were designed to do and how it was affected by unusual conditions. To cite one more example, the pilot sometimes had to balance the probabilities of icing or head winds at certain altitudes against the true air speed and rate of fuel consumption. Though he received advice from his crew members, the final decisions rested with him. Having the proper knowledge did not insure sound judgment, but without a certain minimum of knowledge, good judgment was impossible.

In constructing printed tests to measure flying information, the goal of the Pilot Project was to measure the types of information needed by combat pilots, irrespective of whether or not this information was covered in the present curriculum. Test items were constructed with the aid of experts in each of the relevant areas of information. The items were given a trial administration, distributions of responses to each alternative and the internal consistency of these responses were determined. These data were discussed with the experts and the items were revised. Finally, the reliability of the total test and the reliability and intercorrelations of each of its subsections were determined.

Tests Constructed

Two main tests were developed. One of these, Pilot Information Test, Form 3, contained subsections dealing with flying, navigation, aeroequipment, instrument flying, weather, and personal equipment (covering the use of oxygen, etc.). This test required approximately 3 hours total administration time and had a satisfactorily high reliability of 0.85, determined by the Hoyt method. The score on this test, along with other variables such as flying grades, was used in selecting the best pilots to be grouped with the best navigators, bombardiers, and flight engineers, for training as potential lead crews.

Two alternate 150-item forms of a Pilot Instrument Flying Information Test were also developed. This test was developed to supplement the oral examination which had been used along with the flight check to determine whether the pilot should be given an instrument card allowing him to fly in instrument weather. The AAF Instrument Flying Standardization Board at Bryan recommended these forms for this use throughout the Army Air Forces, and at the time this report was written, they had been adopted for standard use throughout the Training Command.

A 150-item Pilot Information Test was also developed for experimental use in the battery of instructor selection tests which were administered to combat returnees at redistribution stations.

Recommendations

As a result of the experience gained in developing and administering these tests, the Pilot Project recommended that a central testing agency be established for the Army Air Forces. This agency should have a staff of experts on tests and measurement. It should be charged with the responsibility of supervising all printed tests used in the Army Air Forces and of using such tests to evaluate and standardize the efficiency at all levels. Members of this agency should visit stations to give them technical advice on test construction and to inspect their testing programs. Whenever there is a need to test a large number of students on the same subjects at different stations, it should consult relevant experts and use its superior central facilities for the construction of the tests to distribute to the stations. It should conduct research studies, involving item- analyzing, intercorrelating, and determining the reliability of its tests. It should also be responsible for constructing tests to measure the learning and retention of subject material considered to be especially important for safe flying. Reports of periodic surveys on the proficiency of students, of instructors and of operational pilots should be presented to the Office of Flying Safety and to Headquarters AAF. In time of war, this agency should have experts in the field to assist the various combat air forces in making check-out tests especially adapted to the latest conditions in each theater of operation.

EVALUATION AND SELECTION OF FLYING INSTRUCTORS

Qualities of the Good Instructor

Supervisors, instructors, and students all agreed that the following qualities were most important for the good flying instructor: patience and self control; ability to adapt teaching methods to individual students; flying ability; ability to analyze errors; and ability to express himself.

Work on Civilian Flying Instructors

On the basis of the qualities which job analysis studies had shown were most important, a 16-item scale was developed for supervisors to use in rating the civilian flying instructors at primary schools. The correlation between the total scores of ratings by supervisors and assistant supervisors was 0.65, a figure which is too high as an estimate of reliability since these two raters had probably influenced each other's opinion of the instructor. The intercorrelations of the items in the scale ranged from 0.26 to 0.90, indicating a considerable halo effect among some items and a fair amount of independence for others.

This scale was used to validate the scores on a battery of nine instructor selection tests and inventories which had been constructed on the basis of the job analysis studies. The multiple correlation between the score on the 16-item scale and the battery of nine tests was 0.35 for a sample of 433 instructors. The correlation of the tests with a single over-all rating of teaching proficiency was in the same range as the correlation with the total score based on the 16 separate ratings. The value of securing ratings on separate items, however, is shown by the fact that the battery of tests correlated with the teaching qualities indicated by some items on the scale, while it left those indicated by other items completely unmeasured. Thus the use of the separate items suggested areas in which further test development is particularly needed. Although these tests were designed to predict teaching ability, it is interesting to note that their correlation with the rating on pilot ability was higher (0.46) than their correlation with any other item or with the total score on teaching proficiency.

Student rating scale.—An 18-item scale was also developed in order to secure ratings on advanced and transition school instructors by their students. When each instructor's students in a given class were divided into equal groups and the ratings of these groups were correlated with each other, the estimate of split-group reliability (corrected for the full number of students) was 0.75 for a sample of 1,395 instructors. Correlating the average of the ratings by students in one class with that of students in a different class at least 5 weeks later, yielded a test-retest reliability of 0.36 on a comparable, but smaller sample of 277 instructors. Both of these figures are based on an average of 3.25 students per class for each instructor. The marked reduction of the size of the correlation when the students came from the different classes was probably caused by removing the opportunity for the students in the different groups to influence each other. It is also possible that the instructors changed somewhat during the period of five or more weeks between the two ratings.

One of the 18 items on the scale was an over-all rating. For different classes the test-retest reliability of this item was 0.23. The fact that the reliability of the total score was definitely higher than that of the single over-all rating, indicated the value of rating different aspects of performance separately and then adding up these ratings.

The intercorrelations among the items of the scale ranged from 0.12 to 0.76, with the majority of them below 0.40, indicating that the different items were reasonably independent.

The correlation between the total score on the student scale and the supervisor's over-all rating of instructor proficiency was only 0.08 on a sample of 1,141 instructors in advanced and transition schools. The Navy also found a similar lack of relationship between students' and supervisors' ratings of flying instructors.

Tests failed to validate.—A battery of 6 instructor selection tests and inventories, some of them quite similar to those which seemed to be best in the battery administered to civilian instructors, was validated against the total score on the student rating scale for a sample of 215 instructors, and against supervisor's ratings for a sample of 179. None of the tests in this battery, nor the weighted score based on all of the tests, showed any appreciable positive correlation with either of these two criteria. The fact that these tests showed no ability whatsoever to predict the teaching success of combat returnees, while similar tests administered to the civilian flying instructors had shown at least a slight relationship to their teaching ability, indicates that the problems involved in devising and validating a satisfactory battery of tests to select flying instructors are still far from solved.

Tests devised by the Navigator Project also showed no ability to predict on-the-job measures of the teaching performance of combat returnees assigned to teach navigation. The Bombardier Project, however, found that similar tests had an appreciable ability to predict teaching performance of bombardier returnees.[96]

The Pilot Project found also that preference expressed at a Redistribution station did not predict how well a pilot returnee would like instructing after he arrived at a pilot school and started to teach flying. Once he was on the job, however, the flying instructor's like or dislike for his assignment was quite stable. The correlation between two ratings of liking for the job made at least 5 weeks apart was 0.82 for a sample of 277 instructors.

Relationship of Biographical Factors to Good Flying Instruction

Although no relationship was found between the test battery and ratings of teaching proficiency, interesting relationships were found between certain items of biographical and personal data and the students' and supervisors' ratings of teaching proficiency. Data were secured on 1,284 instructors. It was found that the older flying instructors received reliably higher ratings as teachers than the younger ones from both their students and their supervisors. The instructors with more education and those with a higher military rank were also rated as reliably better by their students, but not by their supervisors. Students gave combat returnees better ratings as instructors than nonreturnees but supervisors gave the nonreturnees a better rating in both teaching and flying. Instructors who liked their assignments were rated as better by the supervisors but not reliably so by their students.

The relationship between length of experience as a flying instructor and the supervisor's rating on teaching contained a reversal; the ratings became higher with more experience up to 2 years, but were somewhat lower beyond that point. The reversal was statistically reliable. The total score on the Student Rating Scale showed no statistically reliable difference for instructors with different amounts of experience. An analysis of each item of this scale, however, showed that the instructors who had taught flying for a longer period of time received reliably higher ratings on technical aspects of teaching ability, such as analysis of errors and ability to express themselves, and reliably lower ratings on personality items, such as interest in the student, patience and self control. Apparently, long periods of teaching flying improved the instructor's technical ability but spoiled his patience and personality.

Recommendations

On the basis of the experience gained in this research, a number of recommendations are made. It is believed that teaching cannot be

[96] See Reports No. 9 and 10 in this series, Psychological Research on Bombardier Training and Psychological Research on Navigator Training, respectively.

learned by lectures and demonstrations alone, without supervised practice, any more than flying can. Therefore, if Central Instructors Schools are going to emphasize teacher training in addition to standardized flying, it is believed that it will be necessary for them to contain an experimental school in which the student instructor can practice teaching cadets under expert supervision. Research should be devoted to developing special techniques for grading the student instructor on his practice teaching so that elimination from Central Instructors School may be based on teaching as well as flying ability. Grades on teaching ability at the Central Instructors School should be validated against subsequent performance in the pilot-training school to which the instructor is assigned.

It is believed that it would be useful to use the Student (and perhaps also the Supervisor) Rating Scale to allow the instructors to see themselves as others see them.

In order to make any fundamental improvements in the supervisor's rating of instructor proficiency, it will probably be necessary to make systematic attempts to see that he gets a better opportunity to observe his instructors in the actual teaching situation. In the reduced peacetime training program, it should be possible for two-engine supervisors to ride with their instructors and students more frequently and for single-engine supervisors to sample instruction through use of air-to-ground radio or magnetic wire recorders.

Since the ultimate test of the instructor is the quality of the students he produces, it will be highly desirable to validate and item analyze both the Student and the Supervisor Rating Scales against the quality of students produced under controlled conditions. It is believed that improved Student and Supervisor Rating Scales will have the following useful functions: (1) To improve instructors on the job by letting them know their weak points; (2) to validate instructor selection tests and Central Instructors School grades; (3) to study the way in which factors, such as length of assignment as instructor and other biographical data, are related to teaching proficiency; and (4) to measure the results of training experiments conducted to evaluate the effects of procedures, such as adding variety to the life of the flying instructor by temporarily rotating him to other assignments.

TRAINING EXPERIMENTS

Planning the best program of pilot training involves many problems which can be solved conclusively only by training experiments. The more scientifically these experiments are designed, the easier it is to get accurate, decisive results.

While the essential idea behind the experimental approach is trying something out to see how it works, a truly scientific experiment involves far more than this. Its great power lies in the fact that conditions are arranged so that the effects of disturbing factors are controlled or cancelled out and the results are decisive and unambiguous. It is also necessary to devise some way to measure these results so that they can be stated clearly and interpreted in the same way by everyone.

It is believed that one of the more useful functions of psychologically trained personnel working with the Army Air Forces should ultimately be to serve as consultants on the scientific design of experiments for evaluating training methods and devices. This was not possible during the war because of the concentration of psychological personnel on tasks more immediately related to the original classification testing program. Two minor experiments, however, and one semiexperimental study, were undertaken in the area of pilot training. These are briefly described to illustrate the experimental approach to training problems even though, because of their incidental nature, it was not possible to control conditions in such a way as to secure ideal experimental design.

One of these experiments indicated that students who had received part of their basic flying training on a two-engine airplane performed better in advanced two-engine training than those who had received all of their basic training on a single-engine airplane.

The other experiment compared the value of skeet shooting as a training aid under two conditions: when the conventional shotgun sight was used, and when a new optical sight was used. The optical sight had the advantage of simulating the sight pictures which the students would later employ in fixed gunnery missions. It was found that the students trained on a skeet gun equipped with an optical sight made reliably better scores in subsequent air-to-air firing than those who received their skeet training with the conventional type of shotgun.

The fact that the objective measures of advanced two-engine flying taken during a temporary training freeze happened to coincide with a change of airplanes in some schools enabled a semiexperimental comparison to be made. Because the different types of training compared were ones which happened to occur for other reasons instead of being deliberately designed for an experimental purpose, all of the conditions were not controlled and the results must be accepted with reservation. It was found that the students with a total of 10 hours of two-engine training, all of which was in the higher-powered TB–25, performed better on that airplane than students with a total of 15 hours of two-engine training, 10 of which was on the lower-powered AT–10 or UC–78.

The results of the three studies which have just been described deal with the problem of transfer of training. They all follow the well-known principle that more transfer occurs when the two situations are more similar.

The problem of transfer of training is involved throughout the pilot training program. Whenever a student is trained in one situation with the hope that it will hasten his learning in another, there are three possibilities: (1) The two situations may be so different that little or no transfer occurs; (2) the habits learned in the first situation may be sufficiently different from those acquired in the second so that they will interfere with learning and cause negative transfer; (3) the habits learned in the first situation may be useful in the second and produce positive transfer. Whether positive or negative transfer will occur may depend on the amount of training given in the first situation; one amount of training may be beneficial, another amount may be wasted or even harmful. While the results of experiments which have been performed in psychological laboratories on other types of learning cannot predict exactly what will happen in a complex situation like that found in flying training, they can suggest what to look for and how to design a series of training experiments for deciding what kinds and amounts of training at each level will be most efficient in the long run.

The factor of transfer of training is only one example of the variables which need investigation by the experimental method. Some other examples are determining the most efficient distribution of practice (or in other words, just how much material may be concentrated in a short period of time without defeating the purpose of training), and finding the optimum spacing of review to counteract forgetting. Training aids and equipment, such as training films and the Link Trainer, also should be evaluated by the experimental method. As the Training Command changes over to a peacetime status, one of its important functions should be research to assure this nation leadership in training methods and equipment.

CONDITIONS NECESSARY FOR EFFECTIVE RESEARCH

On the basis of their army and civilian research experience, the members of the Psychological Research Project (Pilot) have outlined the following conditions as essential to efficient research. Though many of these conditions are quite apparent, they are all listed because experience has shown that creative scientific work becomes inefficient whenever any one of them is not fulfilled.[**]

Personnel.—Scientific research requires specially trained, experienced personnel. It demands originality and intelligence in addition to approximately 3 years of postgraduate training in scientific principles and techniques after the completion of the ordinary 4-year college course. After this training, skill in planning and supervising research is acquired through experience. Furthermore, in working in an area such as pilot training which has its own complex technical aspects, the investigator needs a thorough understanding of the field to which he is applying his scientific techniques. It is highly desirable,

[**] See also the section on "Conditions of Research" in chapter 1 of this report.

therefore, that psychologists working on problems of pilot training receive systematic flying training, or that carefully selected pilots who are interested in training research be sent to universities for enough graduate training to receive the Ph. D. degree in psychology.

Since scientists receive stimulation from the discussions and criticisms of their colleagues, it is desirable to have a number of them working at the same place and to bring larger groups together for periodic conferences. In order to allow them to keep up with the latest technical developments in their field, attendance at scientific meetings and visits to various universities should be facilitated.

As increasing numbers of highly trained specialists are employed by the army, each type of specifically military training and duty should be examined to determine whether or not it is efficient to have the specialized personnel spend their time on it.

Since the Military Occupational Specialty number (MOS) plays such an important role in the administration of army personnel policies, research workers should have different MOS numbers adequately reflecting their specialization, even though the number of individuals involved may be relatively small in comparison with the rest of the army categories.

Chain of command.—Since research is such a highly technical specialty, it is just as necessary to have the chain of command for research filled by officers with rigorous training and experience in this specialty as it is to have a pilot as the Director of Training of a flying school. The technical direction of research, and also control over the assignment, transfer, and promotion of the personnel, should be exercised by men with the special training and experience necessary to understand the technical details of research operations.

Procurement.—Research frequently has sudden requirements for unusual supplies or services which cannot be procured by routine army procedures. For example, in designing and constructing an apparatus to accomplish something which has never been done before, the scientist may find that certain parts do not work satisfactorily so that they have to be unexpectedly redesigned. He cannot anticipate his exact needs months in advance in the same way that a mess officer can count on a special dinner for Thanksgiving Day. When he is making arrangements for a firm to develop something completely new, he cannot always submit complete plans in advance to three bidders. He needs special procurement authorities which are far more flexible than those required for routine army operations. In time of war, he needs high priorities in order quickly to procure special equipment. Expediting the procurement of a relatively small amount of crucial research equipment may pay big dividends in the long run.

Authority to establish experimental conditions.—In order to determine the most efficient method of training, it is frequently necessary to introduce controlled changes into training or to be able to procure subjects to serve in special experiments. The investigator must be able to secure authority to establish experimental groups. Although setting up a scientific experiment with the proper controls may sometimes be a temporary inconvenience, it is much more efficient in the long run than just trying something out to see what happens.

Liaison.—A research program needs representatives at all echelons. Representation at the highest level is especially important to keep the program informed of problems that are relevant in the light of the planning which is being conducted there, and to have its results presented where they can be most useful in determining matters of general policy. Most of the actual research, however, must be conducted in direct contact with the practical situation at the lower operational levels. In order to secure the rapid exchange of information which is frequently essential to efficient research, direct communication on technical matters is needed between representatives at different levels and in different commands. Since many types of information can only be secured first-hand, arrangements must be made to facilitate the travel of research personnel to the sources of this information.

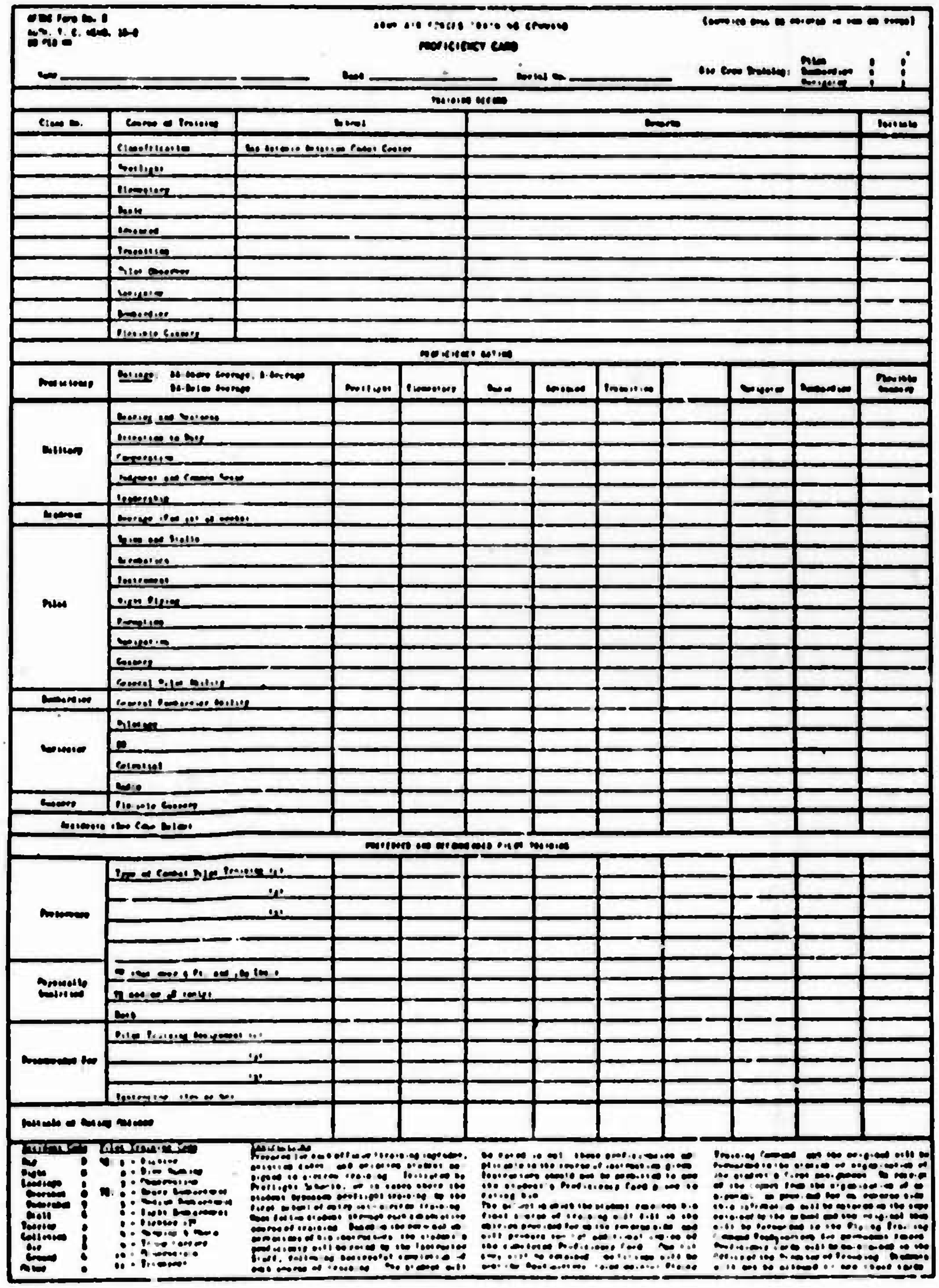
Relationship to operations.—The mission of a research organization should be to gather facts and to develop and evaluate new procedures. It should present recommendations and evidence to those responsible for determining policies and procedures but should not have any direct command responsibility for operations. It may be desirable for research personnel to assist in training operational personnel in new procedures and to monitor these new procedures for a period of time, but they should not be directly responsible for the operations. A separation of function is desirable in order to help the research personnel to preserve their independence and objectivity, and to allow each type of personnel (research and operating) to perform the functions for which they have special training and experience.

Another reason for keeping the two functions administratively separate (although the two types of personnel should work closely together) is to prevent the natural tendency for operations to swallow research. If the two are together, there is a natural tendency to meet various emergencies by cutting back research in order to maintain operations, since the effects of reducing operations are so immediately apparent while the more serious consequences of reducing research (like those of grinding up seed corn to make bread) do not show up till later.

A Research Corps.—It is believed that the special conditions essential to efficient research can be achieved most economically by bringing all different types of research personnel together in a Research Corps. This should result in economy of administration, making it easier to establish a chain of command with specially trained research personnel at all echelons. It should also effect a desirable coordination of the work of different specialists. For example, in developing a new piece of equipment, the question of whether its design is well adapted to human perceptual and motor abilities should be investigated at the same time that the engineering aspects of its development are being considered. Furthermore, those who will have to work on the special problems of selecting and training the personnel to operate it should be kept informed on the significant details of its development. It is believed that psychological research on pilot training can operate most efficiently if it is a part of a Research Corps.

APPENDIX TO CHAPTER FIVE

Studies of Subjective Measures of Flying Proficiency



(Last name) (First name) (Middle name) (Army serial number)

ARMY AIR FORCES TRAINING COMMAND
TRANSMITTAL OF PROFICIENCY CARD

SUBJECT: (Officer's Rank, Name and Serial Number)

TO: (Date)

1. The information shown on the opposite side of this form represents the training record of the subject officer, who has been assigned to your organization, together with an appraisal of his proficiency based on the personal observations of his instructors.

2. It is requested that, at such time as you have had sufficient opportunity to observe this officer's work, you fill in the information provided for below, together with any comments, criticisms or recommendations you may wish to make, and return this card.

REPORT ON SUBJECT OFFICER

TO: (Date)

1. Duty Assigned:
2. Military Discipline and Deportment:
3. Manner of Performing Assigned Duties:
4. General Officer Rating: ( ) Above Average ( ) Average ( ) Below Average
5. Type of Aircraft Flown (if pilot):
6. Accidents
7. Remarks

(Organization)

(Station)

(Signature of Reporting Officer)

Table A5.1.—*Correlation of categorized comments from the daily grade slip with primary pass-fail*

Class 44–K, Jones Field (N_g=59, N_e=33, P_g=.641)

Category [1]	M_e	M_g	$S. D._t$	r_{bis}
Motor technique [2]	0.15	1.10	0.62	0.87
Headwork	.39	1.65	.93	.84
Perception	.10	.61	.44	.72
Emotional difficulties	.42	1.21	.84	.58

[1] For this preliminary sample comments were grouped first into 81 subcategories. Since only those subcategories were used here which were positively related to pass-fail (i. e., eliminee group received more comments than graduates) the absolute size of these correlations is spuriously high, but the relative comparisons among categories are probably meaningful.

[2] Correlation coefficients have been changed in sign to indicate positive association of good performance.

Table A5.2.—*Intercorrelations of categorized comments from the daily grade slip*

Based on 92 Students in Class 44–K, Jones Field

Category	M	$S. D._t$	1	2	3	4
1. Motor technique [1]	0.59	0.62	....	0.48	0.56	0.64
2. Perception	.40	.44	0.48	....	.47	.36
3. Headwork	1.03	.93	.56	.47	....	.45
4. Emotional difficulties	.90	.84	.64	.36	.45	....

[1] Correlation coefficients have been changed in sign to indicate association of good performance positively.

TABLE A5.3.—*Intercorrelations of categories of grade slip comments, classification tests, pilot stanine and primary check-ride grade*

Based on 517 primary students in classes 44–H and 44–K. Correlations with tests and stanine are corrected for restriction in range to an assumed unrestricted standard deviation of 1.87 for the pilot stanine

Variables [1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1 Motor technique—graduate and eliminees		33	53	34	33	28	49	51	19	44	30	03	06	06	08	03	09	06	13	06	10	16	15	17	16	18	09	19
2 Motor technique—graduates	33		32	21	22	13	30	25	15	22	07	−02	−01	00	−02	04	00	−05	−03	−07	−07	−02	03	−01	02	07	01	−03
3 Motor technique—eliminees	53	32		34	40	25	45	48	31	44	23	08	08	08	14	10	10	00	05	03	08	12	18	16	10	07	15	16
4 Perception—graduates and eliminees	34	21	34		23	14	33	43	28	24	16	−04	05	01	01	−02	01	−01	11	−04	07	09	03	04	−08	11	−02	08
5 Perception—eliminees	33	22	40	23		34	20	41	38	25	28	04	08	08	11	00	11	00	05	07	03	12	09	13	06	06	10	12
6 Motivation—eliminees	28	13	25	14	34		19	43	46	23	37	05	18	01	04	−08	−04	−07	16	08	02	06	05	11	−01	06	−03	18
7 Headwork—graduates and eliminees	49	30	45	33	20	19		46	20	37	24	05	−01	09	06	14	15	06	15	14	12	16	21	15	12	18	11	22
8 Headwork—eliminees	51	25	48	43	41	43	46		41	42	33	03	06	05	09	06	13	02	18	15	15	21	14	19	16	12	16	26
9 Emotional difficulties—eliminees	19	15	31	28	38	46	20	41		28	26	02	12	02	08	06	08	04	17	16	08	13	10	11	04	08	06	21
10 Rough, abrupt, or tense controls	44	22	44	24	25	23	37	42	28		33	03	11	12	05	04	07	00	06	07	10	18	18	09	09	11	08	20
11 Primary check ride	30	07	23	16	28	37	24	33	26	33		00	14	16	14	13	10	05	16	16	18	25	21	29	17	08	16	31
12 Biographical data—navigator	03	−02	08	−04	04	05	05	03	02	03	00		17	14	05	16	21	−02	−11	06	02	09	09	03	−04	02	14	06
13 Biographical data—pilot	06	−01	08	05	08	18	−01	06	12	11	14	17		10	09	−05	−07	05	11	01	04	12	05	16	04	12	−02	40
14 Spatial orientation I	06	00	08	01	08	01	09	05	02	12	16	14	10		52	14	42	14	13	30	25	17	40	14	07	15	13	43
15 Spatial orientation II	08	−02	14	01	11	04	06	09	08	05	14	05	09	52		22	31	08	16	24	25	33	39	22	14	11	24	45
16 Reading comprehension	03	04	10	−02	00	−08	14	06	06	04	13	16	−05	14	22		33	05	05	23	16	38	27	46	03	01	54	30
17 Dial and table reading	09	00	10	01	11	−04	15	13	08	07	10	21	−07	42	31	33		14	10	31	28	15	43	13	05	04	42	36
18 Finger dexterity	06	−05	00	−01	00	−07	06	02	04	00	05	−02	05	14	08	05	14		28	17	27	11	13	06	08	26	02	25
19 Two-hand coordination	13	−03	05	11	05	16	15	18	17	06	16	−11	11	13	16	05	10	28		25	40	41	19	18	16	38	07	53
20 Discrimination reaction time	06	−07	03	−04	07	08	14	15	16	07	16	06	01	30	24	23	31	17	25		31	19	30	11	13	11	20	42
21 Complex coordination	10	−07	08	07	03	02	12	15	08	10	18	02	04	25	25	16	28	27	40	31		24	32	12	15	32	12	62
22 Mechanical principles	16	−02	12	09	12	06	16	21	13	18	25	09	12	[illegible]	33	38	15	11	41	19	24		30	45	15	24	38	50
23 Instrument comprehension II	15	03	18	03	09	05	21	14	10	18	21	09	05	40	39	27	43	13	19	30	32	30		27	09	21	30	54
24 General information	17	−01	16	04	13	11	15	19	11	09	29	03	16	14	22	46	13	06	18	11	12	45	27		09	13	36	50
25 Rudder control	16	02	10	−08	06	−01	12	16	04	09	17	−04	04	07	14	03	05	06	16	13	15	18	09	09		09	04	30
26 Rotary pursuit	18	07	07	11	06	06	18	12	08	11	08	02	12	15	11	01	04	26	38	11	22	24	21	13	09		−01	44
27 Math B	09	01	15	−02	10	−03	11	16	06	08	16	14	−02	13	24	54	42	02	07	20	12	38	30	36	04	−01		30
28 Pilot stanine	19	−03	16	08	12	18	22	26	21	20	31	06	40	43	45	30	36	25	53	42	62	50	54	50	30	44	30	

[1] Variables 1–9 are the same as variables A to I in Table 5.11 on page 102.

Table A5.4.—*Intercorrelations of items on rating scale for aviation cadets (Form A-Pilot) grouped by cluster patterns*

Based on 369 students in class 42-K rated between the 8th and 10th hours of primary training

Item	1	2	10	3	15	4	8	5	6	7	13	9	11	12	16	17	18	19	20	14
Cluster I—Headwork:																				
1 Judgment		80	74	70	69	66	67	66	66	63	64	56	55	56	60	69	63	63	40	54
2 Foresight and planning	80		76	70	71	53	70	68	67	64	62	55	51	57	58	69	65	61	38	52
10 Speed of decision and reaction	74	76		64	67	64	67	66	67	63	62	56	58	63	54	66	62	59	41	51
3 Memory	70	70	64		68	74	67	64	60	55	57	48	54	60	53	61	57	55	45	56
15 Progress in developing technique	69	71	67	68		66	71	68	64	64	67	55	61	66	61	64	61	56	47	65
4 Comprehension	66	53	64	74	66		62	61	58	58	53	56	53	60	51	56	60	58	43	53
8 Division of attention	67	70	67	67	71	62		69	65	63	64	56	58	62	64	63	59	56	39	55
5 Visualization of flight course	66	68	66	64	68	61	69		69	64	62	52	61	64	56	60	55	53	36	58
6 Estimation of speed and distance	66	67	67	60	64	58	65	69		69	69	55	59	62	56	62	58	50	38	52
Cluster II—Perception:																				
7 Sense of sustentation	63	64	63	55	64	58	63	64	69		76	54	63	63	57	55	56	47	32	56
13 Feel of the controls	64	62	62	57	67	53	64	62	69	76		50	68	68	56	53	52	45	33	64
Cluster III—Orientation:																				
9 Orientation	56	55	56	48	55	56	56	52	55	54	50		47	47	46	56	52	45	31	42
Cluster IV—Coordination:																				
11 Coordination	55	51	58	54	61	53	58	61	59	63	68	47		76	52	50	42	42	34	61
12 Appropriateness of controls used	56	57	63	60	66	60	62	64	62	63	68	47	76		49	52	52	47	33	58
Cluster V—Emotion:																				
16 Absence of tenseness	60	58	54	53	61	51	64	56	56	57	56	46	52	49		75	63	47	29	54
17 Absence of confusion and nervousness	69	60	66	61	64	56	63	60	62	55	53	56	50	52	75		76	58	37	48
18 Absence of fear and apprehension	63	65	62	57	61	60	59	55	58	56	52	52	42	52	63	76		61	48	42
Cluster VI—Attitude:																				
19 Suitable temperament	63	61	59	55	56	58	56	53	50	47	45	45	42	47	47	58	61		61	45
20 Motivation and attitudes	40	38	41	45	47	43	39	36	38	32	33	31	34	33	29	37	48	61		32
Omit:																				
14 Smoothness of control movement	54	52	51	56	65	53	55	58	52	56	64	42	61	58	54	48	42	45	32	

APPENDIX TO CHAPTER SIX

Objective Measures of Flying Skill For the Primary Level of Pilot Training

For the item-analysis studies, I, II, III, IV, and V, each of the tables of results (tables A6.1 through A6.13) is divided into sections A, B, and C. Section A shows the summary data and contains the biserial correlations of the measure scores with graduation elimination from primary school; the point-biserial correlation with scores of the upper group (more hours of flying training) and the lower group (fewer hours of flying training) on the first day of testing; similar point-biserial correlations for the same groups on the second day of testing; and the test-retest reliability, which is a combination of test-retest reliabilities for the upper and lower groups by the z-transformation technique.

Section B shows the basic data for the biserial correlations with graduation and elimination from Primary. Section C shows the basic data for the point-biserial correlations on the two days of testing and for the test-retest reliabilities of upper and lower groups.

Beginning with table A6.14, the necessary data for studies VI and VII are shown.

In all of the tables the sign of the coefficients has been adjusted so that goodness of performance in both variables is indicated by a positive sign.

The asterisk on the coefficients in the summary data tables indicate the chances in 100 of obtaining positive coefficients as large as these, if there were no relationship between the variables. This is done according to the following system:

No asterisk = More than 10 percent.
* = 10 percent.
** = 2.5 percent.
*** = 0.5 percent.

The footnotes for all of the tables are as follows:

[b] Lower score is better score.
[c] Corrected for broad categories used in the large-scale study of the effects of additional training on flying skill, the complete report of which is in chapter 10.
[d] Tetrachoric correlation.
[e] Coded score.

TABLE A6.1—*Take-off, climb, and first turn*

Section A: Summary data

Code (Study, maneuver and measure)	Validity			Reliability
	1	2	3	4
	r_{bis} grad.—elim.	r_{pbis} U. versus L. first day	r_{pbis} U. versus L. second day	r_{11} first versus second day
VI-A1. Smooth application of throttle in take-off	***0.55			***0.45
III-K2. Ground control during 45 m. p. h. tail high run: kept line under left wing until take-off	***.31	0.12	−0.11	***.27
III-M2. Ground control during acceleration to 55 m. p. h. and deceleration to full stop: kept line under wing throughout	−.17	.02	.00	*.16
VI-A2. Deviation of plane from straight path during take-off roll	***.48			***.66
II-A1. A/S when plane leaves ground; corrected for instrument error (instruments covered)		.12	−.09	.13
III-A1. A/S when plane leaves ground; corrected for instrument error	.01	***.30	*.14	**.22
IV-A3. A/S reading when plane leaves ground	.05			***.68
VI-A4. Instructor's rating of plane fly off	***.75			***.52
I-A2. A/S range in climb to first turn	*.18	.01	−.08	.12
II-A2. A/S range in climb to first turn (instruments covered)		***.31	.04	.07
III-A2. A/S range in climb to first turn	−.29	.03	−.02	−.01
VI-A5. A/S range in climb to first turn	***.66			***.50
VI-A6. Plane drift during climb	***.48			***.58
II-A3. Altitude range during first turn (instruments covered)		**.27	.01	.09
III-A3. Altitude range during first turn	*.20	*.17	***.33	*.14
VI-A7. Altitude deviation from 300 feet during first turn	***.77			***.66
III-A4. Ball-bank deviation during first turn	.05	*.14	.03	−.12

Section B: Data for r_{bis} graduation-elimination

Code	N_t	p_g	M_g	M_e	SD_t	r_{bis}
VI-A1	73	73	1.74	1.44	0.33	0.55
IIIK2	42	69	1.28	.92	.69	.31
IIIM2	42	69	2.14	2.54	1.38	−.17
VI-A2	73	73	2.52	1.91	.75	.48
IIIA1	42	69	[b] 11.03	11.15	5.73	.01
VIA3	71	73	2.97	2.92	.58	.05
VIA4	73	73	1.61	1.17	.35	.75
IA2	35	66	[b] 11.96	13.33	4.69	.18
IIIA2	42	69	[b] 17.59	14.23	7.20	−.29
VIA5	73	73	5.10	4.27	.75	.66
VIA6	73	73	2.09	1.51	.71	.48
IIIA3	42	69	[b] 58.62	72.21	41.02	.20
VIA7	71	73	6.22	4.48	1.35	.77
IIIA4	42	69	5.31	5.23	.85	.05

[b] Lower score is better score.

TABLE A6.1—*Take-off, climb, and first turn*—Continued

Section C: Data for r_{pbis} and for reliability r_{11}

Code	Upper or lower	First day r_{pbis}	First day N	First day M_s	First day SD	Second day r_{pbis}	Second day N	Second day M_s	Second day SD	r_{11}	N	Reliability r_{11} test—retest s-comb.
IIIK2	Upper	0.12	46	0.59	0.49	−0.11	46	0.59	0.49	0.44	46	0.27
	Lower		45	.47	.50		45	.60	.46	.09	45	
IIIM2	Upper	.02	46	1.07	.82	.00	46	1.13	.90	.11	46	.16
	Lower		45	1.04	.92		45	1.13	.91	.21	45	
IIA1	Upper	.12	46	[b] 4.24	4.17	−.09	46	[b] 5.87	4.92	.06	46	.13
	Lower		43	5.23	4.31		43	5.12	3.81	.21	43	
IIIA1	Upper	.30	46	[b] 3.80	3.79	.14	46	[b] 3.70	3.96	.40	46	.22
	Lower		45	6.22	3.97		45	4.78	3.79	.02	45	
IA2	Upper	.01	41	[b] 6.22	3.95	−.08	41	[b] 6.71	2.84	.25	41	.12
	Lower		38	6.32	3.58		38	6.18	3.33	−.02	38	
IIA2	Upper	.31	46	[b] 5.22	3.12	.01	46	[b] 6.20	3.93	.16	46	.07
	Lower		42	7.74	4.53		42	6.43	2.51	−.03	42	
IIIA2	Upper	.03	46	[b] 8.15	4.82	−.02	46	[b] 8.26	4.33	−.15	46	−.01
	Lower		45	8.44	4.57		45	8.11	4.75	.15	45	
IIA3	Upper	.27	44	[b] 20.9	22.3	.01	44	[b] 21.8	20.4	.12	44	.09
	Lower		42	36.0	30.9		42	22.4	23.2	.07	42	
IIIA3	Upper	.17	46	[b] 23.91	22.70	.33	46	[b] 15.65	14.98	−.03	46	.14
	Lower		45	32.45	25.83		45	29.33	23.70	.31	45	
IIIA4	Upper	.14	46	2.80	.45	.03	46	2.76	.82	−.01	46	−.12
	Lower		44	2.64	.71		44	2.73	.49	−.22	44	

[b] Lower score is better score.

TABLE A6.2.—*Low-altitude maneuvers: (S-turns: Rectangular course; altitude, A/S, RPM estimation)*

Section A: Summary data

Code (Study, maneuver, and measure)	Validity 1 r_{bis} grad.—elim.	Validity 2 r_{pbis} U. versus L. first day	Validity 3 r_{pbis} U. versus L. second day	Reliability 4 r_{11} first versus second day
III-G1. S-turns: Altitude range in 4 loops	*0.34	**0.22	**0.25	***0.42
III-G2. S-turns: Drift corrections; sum of time for 2 downwind loops (1-3), minus sum of time for 2 upwind loops (2+4)	***.41	−.09	.02	***.29
III-G3. S-turns: Bank changed through level over road at end of first loop	*.28	.09	.11	.00
III-G4. S-turns: Bank changed through level over road at end of second loop	.15	−.04	.02	−.09
III-G5. S-turns: Bank changed through level over road at end of third loop	.09	.06	*.16	.01
III-G6. S-turns: Bank changed through level over road at end of fourth loop	.11	.03	*.16	−.13
III-H1. Rectangular course: Altitude range on 4 legs of rectangular course	*.35	***.31	*.14	.10
III-H2. Rectangular course: Drift correction; turn onto first leg correct or proper correction made	....	−.19	.12	−.08
III-H3. Rectangular course: Drift correction; turn onto second leg correct or proper correction made	....	−.03	*.14	*.18
III-H4. Rectangular course: Drift correction; turn onto third leg correct or proper correction made	....	.04	.01	.07
III-H5. Rectangular course: Drift correction; turn onto fourth leg correct or proper correction made (r_{bis} for sum of all 4 drift correction measures)	.16	−.09	***.28	***.30
II-F1. Instruments covered: Placed plane at 500-foot altitude	....	.07	.06	*.17
II-F2. Instruments covered: Set plane at cruising A/S; deviation from cruising A/S	....	.06	−.07	−.06
II-F3. Instruments covered: Set plane at cruising r. p. m.; deviation from r. p. m.	....	.04	−.04	*.15
II-F4. Instruments covered: Altitude range during 30 seconds straight and level at 500-foot altitude	....	−.07	.10	.05

TABLE A6.2.—*Low-altitude maneuvers: (S-turns: Rectangular course; altitude, A/S, RPM estimation*—Continued

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_t	M_s	M_e	SD_t	r_{bis}
IIIG1	41	68	[b] 206.43	298.46	161.96	0.34
IIIG2	42	69	[b] 14.07	19.92	8.85	.41
IIIG3	41	71	1.38	1.17	.70	.28
IIIG4	41	71	1.31	1.25	.66	.15
IIIG5	38	71	1.37	1.27	.62	.09
IIIG6	37	70	1.46	1.36	.60	.11
IIIH1	38	66	[b] 245.60	341.54	167.23	.35
IIIH2, 3, 4, 5	41	71	12.21	11.00	4.62	.16

[b] Lower score is better score.

Section C: Data for r_{pbis} and for reliability r_{11}

Code	Upper or lower	First day r_{pbis}	First day N	First day M_a	First day SD	Second day r_{pbis}	Second day N	Second day M_a	Second day SD	r_{11}	Reliability test-retest z-corab.
IIIG1	Upper	0.22	40	[b] 85.00	39.43	0.25	40	[b] 84.50	36.74	−0.06	0.42
	Lower		44	113.41	80.60		44	119.77	88.89	.72	
IIIG2	Upper	.09	46	7.48	4.38	.02	46	7.60	5.34	.32	.29
	Lower		45	8.33	5.42		45	7.60	5.77	.25	
IIIG3	Upper	.08	46	.67	.47	.11	46	.78	.41	[c] −.04	[c] .00
	Lower		44	.59	.49		44	.68	.47	[c] .05	
IIIG4	Upper	−.04	45	.73	.44	.02	45	.56	.50	[c] .06	[c] −.09
	Lower		43	.77	.42		43	.53	.50	[c] −.12	
IIIG5	Upper	.06	46	.74	.44	.16	46	.78	.41	[c] .30	[c] .01
	Lower		41	.68	.47		41	.63	.48	[c] −.32	
IIIG6	Upper	.03	42	.83	.37	.16	42	.79	.41	[c] .15	[c] −.13
	Lower		39	.82	.64		39	.64	.48	[c] −.42	
IIIH1	Upper	.31	43	[b] 91.63	60.34	.14	43	[b] 106.98	53.33	−.07	.10
	Lower		41	149.76	114.03		41	127.80	87.89	.27	
IIIH2	Upper	−.19	46	1.11	.91	.12	46	1.67	.98	−.11	−.08
	Lower		45	1.44	.78		45	1.47	.78	−.05	
IIIH3	Upper	−.03	46	1.22	1.00	.14	46	1.67	.98	.18	.18
	Lower		44	1.27	.90		44	1.41	.86	.17	
IIIH4	Upper	.04	46	1.46	1.08	.01	46	1.87	1.01	.13	.07
	Lower		44	1.36	1.15		44	1.61	1.00	.00	
IIIH5	Upper	−.09	46	1.30	.93	.28	46	2.13	.92	.36	.30
	Lower		44	1.50	1.16		44	1.59	.96	.22	
IIF1	Upper	.07	46	[b] 13.39	14.88	.06	46	[b] 10.48	9.97	.28	.17
	Lower		43	15.61	16.11		43	11.74	11.41	−.05	
IIF2	Upper	.06	46	[b] 3.91	3.60	−.07	46	[b] 4.78	4.54	.00	−.05
	Lower		43	4.42	4.97		43	4.19	3.40	−.20	
IIF3	Upper	.04	46	[b] 5.74	4.88	−.04	46	[b] 6.09	5.33	.11	.15
	Lower		38	6.16	5.35		38	5.66	4.38	.19	
IIF4	Upper	−.07	46	[b] 5.91	7.15	.10	46	4.26	3.46	.04	.05
	Lower		43	5.14	3.99		43	5.05	4.06	.07	

[b] Lower score is better score.
[c] Tetrachoric correlation.

Table A6.3.—*Accuracy landing: traffic pattern and approach*

Section A: Summary data

Code (Study, maneuver and measure)	Validity			Reliability
	1	2	3	4
	r_{bis} grad.—elim.	r_{pbis} U. versus L. first day	r_{pbis} U. versus L. second day	r_{tt} first versus second day
VI-B1. Traffic pattern: Enters downwind leg in middle third	***0.43	....	....	***0.45
I-G1. Traffic pattern: Altitude range from pattern entry until throttle is cut	.15	***0.42	***0.32	.02
II-M1. Traffic pattern: Altitude range from pattern entry until throttle is cut	....	***.29	.07	.05
III-N1. Traffic pattern: Altitude range from pattern entry until throttle is cut	.20	***.33	.05	−.10
VI-B2. Traffic pattern: Maximum deviation from altitude of 500 feet in pattern from entry until throttle is cut	***.90	....	....	***.54
VI-B4. Traffic pattern: Turn onto approach leg started 300 feet or above	***.67	....	....	***.22
VI-B5. Final approach: After throttle closed, more power not necessary, or added as required	**.33	....	....	***.26
I-G2. Final approach: A/S range from completion of last turn until student begins to break glide	*.28	.01	.10	***.31
II-M2. Final approach: A/S range from beginning of last turn until student begins to break glide	....	*.20	.02	.03
II-N1. Final approach: A/S range from beginning of last turn until student begins to break glide	....	.10	*.20	−.05
III-N2. Final approach: A/S range from beginning of last turn until student begins to break glide	.18	.08	.03	.10
VI-B3. Final approach: A/S range from beginning of last turn until student begins to break glide	***.35	....	....	***.77
II-M3. Final approach: Changed approach course or S'ed to hit correct landing area	....	.05	**.26	***[a].28
II-N2. Final approach: Changed approach course or S'ed to hit correct landing area	....	.20	***.32	[a].05
III-N3. Final approach: Changed approach course or S'ed to hit correct landing area	.25	***.33	**.23	[a]−.05
VI-B6. Final approach: Maintain glide track parallel with "T"	***.39	....	....	***.36

[a] Tetrachoric correlation.

Section B: Data for r_{bis} graduation-elimination

Code	N_t	p_s	M_s	M_e	SD_t	r_{bis}
VIB1	70	74	1.80	1.57	0.31	0.43
IG1	35	66	[b] 3.91	4.26	1.38	.15
IIIN1	41	68	[b] 135.71	155.38	69.89	.20
VIB2	71	73	6.20	4.10	1.39	.90
VIB4	73	73	1.96	1.76	.18	.67
VIB5	73	73	2.45	2.16	.50	.33
IG2	35	66	[b] 3.66	4.37	1.66	.28
IIIN2	39	67	[b] 17.12	18.85	5.76	.18
VIB3	73	73	5.12	4.41	.78	.35
IIIN3	38	71	.96	.64	.77	.25
VIB6	73	73	1.60	1.40	.32	.39

[b] Lower score is better score.

Table A6.3.—*Accuracy landing: traffic pattern and approach*—Continued

Section C: Data for r_{pbis} and for Reliability r_{11}

Code	Upper or lower	First day r_{pbis}	First day N	First day M	First day SD	Second day r_{pbis}	Second day N	Second day M	Second day SD	r_{11}	Reliability test-retest z-comb.
IO1	Upper	0.42	41	ᵇ44.02	24.77	0.32	41	ᵇ49.88	22.70	0.23	0.02
	Lower		38	78.03	46.02		38	71.32	39.08	−.20	
IIM1	Upper	.29	46	ᵇ61.30	34.17	.07	46	ᵇ53.70	30.60	.06	.05
	Lower		43	87.67	50.87		43	58.37	41.03	.03	
IIIN1	Upper	.33	46	ᵇ55.70	34.40	.05	46	ᵇ56.50	36.90	−.01	−.10
	Lower		43	83.30	45.20		43	60.90	49.20	−.19	
IO2	Upper	.01	41	ᵇ7.44	4.15	.10	41	ᵇ6.46	4.02	.33	.31
	Lower		38	8.29	4.02		38	7.24	3.92	.28	
IIM2	Upper	.20	42	ᵇ6.79	4.34	.02	42	ᵇ7.50	2.94	−.07	.03
	Lower		41	8.54	4.17		41	7.68	4.95	.13	
IIN1	Upper	.10	43	ᵇ7.33	3.79	.20	43	ᵇ6.51	3.33	.04	−.05
	Lower		42	8.10	4.08		42	8.10	4.36	−.14	
IIIN2	Upper	.08	45	ᵇ7.89	4.15	.03	45	ᵇ8.67	3.86	.24	.10
	Lower		42	8.67	3.98		42	8.93	4.44	−.05	
IIM3	Upper	.05	41	.66	.47	.26	41	.83	.38	.40	.28
	Lower		39	.62	.49		39	.59	.49	.14	
IIN2	Upper	.20	43	.79	.41	.32	43	.88	.32	−.02	.05
	Lower		41	.61	.49		41	.61	.49	.13	
IIIN3	Upper	.33	42	.74	.44	.23	42	.67	.45	−.35	−.06
	Lower		41	.41	.49		41	.44	.50	.24	

ᵇ Lower score is better score.

Table A6.4.—*Accuracy landing: Place and manner of ground contact*

Section A: Summary data

Code (Study, maneuver and measure)	Validity 1: r_{bis} grad.—elim.	Validity 2: r_{pbis} U. versus L. first day	Validity 3: r_{pbis} U. versus L. second day	Reliability 4: r_{11} first versus second day
VI-B7. Ground contact: Landed in crab	*0.29			***0.42
VI-B8. Ground contact: Landed drifting	***.44			***.52
I-G6. Place of contact: Goal, first ⅓ of field	−.19	***0.32	***0.32	−.10
II-M5. Place of contact: Goal, middle of zone 200 feet wide, 600 feet long		*.19	*.24	−.06
II-N4. Place of contact: Goal, middle of zone 200 feet wide, 600 feet long		.10	.16	*.20
III-N4. Place of contact: Goal, middle of zone 200 feet wide, 600 feet long	***.42	*.20	.13	.07
VII-E1. Place of contact: Goal, middle of zone 200 feet wide, 600 feet long		.02		*.02
VII-F1. Place of contact: Goal, middle of zone 200 feet wide, 600 feet long		.03		*.01
VI-B11. Place of contact: Goal, first ⅓ of field	−.23			***.37
II-N5. Place of contact on second successive landing: Corrected for previous over- or under-shooting		−.01	−.05	.00
II-M6. Landing attitude: Smooth 3-point, tail first, wheels first, 3-point but dropped in 3 feet or more		**.26	.09	.03
II-N6. Landing attitude: Smooth 3-point, tail first, wheels first, 3-point but dropped in 3 feet or more		*.17	.10	−.01
III-N5. Landing attitude: Smooth 3-point, tail first, wheels first, 3-point but dropped in 3 feet or more	***.56	**.23	.03	.12
VII-E2. Landing attitude: 3-point or slightly tail first, wheels first with tail below horizontal, tail horizontal or higher		.02		*−.09
VII-F2. Landing attitude: 3-point or slightly tail first, wheels first with tail below horizontal, tail horizontal or higher		.04		***.20
VI-B9. Landing attitude: Smooth 3-point, tail first, wheels first, 3-point but dropped in 3 feet or more	***.32			***.43
I-G5. Plane bounced	−.13	***.34	*.21	.00
II-M7. Plane bounced		.11	−.06	−.12
II-N7. Plane bounced		−.05	−.10	−.03
III-N6. Plane bounced	***.69	**.24	.01	−.09
VII-E3. Plane bounced or dropped in		.00		*−.01
VII-F3. Plane bounced or dropped in		.01		*−.01
VI-B10. Plane bounced	***.60			***.46

Table A6.4—*Accuracy landing: Place and manner of contact*—Continued

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_t	M_t	M_e	SD_t	r_{bis}
VIB7	72	72	1.83	1.69	0.28	0.29
VIB8	71	72	1.69	1.45	.32	.44
IO6	35	66	[b] 6.11	5.69	1.34	−.19
IIIN4	37	65	1.08	.31	1.14	.42
VIB11	73	73	2.73	2.85	.31	−.23
IIIN5	35	66	3.48	2.08	1.53	.56
VIB9	73	73	2.57	2.21	.67	.32
IO5	35	66	[b] 4.33	4.08	1.27	−.13
IIIN6	31	71	2.82	1.78	.91	.69
VIB10	73	73	2.32	1.78	.53	.60

[b] Lower score is better score.

Section C: Data for r_{pbis} and for reliability r_{11}

Code	Upper or lower	First day r_{pbis}	First day N	First day M	First day SD	Second day r_{pbis}	Second day N	Second day M	Second day SD	r_{11}	N	Reliability test-retest z-comb.
IO6	Upper	0.32	41	2.73	0.44	0.32	41	2.83	0.38	0.02	41	−0.10
	Lower		29	2.38	.61		29	2.52	.56	−.27	29	
IIM5	Upper	.19	24	.83	.80	.24	24	.92	.76	−.09	24	−.06
	Lower		25	.52	.81		25	.56	.70	−.02	25	
IIN4	Upper	.10	28	.82	.76	.15	28	1.25	.69	.15	28	.20
	Lower		30	.67	.79		30	1.03	.73	.21	30	
IIIN4	Upper	.20	45	.53	.75	.13	45	.47	.75	−.05	45	.07
	Lower		42	.26	.56		42	.29	.59	.20	42	
VIIE1	85 hrs	.02	954	1.68	.76					(r_{11} N=170)		•.01
	60 hrs		211	1.64	.77							
VIIF1	85 hrs	.03	954	1.85	.79					(r_{11} N=170)		•.01
	60 hrs		211	1.78	.77							
IIN5	Upper	−.01	22	.55	.50	−.05	22	.45	.50	.44	22	.00
	Lower		22	.59	.50		22	.50	.50	−.44	22	
IIM6	Upper	.26	31	1.61	1.01	.09	31	1.90	1.09	−.21	31	.03
	Lower		27	1.07	1.02		27	1.70	1.15	.31	27	
IIN6	Upper	.17	38	1.71	1.10	.10	38	2.11	1.07	−.06	38	.01
	Lower		36	1.31	1.22		36	1.89	1.17	.04	36	
IIIN5	Upper	.23	46	1.52	1.35	.03	46	1.33	1.30	.25	46	.12
	Lower		44	.91	1.20		44	1.25	1.28	−.02	44	
VIIE2	85 hrs	.02	954	2.44	.75					(r_{11} N=170)		•−.09
	60 hrs		210	2.40	.72							
VIIF2	85 hrs	.04	953	2.55	.68					(r_{11} N=170)		•.20
	60 hrs		211	2.47	.66							
IO5	Upper	.34	34	3.50	.75	.21	38	3.45	.71	.07	38	.00
	Lower		35	2.80	1.14		35	3.06	1.07	−.08	35	
IIIN7	Upper	.14	29	1.31	.70	−.00	29	1.38	.67	−.11	29	−.12
	Lower		26	1.12	.64		26	1.46	.63	−.13	26	
IIN7	Upper	−.05	33	1.33	.59	−.10	33	1.55	.66	−.16	33	−.03
	Lower		30	1.40	.66		30	1.67	.47	.11	30	
IIIN6	Upper	.21	45	1.16	.82	−.01	45	1.09	.87	.04	45	−.09
	Lower		43	.74	.87		45	1.07	.87	−.22	43	
VIIE3	85 hrs	.00	952	2.37	.80					(r_{11} N=170)		•−.01
	60 hrs		210	2.36	.79							
VIIF3	85 hrs	.01	954	2.44	.77					(r_{11} N=170)		•.01
	60 hrs		211	2.41	.77							

Table A6.5.—*Recovery landing after bounce or zoom*

Section A: Summary data

Code (Study, maneuver and measure)	Validity: 1 r_{bis} grad.—elim.	Validity: 2 r_{pbis} U. versus L. first day	Validity: 3 r_{pbis} U. versus L. second day	Reliability: 4 r_{11} first versus second day
III-J2. Recovery after bounce, student's landing attitude: Smooth 3-point, tail first, wheels first, 3-point but dropped in 3 feet or more	0.04	**0.23	0.02	0.03
III-J1. Recovery after bounce, student's landing bounced: Not at all, under 3 feet, more than 3 feet	*.30	*.17	.06	.10
III-L1. Recovery after zoom, student's landing attitude: Smooth 3-point, tail first, wheels first, 3-point but dropped in 3 feet or more	−.06	.13	.07	.04
III-L3. Recovery after zoom, student's landing bounced: Not at all, under 3 feet, more than 3 feet	−.05	.10	.06	.06

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_t	M_t	M_e	SD_t	r_{bis}
IIIJ2	41	71	2.48	2.33	2.07	0.04
IIIJ1	41	68	2.07	1.38	1.37	.30
IIIL1	41	69	1.93	2.15	1.83	−.06
IIIL3	41	69	1.72	1.85	1.36	−.05

Section C: Data for r_{pbis} and for reliability r_{11}

Code	Upper or lower	First day r_{pbis}	First day N	First day M_e	First day SD	Second day r_{pbis}	Second day N	Second day M_e	Second day SD	r_{11}	Reliability test-retest z-comb.
IIIJ2	Upper	0.23	46	1.70	1.20	0.02	46	1.60	1.19	−0.17	0.03
	Lower		44	1.11	1.30		44	1.45	1.34	.24	
IIIJ1	Upper	.17	46	1.24	.81	.06	46	1.15	.78	−.09	.10
	Lower		44	.95	.85		44	1.02	.89	.30	
IIIL1	Upper	.13	46	1.39	1.31	.07	46	1.15	1.25	−.06	.04
	Lower		45	1.04	1.32		45	.98	1.18	.14	
IIIL3	Upper	.10	46	1.07	.79	.06	46	1.04	.83	−.04	.06
	Lower		45	.89	.90		45	.93	.88	.16	

Table A6.6.—*Simulated forced landing*

Section A: Summary data

Code (Study, maneuver, and measure)	Validity: 1 r_{bis} grad.—elim.	Validity: 2 r_{pbis} U. versus L. first day	Validity: 3 r_{pbis} U. versus L. second day	Reliability: 4 r_{11} first versus second day
II-L3. A/S range in glide	....	***0.39	0.06	0.07
III-F3. A/S range in glide	***0.50	**.25	.05	.08
II-L2. Simulates base leg and 90° approach	....	*.15	.04	*−.09
III-F2. Simulates base leg and 90° approach	***.62	.04	.12	*.17
VII-D1. Simulates base leg and 90° approach	....	**.07	....	*.03
II-L4. Comes in upwind, crosswind, downwind	....	.04	.07	.07
III-F4. Comes in upwind, crosswind, downwind	***.48	.06	*.20	**.22
II-L1. Would have landed in first ⅓ of field	....	*.20	.13	*.19
III-F1. Would have landed in first ⅓ of field	.20	.11	−.08	.01
VII-D2. Would have landed in first ⅓ of field	....	*.04	....	*.07

TABLE A6.6.—*Simulated forced landing*—Continued

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_e	M_t	M_e	SD_t	r_{bis}
IIIF3	42	69	[b]21.03	30.00	9.12	0.59
IIIF2	40	72	1.41	.64	.75	.62
IIIF4	42	69	3.76	3.08	.93	.48
IIIF1	41	71	2.21	1.83	1.10	.20

[b] Lower score is better score.

Section C: Data for r_{pbis} and for reliability r_{tt}

Code	Upper or lower	First day r_{pbis}	First day N	First day M_e	First day SD	Second day r_{pbis}	Second day N	Second day M_e	Second day SD	r_{tt}	Reliability test-retest z-comb.
IIL3	Upper	0.39	45	[b]6.67	4.35	0.00	45	[b]7.33	3.89	−0.03	0.07
	Lower		42	10.71	5.18		42	7.86	4.77	.18	
IIIF3	Upper	.25	46	[b]10.22	5.51	.05	40	[b]10.00	5.21	−.13	−.08
	Lower		45	13.56	7.27		45	10.56	5.50	−.04	
IIL2	Upper	.15	46	.67	.40	.04	46	.85	.37	−.07	−.08
	Lower		38	.53	.42		38	.82	.66	−.09	
IIIF2	Upper	−.04	46	.52	.50	.12	46	.72	.45	.12	.17
	Lower		43	.56	.50		43	.60	.49	.22	
VIID1	85 hrs	.07	953	2.76	.65						*.03
	60 hrs		211	2.04	.77						
IIL4	Upper	.04	44	1.77	.56	.07	44	1.86	.46	.24	.07
	Lower		40	1.73	.59		40	1.80	.46	.11	
IIIF4	Upper	.06	44	1.80	.46	.20	44	1.93	.25	.08	.22
	Lower		45	1.73	.53		45	1.76	.60	.35	
IIL1	Upper	.20	45	1.27	.80	.13	45	1.38	.82	−.15	.19
	Lower		42	.91	.95		42	1.14	.92	.51	
IIIF1	Upper	.11	46	1.00	.91	.08	46	1.09	.80	.21	.01
	Lower		44	.80	.87		44	1.23	.90	−.20	
VIID2	85 hrs	.04	953	2.40	.92						*.07
	60 hrs		211	2.31	.95						

TABLE A6.7.—*Four 180° climbing turns 1. Left 2. Right 3. Left 4. Right*

Section A: Summary data

Code (Study, maneuver and measure)	Validity 1: r_{bis} grad.—elim.	Validity 2: r_{pbis} U. versus L. first day	Validity 3: r_{pbis} U. versus L. second day	Reliability 4: r_{tt} first versus second day
II-B1. R. p. m. in straight climb (instruments covered), deviation from correct		0.08	0.04	**0.23
II-B2. M. p. h. in straight climb (instruments covered), deviation from correct		−.05	−.10	**.20
I-B1. Four 180° climbing turns: Sum of time (in seconds) difference between turns 1 and 2 and 3 and 4, disregarding sign	−.17	.02	.04	***.26
II-G1. Sum of time (in seconds) difference between turns 1 and 2 and 3 and 4, disregarding sign		.04	.13	***.36
III-B1. Sum of time (in seconds) difference between turns 1 and 2 and 3 and 4, disregarding sign	**.43	*.17	***.31	***.26
II-B3. Sum of time (in seconds) difference between turns 1 and 2 and 3 and 4, disregarding sign (Instruments covered)		*.15	**.21	**.22
I-B2. Four 180° climbing turns; A/S range, sum of all 4 turns	.18	.13	**.21	*.18
II-G2. A/S range, sum of all 4 turns		***.37	.13	***.31
III-B2. A/S range, sum of all 4 turns	.17	**.25	*.20	−.06
II-B4. A/S range, sum of all 4 turns. (Instruments covered)		.04	−.07	−.07
III-B3. First 180° climbing turn, left: Coordination, in terms of deflection of ball in ball-bank	−.22	**.25	**.21	**.21
III-B4. Second 180° climbing turn, right: Coordination, in terms of deflection of ball in ball-bank		.13	.10	.01
III-B5. Third 180° climbing turn, left: Coordination, in terms of deflection of ball in ball-bank		.14	.05	.15
III-B6. Fourth 180° climbing turn, right: Coordination, in terms of deflection of ball in ball-bank		***.28	.06	.09

TABLE A6.7.—*Four 180° climbing turns 1. Left 2. Right 3. Left 4. Right*—Con.

Section B: Data for r_{bis} graduation-elimination.

Code	N_t	p_e	M_e	M_s	SD_t	r_{bis}
IB1	35	66	[b] 47.43	41.42	21.63	−0.17
IIIB1	41	68	[b] 53.50	73.00	27.20	.43
IB2	35	66	[b] 4.66	5.06	1.55	.15
IIIB2	40	67	[b] 52.04	56.54	16.25	.17
IIIB3	40	67	[b] 10.00	9.31	1.96	−.22

[b] Lower score is better score.

Section C: Data for r_{pbis} and for reliability r_{II}

Code	Upper or lower	First day				Second day				r_{II}	Reliability test-retest z-comb.
		r_{pbis}	N	M_s	SD	r_{pbis}	N	M_s	SD		
IIB1	Upper	0.08	46	[b]62.80	45.37	0.04	46	[b]67.93	39.88	0.18	0.23
	Lower		37	70.95	63.81		37	71.62	51.95	.27	
IIB2	Upper	.05	46	[b]4.78	4.77	.10	46	[b]4.63	4.83	.29	.20
	Lower		41	4.39	3.69		41	3.80	3.54	.07	
IB1	Upper	.02	39	[b]22.97	13.68	.04	39	[b]20.92	11.74	.40	.26
	Lower		38	23.66	15.75		38	21.87	11.87	.11	
IIO1	Upper	.04	41	[b]20.16	16.83	.13	44	[b]23.80	13.99	.35	.36
	Lower		42	27.52	19.68		42	27.62	16.21	.37	
IIIB1	Upper	.17	46	[b]24.80	17.13	.31	46	[b]21.15	11.10	.20	.26
	Lower		41	31.05	19.97		44	29.36	14.10	.31	
IIB3	Upper	.15	46	[b]22.11	18.30	.21	46	[b]19.20	12.20	.04	.22
	Lower		42	26.83	15.53		42	[illegible]	13.68	.40	
IB2	Upper	.13	41	[b]19.15	7.96	.21	41	[b]16.22	9.16	.16	.18
	Lower		38	21.18	8.15		38	20.00	8.35	.19	
IIO2	Upper	.37	46	[b]12.07	6.39	.13	46	[b]14.13	10.28	.18	.31
	Lower		43	18.02	8.43		43	16.81	7.12	.44	
IIIB2	Upper	.25	46	[b]21.96	9.86	.20	46	[b]20.98	10.76	−.05	−.06
	Lower		43	27.79	12.64		43	25.35	10.91	−.07	
IIB4	Upper	.04	46	[b]16.90	7.77	.07	46	[b]18.91	8.78	−.32	−.07
	Lower		43	17.67	8.85		43	17.91	6.92	−.01	
IIIB3	Upper	.25	46	[b]1.11	.31	.21	46	[b]1.11	.31	.33	.21
	Lower		45	1.31	.46		45	1.20	.50	.09	
IIIB4	Upper	.13	46	[b] 1.13	.34	.10	46	[b] 1.11	.31	−.14	.01
	Lower		44	1.23	.42		44	1.18	.39	.17	
IIIB5	Upper	.14	46	[b] 1.07	.32	.05	46	[b] 1.11	.31	−.07	.15
	Lower		43	1.16	.37		43	1.14	.35	.37	
IIIB6	Upper	.28	46	[b] 1.02	.15	.08	46	[b] 1.09	.28	−.05	.09
	Lower		45	1.22	.47		45	1.13	.34	.23	

[b] Lower score is better score.

TABLE A6.8.—*Four 180° gliding turns 1. Left 2. Right 3. Left 4. Right*

Section A: Summary data

Code (Study, maneuver, and measure)	Validity			Reliability
	1	2	3	4
	r_{bis} grad.—elim.	r_{pbis} U. versus L. first day	r_{pbis} U. versus L. second day	r_{tt} first versus second day
II-E1. A/S in straight normal glide, deviation from correct A/S (instruments covered)	....	0.04	−0.02	0.11
I-F1. Four 180° gliding turns: Sum of time (in seconds) difference between turns, (1-2) plus (3-4), disregarding sign	*0.35	.01	*.19	***.61
II-K1. Sum of time (in seconds) difference between turns, (1-2) plus (3-4), disregarding sign	....	.04	.06	***.26
II-E2. Sum of time (in seconds) difference between turns, (1-2) plus (3-4), disregarding sign, instruments covered	....	.14	.04	*.14
III-E1. Sum of time (in seconds) difference between turns, (1-2) plus (3-4), disregarding sign	*.28	−.10	−.09	**.23
I-F2. Four 180° gliding turns: A/S range, sum of all 4 turns	.12	***.30	***.36	*.17
II-K2. A/S range, sum of all 4 turns	....	*.17	**.23	.09
II-E3. A/S range, sum of all 4 turns, instruments covered	....	***.29	*.20	**.23
III-E2. A/S range, sum of all 4 turns	−.23	***.45	***.35	***.26
III-E3. First 180° gliding turn, left: Coordination, in terms of deflection of ball in ball-bank	....	−.08	.03	.11
III-E4. Second 180° gliding turn, right: Coordination, in terms of deflection of ball in ball-bank	....	−.05	−.02	.09
III-E5. Third 180° gliding turn, left: Coordination, in terms of deflection of ball in ball-bank	....	.04	−.02	.23
III-E6. Fourth 180° gliding turn, right: Coordination, in terms of deflection of ball in ball-bank (r_{bis} for 4 turns combined)	−.05	.03	.07	*.19

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_s	M_s	M_e	SD_t	r_{bis}
IF1	35	66	[b] 21.30	29.25	13.96	0.35
IIIE1	41	71	[b] 8.03	10.00	4.26	.28
IF2	35	66	[a] [b] 4.51	4.83	1.46	.12
IIIE2	40	72	[b] 50.66	51.82	21.07	−.23
IIIE3, 4, 5, 6	40	72	21.83	22.09	2.73	−.05

[b] Lower score is better score.
[a] Coded score.

Table A6.8.—*Four 180° gliding turns 1. Left 2. Right 3. Left 4. Right*—Con.

Section C: Data for r_{pbis} and for reliability r_{11}

Code	Upper or lower	First day r_{pbis}	First day N	First day M	First day SD	Second day r_{pbis}	Second day N	Second day M	Second day SD	r_{11}	Reliability test-retest z-comb.
IF1	Upper	0.01	40	b 11.10	7.93	0.19	40	b 9.43	6.28	.38	0.51
	Lower		38	11.29	7.77		39	12.03	7.47	.64	
IIK1	Upper	.04	46	b 3.52	2.14	.06	46	b 3.65	2.72	.21	.26
	Lower		43	3.72	2.37		43	3.95	2.62	.32	
IIE2	Upper	.14	46	b 3.70	3.01	.04	46	b 3.50	2.10	.21	.14
	Lower		42	4.48	2.35		42	3.62	2.09	.08	
IIIE1	Upper	.10	46	b 4.37	2.96	−.09	46	b 5.30	2.67	.29	.23
	Lower		44	3.81	3.31		44	4.77	3.27	.16	
IIE1	Upper	.04	44	b 3.80	3.51	−.02	44	b 3.52	3.47	.04	.11
	Lower		41	4.15	3.29		41	3.41	3.39	.18	
IF2	Upper	.30	41	b 21.34	7.57	.30	40	b 17.13	9.01	.20	.17
	Lower		38	26.58	9.11		38	24.87	11.21	.14	
IIK2	Upper	.17	46	b 12.39	6.32	.23	46	b 12.50	7.64	.10	.09
	Lower		43	16.63	8.91		43	16.28	8.22	.08	
IIE3	Upper	.29	46	b 14.13	7.40	.20	46	b 14.51	7.58	.18	.23
	Lower		43	19.07	8.71		43	18.49	11.64	.30	
IIIE2	Upper	.45	46	b 20.65	8.57	.35	46	b 19.35	8.82	.14	.26
	Lower		43	31.16	11.85		43	27.91	13.52	.39	
IIIE3	Upper	−.08	46	2.65	.52	.03	46	2.80	.45	.08	.11
	Lower		44	2.73	.45		44	2.77	.47	.14	
IIIE4	Upper	−.05	46	2.67	.59	−.02	46	2.72	.54	.12	.08
	Lower		44	2.73	.49		44	2.73	.54	.06	
IIIE5	Upper	.04	45	2.80	.45	−.02	45	2.60	.55	.38	.23
	Lower		43	2.77	.47		43	2.70	.67	.07	
IIIE6	Upper	.03	46	2.78	.41	.07	46	2.74	.57	.13	.19
	Lower		44	2.75	.55		44	2.66	.56	.25	

b Lower score is better score.

Table A6.9.—*360° steep turns at cruising throttle*

Section A: Summary data

Code (study, maneuver and measure)	Validity 1: r_{bis} grad.-elim.	Validity 2: r_{pbis} U. versus L. first day	Validity 3: r_{pbis} U. versus L. second day	Reliability 4: r_{11} first versus second day
I-C1. Time in seconds to complete turns: Left and right steep turns	−0.01	*0.14	*0.15	***0.71
II-C1. Time in seconds to complete turn: Left steep turn, instruments covered		*.16	.10	***.51
II-H4. Time in seconds to complete turn: Right steep turn, instruments covered		.02	.07	***.36
II-H1. Time in seconds to complete turn: Left steep turn		.04	.12	***.43
II-C4. Time in seconds to complete turn: Right steep turn, instruments covered		.13	.10	***.36
I-C2. Looks around 50 percent of the time: Left and right steep turns	.09	***.28	***.43	***.38
II-H2. Looks around 50 percent of the time: Left steep turn		***.30	.00	**.27
II-H5. Looks around 50 percent of the time: Right steep turn		*.19	−.05	***.49
II-C2. Looks around 50 percent of the time: Left steep turn, instruments covered		.04	.13	*.07
II-C5. Looks around 50 percent of the time: Right steep turn, instruments covered		−.09	.04	*.04
I-C3. Altitude range: Left and right steep turns	.13	***.62	**.23	*.16
II-H3. Altitude range: Left steep turn		***.36	.11	**.20
II-H6. Altitude range: Right steep turn		***.35	**.22	***.41
II-C3. Altitude range: Left steep turn, instruments covered		***.37	−.06	***.32
II-C6. Altitude range: Right steep turn, instruments covered		***.32	*.15	.11

TABLE A6.9.—*360° steep turns at cruising throttle*—Continued

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_s	M_s	M_e	SD_t	r_{bis}
IC1	35	66	[a, b] 6.21	6.17	2.28	−.01
IC2	35	66	[b] 4.30	4.85	1.83	.06
IC3	35	66	[a, b] 3.03	3.42	1.77	.13

[b] Lower score is better score.
[a] Coded score.

Section C: Data for r_{pbis} and for reliability r_{11}

Code	Upper or lower	First day r_{pbis}	First day N	First day M	First day SD	Second day r_{pbis}	Second day N	Second day M	Second day SD	r_{11}	Reliability test-retest z-comb.
IC1	Upper	0.14	41	[b] 59.46	15.81	0.15	41	[b] 56.10	16.06	0.61	0.71
	Lower		38	65.97	27.84		38	61.65	20.53	.70	
IIC1	Upper	.16	46	[b] 26.37	8.76	.10	46	[b] 25.98	8.87	.34	.51
	Lower		42	29.40	11.66		42	27.48	9.11	.65	
IIH4	Upper	.02	45	[b] 27.20	7.44	.07	45	[b] 26.91	8.93	.57	.36
	Lower		41	27.73	10.60		41	28.00	8.60	.06	
IIH1	Upper	.04	45	[b] 26.89	7.16	.12	45	[b] 25.53	8.76	.39	.43
	Lower		43	27.51	9.03		43	27.21	8.22	.46	
IIC4	Upper	.13	44	[b] 27.18	7.31	.10	44	[b] 26.70	7.16	.40	.56
	Lower		43	29.81	12.21		43	28.37	9.28	.68	
IC2	Upper	.28	40	2.70	.68	.43	40	2.80	.60	.34	.38
	Lower		36	2.25	.86		36	2.08	.94	.43	
IIH2	Upper	.30	44	.77	.42	.00	44	.66	.47	.48	.37
	Lower		41	.49	.50		41	.66	.47	−.01	
IIH5	Upper	.19	42	.76	.43	−.05	42	.69	.46	.71	.49
	Lower		41	.59	.41		41	.73	.44	.09	
IIC2	Upper	.04	44	.77	.42	.13	44	.82	.39	.31	.07
	Lower		38	.74	.44		38	.71	.45	−.21	
IIC5	Upper	−.09	43	.77	.42	.04	43	.79	.41	.22	.04
	Lower		37	.84	.37		37	.76	.43	−.18	
IC3	Upper	.62	41	[b] 137.80	107.47	.23	41	[b] 148.84	99.74	.21	.16
	Lower		38	273.03	216.11		38	206.05	144.89	.10	
IIH3	Upper	.36	46	[b] 45.87	44.36	.11	46	[b] 54.13	44.80	.23	.20
	Lower		43	97.67	83.18		43	64.10	45.06	.17	
IIH6	Upper	.35	45	[b] 41.56	36.75	.22	45	[b] 47.11	34.92	.41	.41
	Lower		43	88.37	82.71		43	68.37	57.60	.40	
IIC3	Upper	.37	46	[b] 60.2	51.4	−.05	46	[b] 71.1	59.3	.24	.31
	Lower		43	114.9	81.3		43	68.6	47.1	.38	
IIC6	Upper	.32	46	[b] 55.9	45.6	.15	46	[b] 59.1	50.0	.23	.11
	Lower		43	105.8	97.4		43	77.0	70.1	−.03	

[b] Lower score is better score.

Table A6.10.—*360° steep turns at full throttle*

Section A: Summary data

Code (study, maneuver, and measure)	Validity: 1. r_{bis} grad.—elim.	Validity: 2. r_{bis} U. versus L. first day	Validity: 3. r_{bis} U. versus L. second day	Reliability: 4. r_{tt} first versus second day
I-C4. Time in seconds to complete turns: Left and right steep turns	0.06	**0.24	*0.20	***0.40
II-H7. Time in seconds to complete turn: Left steep turn	....	−.06	.12	***.40
II-H10. Time in seconds to complete turn: Right steep turn	....	−.04	.12	***.65
II-C7. Time in seconds to complete turn: Left steep turn, instruments covered	....	*.19	.15	***.51
II-C10. Time in seconds to complete turn: Right steep turn, instruments covered	....	*.19	.10	***.66
III-D1. Time in seconds to complete turn: Left steep turn	−.09	−.04	−.03	***.43
III-D5. Time in seconds to complete turn: Right steep turn	−.09	.00	−.03	***.52
I-C5. Looks around 50 percent of the time: Left and right steep turns	.00	*.16	***.34	***.42
II-H8. Looks around 50 percent of the time: Left steep turn	....	*.14	−.04	**ʼ.53
II-H11. Looks around 50 percent of the time: Right steep turn	....	*.17	.06	**ʼ.30
II-C8. Looks around 50 percent of the time: Left steep turn, instruments covered	....	.14	.11	ʼ.06
II-C11. Looks around 50 percent of the time: Right steep turn, instruments covered	....	−.03	.11	ʼ—.10
III-D2. Looks around 50 percent of time: Left steep turn	.12	.07	**.23	ʼ—.15
III-D6. Looks around 50 percent of the time: Right steep turn	−.01	.10	***.33	ʼ—.20
I-C6. Altitude range: Left and Right Steep turns	*.35	***.31	***.34	*.20
II-H9. Altitude range: Left steep turn	....	***.31	.07	***.30
II-H12. Altitude range: Right steep turn	....	**.27	.02	*.17
II-C9. Altitude range: Left steep turn, instruments covered	....	**.27	.11	***.36
II-C12. Altitude range, Right steep turn, instruments covered	....	***.44	.14	*.13
III-D3. Altitude range: Left steep turn	.20	***.28	***.32	.04
III-D7. Altitude range: Right steep turn	.13	**.26	***.27	***.27
VII-C1. Maximum deviation from starting altitude: Left and right steep turns	....	***.18	....	*** ʼ.32
VII-C2. Maintained bank of 60° or greater: Left and right steep turns	....	*.05	....	** ʼ.17
VII-C3. Final roll-out within .45° of starting heading: Left and right steep turns	....	.03	....	ʼ.03
III-D4. Coordination, in terms of deflection of ball-bank: Left steep turn	−.12	.04	*.18	**.22
III-D8. Coordination, in terms of deflection of ball-bank: Right steep turn	.07	*.20	*.05	*.19

Section B: Date for r_{bis} graduation-elimination

Code	N_t	p_g	M_g	M_t	SD_t	r_{bis}
IC4	35	66	[b] 6.50	6.74	1.90	0.06
IIID1	42	69	[b] 43.00	41.62	9.16	−.09
IIID5	36	69	[b] 44.32	42.64	11.93	−.09
IC5	35	66	4.47	4.47	1.73	.00
IIID2	42	69	.97	.85	.63	.12
IIID6	38	71	.89	.91	.68	−.01
IC6	35	66	[b] 2.61	3.62	1.78	.35
IIID3	42	69	[b] 292.41	370.69	234.22	.20
IIID7	40	70	[b] 268.21	308.33	191.53	.13
IIID4	40	70	5.25	5.42	.87	−.12
IIID8	41	68	4.96	4.85	1.22	.07

TABLE A6.10.—*360° steep turns at full throttle*—Continued

Section C: Data for r_{bis}: and for reliability r_{II}

Code	Upper or lower	First day r_{bis}	First day N	First day M	First day SD	Second day r_{bis}	Second day N	Second day M	Second day SD	r_{II}	Reliability test-retest s-comb.
IC4	Upper	0.24	40	[b]47.35	12.05	0.20	41	[b]46.95	13.34	.58	0.19
	Lower		36	55.22	18.92		37	53.51	18.52	.39	
IIH7	Upper	—.06	45	[b]21.58	5.66	.12	45	[b]21.01	4.27	.49	.49
	Lower		43	20.91	5.02		43	22.19	4.94	.48	
IIH10	Upper	—.04	46	[b]23.07	6.83	.12	4[illegible]	[b]22.61	5.05	.67	.65
	Lower		40	22.50	6.80		40	21.00	3.42	.62	
IIC7	Upper	.19	46	[b]19.89	4.69	.15	46	[b]19.74	3.74	.40	.51
	Lower		43	22.02	6.10		43	21.33	6.20	.61	
IIC10	Upper	.19	43	[b]21.40	5.71	.10	45	[b]21.71	4.84	.56	.63
	Lower		41	24.22	8.74		41	23.00	8.11	.70	
IIID1	Upper	—.04	46	[b]21.52	6.32	—.03	46	[b]21.76	6.11	.40	.43
	Lower		45	21.09	5.71		45	21.38	5.02	.46	
IIID5	Upper	.00	45	[b]21.73	4.80	—.03	45	[b]22.20	6.52	.55	.52
	Lower		39	21.77	6.54		39	21.79	7.31	.48	
IC5	Upper	.16	39	2.59	.74	.34	39	2.62	.74	.51	.42
	Lower		33	2.33	.88		33	2.03	.90	.29	
IIH8	Upper	.14	43	.70	.46	—.04	43	.67	.47	[a].62	[a].53
	Lower		41	.56	.50		41	.71	.45	[a].31	
IIIH11	Upper	.17	40	.75	.43	—.06	40	.68	.47	[a].36	[a].39
	Lower		37	.59	.49		37	.73	.44	[a].39	
IIC8	Upper	.14	44	.80	.40	.11	44	.80	.40	[a].03	[a].03
	Lower		37	.68	.47		37	.70	.46	[a].09	
IIC11	Upper	—.03	41	.78	.41	.11	41	.85	.35	—[a].11	—[a].16
	Lower		35	.80	.40		35	.77	.42	—[a].21	
IIID2	Upper	.07	45	.49	.50	.23	45	.69	.46	—.03	—.15
	Lower		45	.42	.49		45	.47	.50	—.27	
IIID6	Upper	.10	39	.54	.50	.33	39	.74	.44	—.33	—.20
	Lower		41	.46	.50		41	.41	.49	—.08	
IC6	Upper	.31	41	[b]166.10	132.29	.34	41	[b]140.00	100.24	.32	.23
	Lower		38	293.06	256.71		37	258.22	206.75	.06	
IIH9	Upper	.31	46	[b]44.35	46.23	.07	46	[b]61.96	36.09	.61	.39
	Lower		42	92.62	94.24		42	67.62	41.45	.05	
IIH12	Upper	.27	46	[b]55.65	49.46	.02	46	[b]54.04	47.49	.19	.17
	Lower		41	95.85	91.38		41	59.76	41.11	.14	
IIC9	Upper	.27	46	[b]66.00	60.20	.11	46	[b]75.40	58.70	.47	.36
	Lower		43	98.60	55.70		43	87.90	59.30	.23	
IIC12	Upper	.44	45	[b]50.00	33.60	.14	45	[b]62.40	47.30	.06	.13
	Lower		43	101.60	67.10		43	76.50	53.30	.19	
IIID3	Upper	.28	46	[b]106.50	89.70	.32	46	[b]78.30	62.10	.02	.04
	Lower		45	207.33	225.91		45	141.11	119.11	.06	
IIID7	Upper	.26	46	[b]87.80	90.00	.27	46	[b]80.90	60.00	.11	.27
	Lower		43	164.40	186.50		43	143.00	48.00	.43	
VIIC1	85 hrs	.18	954	1.79	.60	.....	.....	(N=170)			[a].33
	60 hrs		211	1.47	.62						
IIID4	Upper	.04	45	2.47	.78	.18	45	2.60	.59	.12	.22
	Lower		44	2.41	.78		44	2.43	.78	.31	
IIID8	Upper	.20	46	2.76	.48	.05	46	2.80	.65	.27	.19
	Lower		43	2.51	.73		43	2.74	.49	.11	
VIIC2	85 hrs	.05	954	2.72	.58	.....	.....	(N=170)			[a].17
	60 hrs		211	2.64	.64						
VIIC3	85 hrs	.03	952	2.98	.20	.....	.....	(N=170)			.03
	60 hrs		211	2.96	.27						

[b] Lower score is the better score.

TABLE A6.11.—*Maintaining altitude while reducing A/S*

Section A: Summary data

Code (Study, maneuver, and measure)	Validity 1: r_{bis} grad.—elim.	Validity 2: r_{bis} U. versus L. first day	Validity 3: r_{bis} U. versus L. second day	Reliability 4: r_{II} first versus second day
I-D2. Altitude range, sum of two trials	*0.26	***0.34	0.01	***0.27
I-D2. Altitude range, sum of first day's trials	*.30	....	....	....
I-D2. Altitude range, sum of second day's trials	.10	....	....	....

TABLE A6.11.—*Maintaning altitude while reducing A/S*—Continued

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_t	M_g	M_e	SD_t	r_{bis}
ID2	36	66	[b] 2.60	4.43	1.78	0.28
ID2 1st day	35	66	[b] 1.87	2.43	1.19	.30
ID2 2nd day	35	66	[b] 1.82	1.99	.96	.10

[b] Lower score is better score.

Section C: Data for r_{pbis} and for reliability, r_{tt}

Code	Upper or lower	First day				Second day				r_{tt}	Reliability test-retest z-comb.
		r_{pbis}	N	M	SD	r_{pbis}	N	M	SD		
ID2	Upper	0.34	41	[b] 195.37	89.34	0.01	41	[b] 210.73	117.88	0.18	0.27
	Lower		38	286.84	156.54		38	212.16	104.38	.30	

[b] Lower score is better score.

TABLE A6.12.—*Rudder-exercise and power-on stalls*

Section A: Summary data

Code (Study, maneuver, and measure)	Validity			Reliability
	1	2	3	4
	r_{bis} grad.—elim.	r_{pbis} U. versus L. first day	r_{pbis} U. versus L. second day	r_{tt} first versus second day
II-D1. Rudder-exercise stall: Time to lose 500 feet		0.09	**0.26	0.12
II-D2. Rudder-exercise stall: Maximum angle of wings with horizon		***.35	**.22	**.24
II-D3. Rudder-exercise stall: Maximum directional change		***.40	**.24	**.22
III-C1. Power-on stall recovery: Stalled ailerons	*0.38	*.18	.03	(¹)
III-C2. Power-on stall recovery: Time from forward stick to level flight. Short time is best	.16	**.25	*.16	*.15
III-C3. Power-on stall recovery: Throttle forward with stick in recovery	.24	*.19	*.17	(¹)
III-C4. Power-on stall recovery: Maximum A/S in recovery. Low A/S is best	**.43	*.20	**.22	***.41

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_t	M_g	M_e	SD_t	r_{bis}
IIIC1	42	69	1.62	1.23	0.63	0.38
IIIC2	42	69	[b] 15.45	16.15	2.66	.16
IIIC3	42	69	1.75	1.54	.56	.24
IIIC4	42	69	[b] 173.10	190.77	24.35	.43

[b] Lower score is better score.

Table A6.12.—*Rudder-exercise and power-on stalls*—Continued

Section C: Data for r_{pbis}, and for reliability, r_{tt}

Code	Upper or lower	First day r_{pbis}	First day *N*	First day *M*	First day *SD*	Second day r_{pbis}	Second day *N*	Second day *M*	Second day *SD*	r_{tt}	Reliability test-retest s-comb.
IID1	Upper	0.09	43	19.23	8.47	0.26	43	19.72	8.59	0.12	0.12
	Lower		29	18.07	7.14		29	16.76	4.87	.13	
IID2	Upper	.35	46	1.72	1.25	.22	46	1.73	[illegible]	[illegible]	.24
	Lower		43	.79	1.27		43	1.21	1.21	.18	
IID3	Upper	.40	46	2.04	1.32	.24	46	1.96	1.28	.23	.22
	Lower		43	.88	1.35		43	1.33	1.32	.20	
IIIC1	Upper	.18	46	.83	.38	.03	46	.87	.34	(1)	(1)
	Lower		45	.67	.47		45	.84	.36	.16	
IIIC2	Upper	.25	46	b 7.17	1.43	.16	46	b 7.17	1.39	.35	.16
	Lower		45	8.04	1.92		45	7.73	−2.09	.07	
IIIC3	Upper	.19	46	.96	.20	.17	46	.96	.20	(1)	(1)
	Lower		44	.84	.37		44	.86	.34	.37	
IIIC4	Upper	.20	46	b 84.35	8.88	.22	46	b 84.57	6.33	.30	.41
	Lower		45	88.33	10.59		45	90.41	17.85	.52	

[1] Not calculated because of nature of distribution.
[b] Lower score is better score.

Table A6.13.—*Spin recovery*

Section A: Summary data

Code (Study, maneuver and measure)	Validity 1: r_{bis} grad.-elim.	Validity 2: r_{pbis} U. versus L. first day	Validity 3: r_{pbis} U. versus L. second day	Reliability 4: r_{tt} first versus second day
I-E1. Did student apply opposite rudder at 2 turns ±45°?	*0.28	***0.33	*0.17	0.01
II-J1. Did student apply opposite rudder at 1½ turns within ±22°, ±45°, or more?	....	***.31	.13	.12
I-E2. Time in seconds from opposite rudder till nose cuts horizon. Short time is best	*.32	.13	***.33	***.37
II-J2. Time in seconds from opposite rudder till nose cuts horizon. Short time is best	....	.11	.06	−.10
I-E3. Maximum A/S in dive, corrected for indicator error by subtraction of cruising indicated A/S. Low A/S is best. (See text)	*.30	.08	.06	**.24
II-J3. Maximum A/S in dive, corrected for indicator error by subtraction of cruising indicated A/S. Low A/S is best. (See text)	....	.11	−.01	***.30

Section B: Data for r_{bis}, graduation-elimination

Code	N_t	p_g	M_g	M_e	SD_t	r_{bis}
IE1	35	66	a b 6.56	7.35	1.74	0.28
IE2	35	66	b 8.59	9.32	1.39	.32
IE3	35	66	b 88.48	95.45	14.19	.30

[b] Lower score is better score.
[a] Coded score.

Table A6.13.—*Spin recovery*—Continued

Section C: Data for r_{pbis}, and for reliability, r_{II}

Code	Upper or lower	First day r_{pbis}	First day N	First day M	First day SD	Second day r_{pbis}	Second day N	Second day M	Second day SD	r_{II}	Reliability test-retest z-comb.
IE1	Upper	0.33	41	1.95	0.22	0.17	41	1.90	0.30	−0.07	0.01
	Lower		37	1.70	.46		37	1.78	.41	.09	
IIJ1	Upper	.31	46	1.87	.34	.13	46	1.89	.38	.23	.12
	Lower		42	1.50	.73		42	1.76	.57	.00	
IE2	Upper	.13	40	[b] 8.40	2.60	.33	36	[b] 7.28	1.41	.51	.37
	Lower		33	9.09	2.63		33	8.31	1.51	.16	
IIJ2	Upper	.11	46	[b] 8.15	2.57	.06	46	[b] 8.22	2.06	−.23	−.10
	Lower		42	8.76	2.75		42	8.50	2.47	.04	
IE3	Upper	.08	41	46.59	9.72	.06	41	46.59	11.81	.46	.24
	Lower		37	45.00	10.40		37	45.41	9.33	−.03	
IIJ3	Upper	.11	45	[b] 25.44	8.29	−.01	45	[b] 27.22	8.54	.36	.30
	Lower		42	27.50	10.37		42	27.02	9.13	.26	

[b] Lower score is better score.

Table A6.14.—*Validity and reliability coefficients of objective items used in primary check ride administered in study IV*

Section A: Summary data

	1 Validity r_{bis} Pass-fail	2 Validity r_{pbis} Upper-lower	3 Reliability r_{II} Test-retest
Medium Turns:			
a. Agreement on coordination	−0.07	0.14	−0.11
Spins:			
b. Agreement on direction	.09	**.28	−.06
c. Agreement on number of turns	*.30	.16	*.17
d. Agreement on application of stick	−.24	.02	.10
Forward Slips:			
e. Air speed range	*.27	**.27	.10
f. Score on highest air speed	**.39	.14	**.28
g. Part of plane over road	.03	*.21	*.23
h. Slipped ½ ball	*.34	*.25	***.39
i. Instructor took over	−.17	.09	.00
Slipping Turns:			
j. Air speed range	−.08	**.25	**.29
k. Score on highest air speed	.07	*.22	***.31
l. Plane stalled	−.17	.14	***.44
m. Remained in slip	*.30	**.25	***.40
n. Recovery made at ±5° from 90°	*.30	***.45	**.27
o. Instructor took over	.20	**.27	−.04
Power Landings:			
p. Approach pattern	.11	**.26	***.44
q. Place of landing	.00	.14	**.19
r. Air speed range	.00	.11	−.02
s. Score on highest air speed	−.01	*.21	−.01
t. Land in crab	.17	*.17	−.15
u. Land drifting	***.45	**.26	.09
v. Type of landing	*.27	**.27	*.17
w. Plane bounced	−.04	.02	.12
x. Landing roll	***.46	*.24	*−.17
Cross-wind Landings:			
y. Variation from 500 feet in traffic	.12	***.48	−.07
z. Ball deviation on approach turn	.22	***.38	−.01

*Correlation significant at 20-percent level or above.
**Correlation significant at 5-percent level or above.
***Correlation significant at 1-percent level or above.

Table A6.14.—*Validity and reliability coefficients of objective items used in primary check ride administered in study IV*—Continued

Section B: Data for biserial correlation with graduation-elimination

	N_t	p_e	M_g	M_e	SD_t	r_{bis}
a	51	80	7.98	8.20	1.86	−0.07
b	39	82	11.81	11.71	.61	.09
c	50	80	24.73	23.70	1.91	*.30
d	48	79	8.32	9.30	2.10	−.24
e	49	80	28.23	26.80	3.01	*.27
f	50	80	13.58	11.10	3.64	**.39
g	51	80	8.54	8.40	2.53	.03
h	51	80	10.68	9.40	2.11	*.34
i	51	80	11.85	12.00	.47	−.17
j	48	79	28.26	28.70	2.87	−.08
k	48	79	15.21	14.80	3.19	.07
l	49	84	11.71	12.00	.96	−.17
m	51	80	9.66	8.30	2.79	*.30
n	51	80	9.95	8.80	2.17	*.30
o	51	80	11.85	11.60	.71	.20
p	42	79	9.00	8.56	2.19	.11
q	45	78	11.40	11.40	.08	.00
r	50	80	20.60	20.60	2.65	.00
s	50	80	10.83	10.90	3.15	−.01
t	50	80	10.55	10.00	1.89	.17
u	48	79	10.79	9.20	2.01	***.45
v	51	80	9.39	8.20	2.45	*.27
w	51	80	8.44	8.60	2.10	−.04
x	51	80	8.78	7.00	2.18	***.46
y	48	81	21.59	20.78	3.78	.12
z	48	81	15.66	14.48	3.00	.22

Section C: Data for point-biserial correlation with upper and lower groups chosen by instructors

	N_u	M_u	SD_u	N_L	M_L	SD_L	r_{pbis}
a	29	8.24	2.13	27	7.70	1.61	0.14
b	24	11.92	.40	18	11.56	.83	**.28
c	28	24.64	1.84	27	24.00	2.04	.16
d	26	8.46	2.17	27	8.37	2.11	.02
e	27	28.67	2.55	27	27.04	3.24	**.27
f	28	13.84	3.22	27	12.48	4.01	.14
g	29	9.03	2.50	27	8.00	2.37	*.21
h	29	11.63	1.71	27	10.00	2.24	**.25
i	29	11.93	.36	27	11.85	.52	.09
j	28	29.04	2.58	25	27.64	2.77	**.25
k	28	15.75	2.50	25	14.36	3.60	*.22
l	29	11.86	.51	25	11.60	1.26	.14
m	29	10.00	2.88	27	8.67	2.31	**.25
n	29	10.55	1.73	27	8.59	2.16	***.45
o	29	12.00	.00	27	11.63	.95	**.27
p	24	9.54	2.58	23	8.39	1.52	**.28
q	24	11.46	.96	25	11.16	1.12	.14
r	29	20.66	2.56	26	20.04	2.90	.11
s	29	11.31	2.93	26	9.06	3.32	*.21
t	28	10.64	1.86	24	10.00	1.84	*.17
u	27	11.04	1.87	25	10.00	2.19	**.28
v	29	9.76	2.70	26	8.42	1.95	**.27
w	29	8.17	2.12	26	8.81	1.94	.02
x	29	8.97	2.39	26	7.92	1.80	*.24
y	27	23.15	3.00	24	19.62	3.36	***.45
z	27	16.86	2.77	26	14.42	2.34	***.35

Table A6.14.—*Validity and reliability coefficients of objective items used in primary check ride administered in study IV*—Continued

Section D: Data for test-retest reliabilities

	First ride			Second ride			r_{12}
	N	M	SD	N	M	SD	
a	56	3.88	1.44	56	4.11	1.44	−0.11
b	42	5.86	.52	42	5.90	.43	−.06
c	55	12.18	1.19	55	12.15	1.37	*.17
d	53	4.34	1.54	53	4.08	1.34	.16
e	54	13.59	2.28	54	14.26	1.77	.10
f	55	6.24	2.39	55	6.78	2.19	**.28
g	56	4.21	1.50	56	4.32	1.68	*.23
h	56	5.07	1.41	56	5.46	1.03	***.39
i	56	6.00	.00	56	5.89	.45	.00
j	53	13.83	1.94	53	14.55	1.49	**.29
k	53	7.15	2.19	53	7.94	1.68	***.31
l	54	5.85	.65	54	5.89	.46	***.44
m	56	4.32	1.80	56	5.01	1.41	***.40
n	56	4.86	1.25	56	4.75	1.49	**.27
o	56	5.93	.37	56	5.89	.59	−.04
p	47	4.30	1.40	47	4.68	1.20	***.44
q	49	5.69	.58	49	5.61	.78	*.19
r	55	13.47	1.88	55	12.89	2.04	−.02
s	55	5.89	2.42	55	4.78	2.11	−.01
t	54	5.33	1.33	54	5.00	1.53	−.15
u	52	5.31	1.35	52	5.23	1.31	.09
v	55	4.42	1.51	55	4.71	1.71	*.17
w	55	4.15	1.26	55	4.33	1.49	.12
x	55	4.29	1.64	55	4.18	1.76	*−.17
y	53	10.43	2.87	53	10.98	2.45	−.07
z	53	7.45	1.88	53	8.06	2.07	−.01

Table A6.15.—*In study IV, correlation between instructor's ratings and graduation-elimination*

	N_t	P_{p-e}	P_{p-L}	P_{f-e}	P_{f-L}	r_{tt}
Graduation-elimination v. Instructor's ratings	51	53	28	.02	17	***0.75

Table A6.16.—*Data for study*

Graduation—Elimination

	N_t	P_g	M_g	M_e	SD_t	r_{bis}
Time [1]	33	79	[b] 66.57	74.25	9.20	**0.48

Upper v. Lower

	N_t	P_u	M_u	M_L	SD_t	r_{bis}
Time [1]	46	33	[b] 66.20	71.10	9.87	*0.30

Test-retest reliability

	First day			Second day			r_{12}
	N	M	SD	N	M	SD	
Time [1]	46	35.07	6.13	46	34.41	6.14	0.16

[1] In minutes—sum of two check rides.

Table A6.17.—*Semi-partial reliability correlations in Study VI*

Formula:

$$r_{(1\cdot 3)(2\cdot 4)} = \frac{r_{12} - r_{24}r_{14} - r_{13}r_{23} + r_{13}r_{24}r_{34}}{\sqrt{1-r^2_{13}}\,\sqrt{1-r^2_{24}}}$$

Variable $_1$ = Odd score. Variable $_3$ = Odd hours.
Variable $_2$ = Even score. Variable $_4$ = Even hours.

Code (Study, maneuver, and measure)	r_{12}	r_{13}	r_{14}	r_{23}	r_{24}	r_{34}	$r_{(1\cdot 3)(2\cdot 4)}$	*$r^*_{(1\cdot 3)(2\cdot 4)}$
V-A1	0.66	0.42	0.48	0.47	0.53	0.95	0.55	0.71
V-A2	.82	.52	.52	.63	.66	.95	.75	.86
V-A3	.75	.01	.07	.13	.18	.95	.76	.86
V-A4	.73	.43	.49	.51	.57	.95	.61	.76
V-A5	.72	.43	.54	.45	.53	.95	.60	.75
V-A6	.77	.44	.48	.51	.54	.95	.68	.81
V-A7	.85	.57	.59	.65	.68	.95	.75	.86
V-B1	.71	.46	.56	.53	.60	.95	.55	.71
V-B2	.86	.66	.60	.74	.75	.95	.64	.78
V-B3	.86	.34	.41	.33	.37	.95	.83	.91
V-B4	.47	.35	.48	.39	.49	.95	.30	.46
V-B5	.51	.51	.59	.40	.44	.95	.34	.51
V-B6	.67	.56	.61	.50	.65	.95	.46	.63
V-B7	.57	.35	.44	.22	.35	.95	.52	.68
V-B8	.69	.56	.60	.38	.44	.95	.61	.76
V-B9	.64	.45	.47	.50	.56	.95	.53	.09
V-B10	.79	.70	.73	.63	.64	.95	.56	.72
V-B11	.50	−.25	−.11	−.15	.00	.95	.47	.64

*$r^*_{(1\cdot 3)(2\cdot 4)}$ is the semi-partial odd-even reliability correlation corrected by the Spearman-Brown formula to 12 administrations.

Table A6.18.—*Data for study VI*

2. Check ride by regular instructor.
3. Check ride by different instructor ("switch"-ride).
4. Check ride by regular instructor.

Code (Study, maneuver and measure)	r_{23}	r_{34}	Combination* of r_{23} and r_{34}	Spearman-Brown for 12 times length
V-A1	0.20	0.64	0.45	0.91
V-A2	.03	−.08	−.03	−.27
V-A3	.49	.19	.34	.86
V-A4	.15	.23	.19	.74
V-A5	.13	.24	.19	.74
V-A6	−.06	−.18	−.12	−.62
V-A7	.20	.28	.24	.79
V-B1	−.06	−.14	−.10	−.57
V-B2	.01	.07	.04	.33
V-B3	.05	.29	.17	.71
V-B4	−.03	−.1[illegible]	−.07	−.47
V-B5	.28	.12	.20	.75
V-B6	.18	−.02	.08	.51
V-B7	.22	.04	.13	.64
V-B8	.10	.07	.09	.54
V-B9	.01	.12	.07	.47
V-B10	−.13	.11	−.01	−.11
V-B11	−.14	−.10	−.12	−.62

*Combined by the z-transformation technique.

TABLE A6.19.—*Data from study VII for test-retest reliabilities*

Code	Garner, N=70					Chickasha, N=100				r	z^{*}_{comb}	r^{*}
	Test		Retest		r	Test		Retest				
	M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*			
VII-A1	2.51	0.79	2.59	0.76	0.07	2.32	0.92	2.39	0.89	0.18	0.14	0.15
VII-B1	2.14	.96	2.31	.89	.00	1.84	.96	2.18	.97	.34	.20	.28
VII-C1	1.81	.62	1.97	.76	.20	1.47	.68	1.46	.61	.25	.23	.32
VII-C2	2.77	.48	2.76	.52	.01	2.63	.64	2.69	.63	.21	.13	.17
VII-C3	2.91	.41	2.86	.52	.06	3.00	.00	2.96	.28	.00	.02	.03
VII-D1	2.97	.24	2.94	.33	.02	2.78	.63	2.86	.51	.03	.03	.03
VII-D2	2.86	.52	2.74	.67	.06	2.66	.75	2.40	.92	.05	.05	.07
VII-E1	1.59	.78	1.77	.81	−.02	1.59	.76	1.45	.74	.04	.02	.03
VII-E2	2.67	.65	2.54	.71	−.23	2.16	.91	2.35	.90	.05	−.07	−.09
VII-E3	2.34	.75	2.36	.83	−.04	2.14	.92	2.12	.91	.01	−.01	−.01
VII-F1	1.83	.76	1.77	.81	.05	1.59	.75	1.58	.78	−.02	.01	.01
VII-F2	2.69	.62	2.66	.63	.24	2.26	.85	2.54	.77	.08	.14	.20
VII-F3	2.41	.82	2.37	.85	.11	2.20	.86	2.37	.81	−.06	.01	.01

The coefficients for the two elementary schools were combined by the z-transformation technique and then corrected for broad categories used in this Freeze Study.

TABLE A6.20.—*Landings in study VII. Test-retest reliabilities Garner and Chickasha data combined N=170*

	r_{tt}	First day			Second day		
		N	*M*	*SD*	*N*	*M*	*SD*
Zone plane lands:							
1st landing, 1st day *v.* 2d landing, 2d day.	0.00	170	1.71	0.79	170	1.66	0.80
2d landing, 1st day *v.* 1st landing, 2d day.	*.10	170	1.60	.77	170	1.88	.70
Landing attitude:							
1st landing, 1st day *v.* 2d landing, 2d day.	−.03	170	2.36	.85	170	2.50	.72
2d landing, 1st day *v.* 1st landing, 2d day.	−.04	170	2.44	.79	170	2.45	.83
Plane bounced or dropped:							
1st landing, 1st day *v.* 2d landing, 2d day.	.07	170	2.24	.85	170	2.35	.84
2d landing, 1st day *v.* 1st landing, 2d day.	*−.13	170	2.29	.84	170	2.23	.86

APPENDIX TO CHAPTER SEVEN

Measures of Two-Engine Flying Skill (Contact)

TABLE A7.1.—*Summary of TB-25 contact objective flying scale*

Maneuver	*Variables measured*
Formation straight and level.	Manifold pressure deviations. Manifold pressure differences in two engines. Climb above lead plane or drop below. Lag. Crowd up.
Formation cross under (and return).	Total time. Maximum lag. Maximum lateral overshoot.
Formation turns (two 180's).	Over-run or crowd up. Lag. Climb above lead plane. Lateral error.
Formation let-down and climb.	Manifold pressure deviations. Over-run. Lag. Climb. Drop.
Steep turns to right and left.	Altitude. Air speed. Bank.
Slow flying—straight and level.	Heading deviations. Altitude changes. Time to reach 100 m. p. h. Air speed variations.
Slow flying turns (right and left).	Air speed deviations. Bank. Altitude changes. Cylinder head temperatures.
Single-engine out of slow flying.	Power control and attitude. Raise gear and flaps. Air speed control. Heading deviations. Altitude control.
Power off stall (gear and flaps up).	Altitude control. Heading deviations. Rate of climb. Secondary stall.
Power off stall (gear and flaps down).	Heading deviations. Altitude control at break and recovery. Bank. Gear and flaps up during recovery.

TABLE A7.1.—*Summary of TB-25 contact objective flying scale*—Continued

Maneuver	*Variables measured*
Short field landings (two).	Air speed during approach. Air speed at initial contact. Landing zone. Addition of power. "Drop in." Where cut power.

TABLE A7.2.—*Sample contact maneuver*

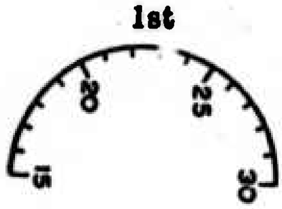
Formation—Straight and Level

Lead plane climbs to 5,000 feet above the ground and flies a straight and level course. Student flies normal position off right wing. Lead plane should fly at 180 m. p. h. with 1,900 r. p. m. Allow the first 30 seconds for student to adjust his position and speed. Make test observations during the next 4 minutes.

A. Manifold pressure left engine during:

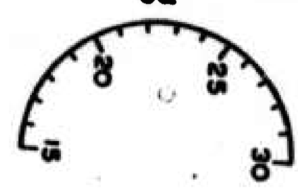
Minutes:

1st



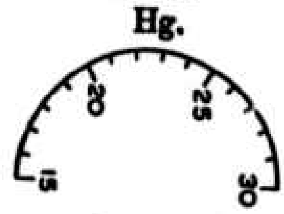
2d

3d



4th

Inches Hg.



Student's score

Deviation	Points
2 inches	(5)
3 inches	(4)
4 inches	(3)
5 inches	(2)
Over 5 inches	(1)

Score:

---- ---- ---- ----

B. During the 4 minutes, the *maximum difference* in manifold pressure between the two engines was:

2 inches or less................. (5)
3 or 4 inches.................... (3)
Over 4 inches.................... (1)

TABLE A7.2.—*Sample contact maneuver*—Continued

C. How many times did student climb above and/or drop below lead plane by 10 feet or more (approximate thickness of fuselage):

0 times	(5)	3 times	(2)
1 time	(4)	4 times or more	(1)
2 times	(3)		

D. Maximum lag at any time was:

Within ½ plane length	(5)
Within full plane length	(3)
More than one plane length	(1)

E. How many times did student crowd up so check pilot could see *leading edge of left vertical stabilizer* of lead plane:

0 times	(5)	3 times	(2)
1 time	(4)	4 times or more	(1)
2 times	(3)		

Table A7.3.—*Intercorrelations among the measures in the Advanced two-engine check ride administered during the temporary training freeze* [1]

N=725 *S.E.* of correlation of zero=.037

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	BB	C	
Single-Engine Procedure:																				
J—Heading (maximum deviation)	0.17	0.17	0.19	0.03	0.08	0.01	0.02	0.02	0.04	0.13	0.05	0.08	0.07	0.09	0.07	0.06	−0.02	0.02	0.07	*J*
K—Air speed (minimum A/S)		.20	.08	.04	.10	.06	.09	.03	.04	.04	−.05	.00	.08	.06	.10	−.03	.03	.07	−.02	*K*
L—Altitude (maximum lost)			.28	.00	.07	.01	.04	.04	−.05	−.03	−.06	.02	.02	.05	.10	.06	.06	.01	.03	*L*
M—Procedure check list				[2].07	.03	[2].06	.07	.05	[2].06	−.01	.00	−.02	−.02	.04	.04	.08	[2].16	.07	.05	*M*
Single-Engine Landing:																				
N—Turn onto approach										.05	.06	.07	.01	.08	.09	.03	[2]−.11	.10	.01	*N*
O—A/S on the approach					[2]					.08	.05	.09	.05	.00	.08	.05	.01	.08	−.01	*O*
P—Was gear checked after lowering? [3]										.04	.03	.04	−.02	.06	.09	−.01	[2].01	.11	−.02	*P*
Q—Where was power cut completely?										.03	.01	.05	−.01	.06	.08	.01	−.08	.08	.01	*Q*
R—Where was landing first made?										.03	.02	.07	−.05	.04	.10	.03	−.07	.17	.07	*R*
S—Landed in skid?										.01	.03	.07	.01	.05	.05	.07	[2]−.02	.08	.03	*S*
Straight and Level Flight (Instrument):																				
T—Heading deviation											.27	.29	.12	.08	.09	−.03	.01	−.04	−.03	*T*
U—Air speed variation												.43	.08	.09	.08	.01	−.01	−.05	.01	*U*
V—Altitude variation													.06	.07	.09	.04	.00	.00	.01	*V*
Instrument Let-Down and Low Approach:																				
W—Position														.30	.17	−.01	.09	−.01	−.01	*W*
X—Time															.37	.15	.01	.09	.04	*X*
Y—Altitude																.01	−.04	.07	.03	*Y*
Power-Off Approach and Landing:																				
Z—Altitude at roll into approach																	.05	.09	−.02	*Z*
AA—Was power added on approach?																		−.10	−.04	*AA*
BB—When was landing first made?																			−.02	*BB*

C=Pilot Stanine.

[1] See table A7.5 for M's and S.D.'s and appendix to chapter 10 for copy of check ride booklet.

[2] Tetrachoric correlations.

[3] See table A7.4.

Table A7.4.—*Intercorrelations of single-engine landing, two-engine Advance training, first training freeze*

	Item	N	O	P	Q	R	S
Lined up with runway	N	....	0.53	[1] 0.73	0.54	0.59	[1] 0.78
A/S on approach	O	0.13	....	.45	.60	.56	.51
Was gear checked?	P	[1] .05	.01	....	.63	.61	[1] .74
Where power was cut	Q	.05	.27	.08	....	.64	.53
Landing zone	R	−.02	.07	.05	.13	....	.64
Landed in skid	S	[1] .13	.06	[1] .01	−.02	.06	....

[1] Tetrachoric correlations.

Note on tables A7.3 and A7.4. These tables give the intercorrelations of the 19 measures and the pilot stanine. In table A7.4 are the intercorrelations for the six measures in the single-engine landing maneuver. None of the coefficients in these two tables was corrected for the effect of scoring in broad categories.

The upper right-hand half of the matrix in table A7.4 is based on all cases of class 44–J, a total of 725 cases. The lower left-hand half of the matrix is based on the same group after removing 214 cases who all received the same score of "1" on all 6 measures in this maneuver since they failed to make a satisfactory landing and had to "go around." It was necessary to make this correction since giving these students the lowest score on all of the measures will, of course, increase the intercorrelation among them and would thus give a false picture of the relations between the measures of the single-engine landing. The uncorrected values are shown in the upper right-hand half of the matrix.

Table A7.5.—*Means and standard deviations for the 19 objective measures presented in tables A7.3 and A7.4*

N=725

Item	Mean	Standard deviation	Means and standard deviations after removing 214 cases with scores of "1." (Table A7.4, lower left-hand half of matrix.) *N*=511	
			Mean	*SD*
J	2.47	0.611	....	....
K	2.12	.686	....	....
L	2.22	.791	....	....
M	2.72	.689	....	....
N	2.10	.995	2.58	0.836
O	1.75	.765	2.07	.704
P	2.06	.997	2.54	.844
Q	2.05	.958	2.47	.779
R	1.94	.731	2.32	.485
S	2.14	.991	2.62	.787
T	2.02	.785	....	....
U	2.60	.577	....	....
V	2.40	.607	....	....
W	1.77	.785	....	....
X	2.30	.790	....	....
Y	1.85	.840	....	....
Z	2.81	.457	....	....
AA	2.40	.916	....	....
BB	2.13	[illegible]	....	....

APPENDIX TO CHAPTER EIGHT

Objective Measures of Multiengine Flying Skill

TABLE A8.1.—*Means and sigmas of scores by check pilots and second observers during the straight and level course with changing air speed*

Supplement to table 8.1

	Altitude					
	Check pilot			Second observer		
	N	*M*	*SD*	*N*	*M*	*SD*
		Feet	*Feet*		*Feet*	*Feet*
Time sample	60	385.85	219.00	60	404.45	202.30
Range	60	86.75	48.02	60	109.75	42.86
Limits	60	58.83	32.40	60	68.75	28.16

	Heading					
	Check pilot			Second observer		
	N	*M*	*SD*	*N*	*M*	*SD*
		Degrees	*Degrees*		*Degrees*	*Degrees*
Time sample	60	13.57	12.36	60	14.03	11.40
Range	60	4.03	2.47	60	4.10	2.28
Limits	60	2.78	1.64	60	2.67	1.36

TABLE A8.2.—*Means and SD's of test and retest*

Supplement to table 8.2

Altitude in steep turn

	First turn			Second turn		
	N	*M*	*SD*	*N*	*M*	*SD*
		Feet	*Feet*		*Feet*	*Feet*
Time sample	60	60.73	49.17	60	38.73	27.46
Range	60	121.25	77.18	60	99.50	53.52
Limits	60	104.75	64.24	60	79.88	45.80

Heading in straight and level course

	1-2 legs			3-4 legs		
	N	*M*	*SD*	*N*	*M*	*SD*
		Degrees	*Degrees*		*Degrees*	*Degrees*
Time sample	60	7.82	7.10	60	6.22	5.45
Range	60	3.40	2.38	60	2.87	1.76
Limits	60	2.35	1.45	60	2.10	1.25

Table A8.2.—*Means and SD's of test and retest*—Continued

Supplement to table 8.2

Altitude in straight and level course

	1-2 legs			3-4 legs		
	N	*M*	*SD*	*N*	*M*	*SD*
		Feet	*Feet*		*Feet*	*Feet*
Time sample	60	225.53	133.21	60	178.53	99.07
Range	60	90.83	35.65	60	76.75	37.93
Limits	60	62.75	26.21	60	53.00	28.61

Table A8.3.—*Means and SD's of the objective measures for the two criterion groups*

Supplement to table 8.3

Low Group

	Range			Limit			Time sample		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Altitude deviation in steep turn (in feet)	30	258.00	122.48	30	211.83	107.31	30	114.76	64.11
Altitude deviation straight and level (in feet)	30	121.83	38.01	30	74.50	20.99	30	437.00	189.13
Heading straight and level (in degrees)	30	4.80	2.40	30	2.87	1.31	30	17.47	14.39
High Group									
Altitude in steep turn (in feet)	30	153.50	54.47	30	156.83	74.11	30	84.17	48.84
Altitude in straight and level (in feet)	30	97.67	[illegible]	30	63.00	32.83	30	371.90	209.84
Heading in straight and level (in degrees)	30	3.40	1.91	30	2.47	1.36	30	10.00	8.76

Table A8.4.—*Comparison of the total of four part scores with the single score based on the largest deviation during the straight and level course with changing air speeds*

A. Observer reliability (N=60)

	Range					Limits				
	Second observer		Check pilot		*r*	Second observer		Check pilot		*r*
	M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	
Heading:	(Degrees deviation)									
Total score	7.7	5.0	7.6	5.4	0.45	5.4	3.3	5.6	4.0	0.51
Single score	4.1	2.3	4.0	2.5	.56	2.7	1.4	2.8	1.8	.54
Altitude:	(Feet deviation)									
Total score	220	81	159	88	.65	147	59	114	62	.73
Single score	110	43	87	48	.69	69	28	59	32	.79

TABLE A8.4.—*Comparison of the total of four part scores with the single score based on the largest deviation during the straight and level course with changing air speeds*—Continued

B. Test-retest reliability

	1-2 legs		3-4 legs		r	1-2 legs		3-4 legs		r
	M	SD	M	SD		M	SD	M	SD	
Heading (degrees):										
Total score	4.2	3.0	3.7	2.5	0.82	3.0	1.9	3.0	2.0	0.90
Single score	3.4	2.4	2.9	1.8	.80	2.4	1.5	2.1	1.3	.75
Altitude (feet):										
Total score	123	50	98	47	.57	93	40	75	38	.58
Single score	91	36	77	33	.56	63	26	53	29	.61

C. Validity coefficients

	Low group		High group		r_{bis}	Low group		High group		r_{bis}
	$N=30$		$N=30$			$N=30$		$N=30$		
	M	SD	M	SD		M	SD	M	SD	
Heading (degrees):										
Total score	9.3	5.1	6.4	4.5	0.29	5.9	3.0	4.9	3.5	0.16
Single score	4.8	2.4	3.4	1.9	.31	2.9	1.3	2.5	1.3	.15
Altitude (feet):										
Total score	242	72	199	54	.26	158	51	137	65	.18
Single score	122	38	98	44	.23	75	21	63	33	.20

TABLE A8.5.—*TB-25 Instrument flight check*

(For research purposes only)

Date .... Time of flight ......

Name of student 1

(Last name first)

Name of instructor 2

Class .... Squadron .... Flight 3-4

Total TB-25 Time .... Inst. dual .... Inst. solo.... 5-6

A. INSTRUMENT TAKE-OFF: 7

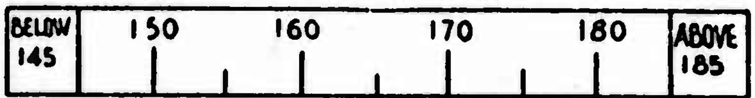
1. Did instructor help to keep aircraft on runway 8-9
during roll? ...... Yes ...... No 10

2. After take-off did aircraft 11-12
---- lose altitude
---- level out ---- climb too steeply 13
---- maintain steady climb 14

3. A/S range (grade from time 165 m. p. h. reached to 4,000 ft. Place mark at lowest *and* highest A/S):

B. CLIMB:

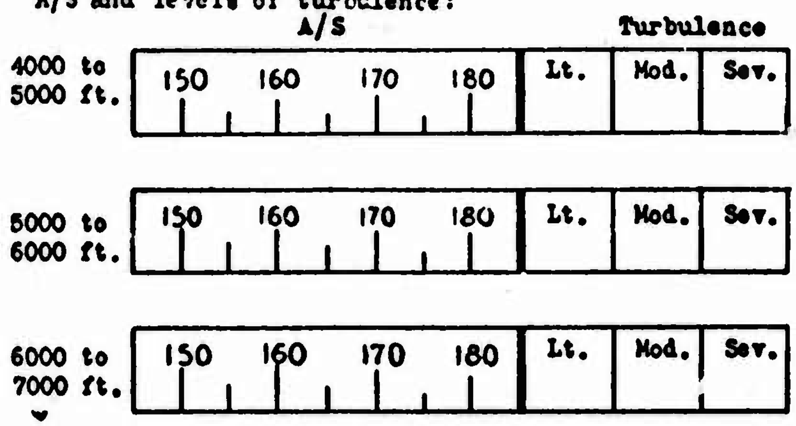
During climb to 7,000 ft. mark lowest *and* highest A/S and levels of turbulence:



16
17
18

B. CLIMB

During climb to 7000 ft. mark lowest and highest A/S and levels of turbulence:



19
20
21

22
23
24
25
26
27
28
29
30
31
32
33
34

35
36
37
38
39

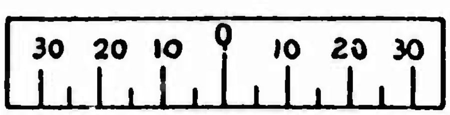
40
41
42
43
44

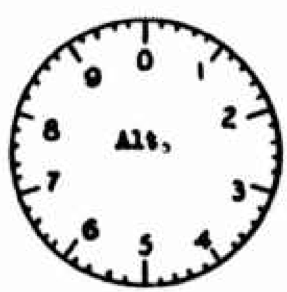
C. LEVEL OUT:

"Level out at .... ft. Maintain rollout heading and altitude."—Student deviated from "level out" heading and altitude by: (mark lowest *and* highest deviations). For one minute.

45
46

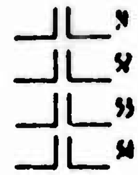
Heading





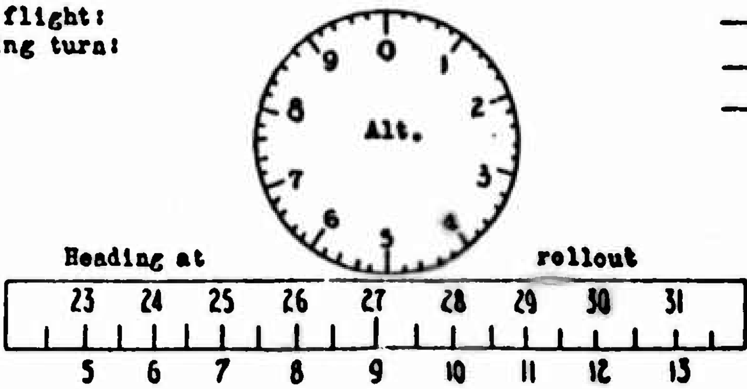
D. 180° TURN:

"Now do a 180° full panel standard rate turn to the (left) (right). Set D/G on 0°. Start with clock on any cardinal heading. Calibrate needle during this turn."—Mark deviations from start of turn to level flight:



level flight:
During turn:

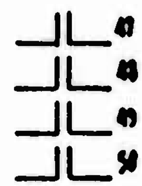
Time error from start of turn to start of rollout.
_____ sec.



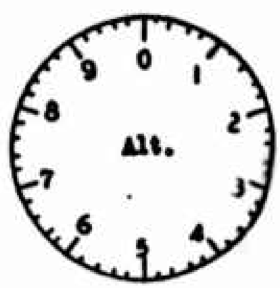
E. PARTIAL PANEL TURN:

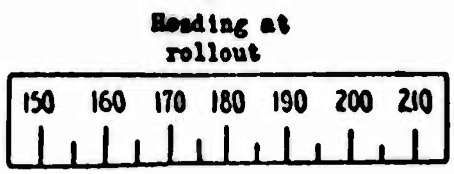
"Cage artificial horizon and D/G. Start on 0°. Now do a standard rate time turn to (right) (left) of 270°. Start time on any cardinal heading of the clock."—Mark deviations from start of turn to level flight:

During turn:



Time error from start of turn to start of rollout.
_____ sec.



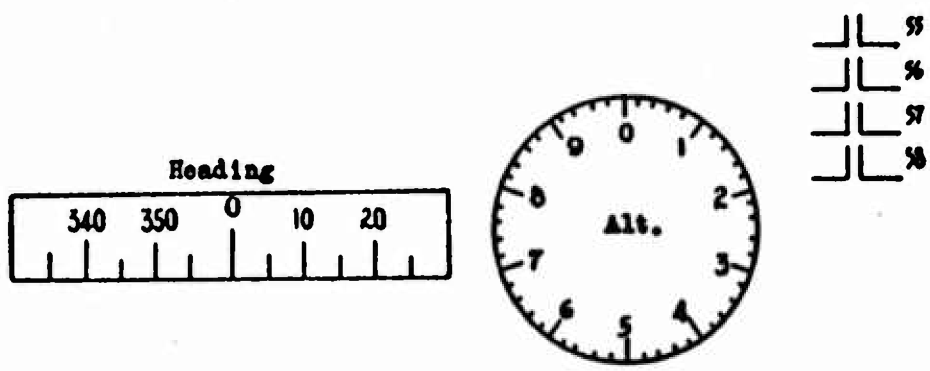


F. POWER LET-DOWN (BASIC) (RADIO):

"Fly at 180 m. p. h., 1,800 r. p. m., set D/G on 0°. Prepare for power letdown, GUMP, using 2200 r. p. m., gear down, 15° flaps and 160 m. p. h. Hold heading and altitude."

1. During change of A/S:

a. Heading and altitude deviations:



b. Did student complete GUMP check?

---------- Yes ---------- No

"Let down 500 feet from starting altitude and then level out. Hold heading and 160 m. p. h. after level out."

2. Level out (from first application of power): Heading, altitude and A/S deviations during the next minute:

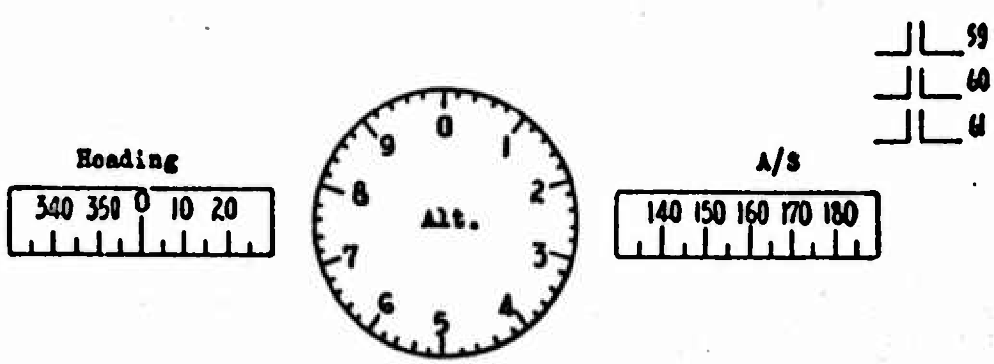
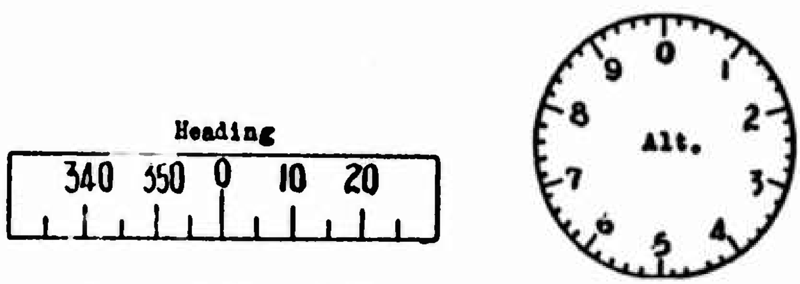


TABLE A8.5.—*TB-25 Instrument flight check*—Continued

G. SINGLE-ENGINE OPERATION:

"Fly at 1,800 r. p. m., 180 m. p. h. normal cruise. Set D/G on 0°. I am going to cut a throttle and you are to set up a single-engine procedure holding heading and altitude. Maintain safe single-engine A/S."—(Take all readings from throttle cut to 1 minute after).

62
63
64



Lowest A/S reached: ---------- m. p. h.

H. RADIO PROCEDURE:

Did student: 65

a. Tune in and identify proper station? 66

---------- Yes ---------- No 67

b. Check d/g against magnetic compass *and* reset? 68

---------- Yes ---------- No 69

During 1st minute, deviations while tuning radio equipment:

I. QUADRANT IDENTIFICATION:

Did student recognize fade or build correctly?

---------- Yes ---------- No 70

J. BEAM CROSSING:

After crossing through beam, did student start his turn in:

---------- 4 signals or less

---------- 8 signals or less 71

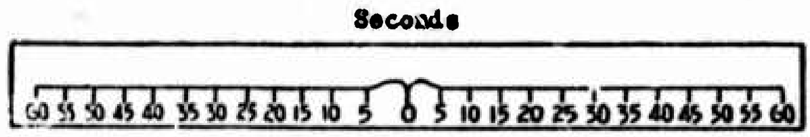
---------- 12 signals or less

---------- more than 12 signals

TABLE A8.5.—*TB-25 Instrument flight check*—Continued

K. STATION IDENTIFICATION:

Position when student signaled he passed station:



72
73

L. LET-DOWN AND LOW APPROACH:

a. Position when student signaled "over field":

.... within field boundary 74
.... within 1 mile of field 75
.... within 2 miles of field
.... over 2 miles from field

b. Time error from low cone to field boundary; .... sec.

M. VOICE PROCEDURE: 76

Did student use voice procedure correctly?

.... Yes .... No

79 77
80 78

TABLE A8.6.—*Mean, sigma and reliability of each objective measure in the TB-25 daily check ride*

(Based on scores of average odd and average even trials for double length)

	N	Means		Standard deviation		Reliability coefficient		P[2]
		Odd	Even	Odd	Even	Uncorrected[1]	Corrected[1]	
A. INSTRUMENT TAKE OFF:								
1. Instructor help keep on runway	68	2.00	1.87	0.82	0.84	0.03	0.06	0.40
2. Attitude in initial climb	70	3.19	3.22	1.39	1.34	.15	.26	.11
3. Air speed range in climb to 4,000 feet	72	3.29	3.27	.93	.86	.17	.29	.08
B. SPIRAL CLIMB:								
4. Air speed in climb 4,000-5,000 feet	74	3.50	3.35	.71	.70	−.12	−.21	......
5. Air speed in climb 5,000-6,000 feet	75	3.66	3.66	.71	.71	−.01	−.02	......
6. Air speed in climb 6,000-7,000 feet	75	3.75	3.91	.70	.63	.09	.17	.23
C. LEVEL OUT:								
7. Heading during level out	76	4.03	4.04	.71	.70	.24	.39	.02
8. Altitude during level out	76	3.57	3.62	.73	.78	.32	.48	.01
D. 180° TURN:								
9. Time error in 180° turn	77	2.44	2.52	.50	.46	.09	.17	.23
10. Altitude range in 180° turn	77	3.09	3.12	.77	.85	.47	.64	.01
11. Heading at roll-out of 180° turn	77	4.22	4.40	.77	.76	−.05	−.10	......
E. PARTIAL PANEL TURN:								
12. Time error in partial panel turn	77	2.79	2.75	.38	.33	.08	.15	.25
13. Altitude range in partial panel turn	75	2.64	2.84	.92	.80	.39	.56	.01
14. Heading at roll-out in partial panel turn	77	2.62	2.66	1.13	1.04	.08	.15	.25
F. POWER LET DOWN:								
15. Heading range during A/S change	71	3.41	3.50	1.00	.83	.18	.31	.07
16. Altitude range during A/S change	76	2.59	2.77	.88	.97	.14	.25	.12
17. Did student complete GUMP check?	76	2.56	2.39	.56	.72	.15	.26	.11
18. Heading range at level out	71	4.05	3.99	.73	.63	.22	.36	.03
19. Altitude range at level out	76	3.84	3.88	.72	.80	.32	.48	.01
20. A/S range at level out	76	3.58	3.69	.75	.69	.13	.23	.14

See footnotes at end of table.

TABLE A8.6.—*Mean, sigma and reliability of each objective measure in the TB-25 daily check ride*—Continued

	N	Means		Standard deviation		Reliability coefficient		P[2]
		Odd	Even	Odd	Even	Uncorrected[1]	Corrected[1]	
G. SINGLE-ENGINE OPERATION:								
21. Heading range during S. E. operation	75	3.01	2.95	.90	.87	.12	.21	.16
22. Altitude range during S. E. operation	75	2.60	2.43	1.08	1.04	.20	.33	.05
23. Lowest A/S reached during S. E. operation	72	2.36	2.34	.78	.71	.18	.31	.07
H. RADIO PROCEDURE:								
24. Tune in and identify proper station	67	2.96	2.94	.27	.28	−.03	−.06	......
25. Check directional gyro	66	2.50	2.26	.73	.81	.39	.56	.01
26. Heading range while tuning radio	61	3.96	3.99	.74	.72	−.03	−.06	......
27. Altitude range while tuning radio	65	3.45	3.72	.97	.93	.21	.35	.05
28. A/S range while tuning radio	60	3.89	3.91	.92	.88	.09	.17	.25
I. QUADRANT IDENTIFICATION:								
29. Recognize fade or build	64	2.75	2.81	.68	.46	.33	.50	.01
J. BEAM CROSSING:								
30. When started turn after crossing beam	66	3.61	3.62	.56	.56	−.15	−.26	......
K. STATION RECOGNITION:								
31. Error in seconds when crossed cone	57	4.01	3.91	1.11	1.34	.23	.37	.04
L. LET-DOWN AND LOW APPROACH:								
32. Position when signaled over field in let-down	27	4.85	4.48	.52	.88	−.17	−.29	......
33. Time error of let-down from cone to field	26	2.10	1.88	.59	.32	.06	.11	.38
M. VOICE PROCEDURE:								
34. Use proper voice procedure	16	2.31	2.31	.92	.92	.12	.12	.33

[1] Uncorrected reliability coefficients are calculated from the average of odd-day ride scores and the average of even-day ride scores. The actual test was a combination of both, or twice as long as the tests represented by each of these averages. The corrected reliability coefficients allow for the increased reliability of the full-length test, as compared with the half-length test from which the uncorrected coefficients were derived.

[2] The probability of obtaining by chance positive, uncorrected correlations of the size observed.

TABLE A8.7.—*Semi-partial correlations to correct for differences in amount of total TB-25 flying time on the odd v. the even days*

Measure	Uncorrected odd-even reliability	Semi-partial correlation
9. Time error in 180° turn	0.09	0.12
13. Altitude range in partial panel turn	.30	.33
15. Heading range during let-down	.18	.16

$$r_{(1\cdot3)(2\cdot4)}=\frac{r_{12}-r_{24}r_{14}-r_{23}r_{13}+r_{13}r_{24}r_{34}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{24}}}$$ [1]

r_{12}=odd score, even score.
r_{14}=odd score, even hours.
r_{24}=even score, even hours.
r_{13}=odd score, odd hours.
r_{23}=even score, odd hours.
r_{34}=odd hours, even hours.

[1] Guilford, J. P., *Psychometric Methods*, p. 404, McGraw-Hill Book Co., Inc., N. Y., 1936.

TABLE A8.8.—*TB-25 Instrument flight check (AAF 50-3)* [1]

Date.......... Squadron.......... Flight..........
Name of student..........
Name of check pilot..........
Total TB-25 time.......... Instrument dual..........
Instrument solo..........

INSTRUCTIONS: Score each measure as it occurs. If the student fails any one or more maneuvers, he has not passed this 50-3 check ride. Appropriate remarks will be made in these cases. Determine the total score for those students that pass all maneuvers in the check ride.

SUMMARY OF STUDENT'S RIDE: The numerical score is the sum of all measures. The radio range scores are added together and multiplied by two thus giving the basic and radio phases approximately equal weight in the total score. The check pilot should write a short descriptive statement pointing out the strong and the weak aspects of the student's instrument flying skill as demonstrated on this ride. Frequent reference can be made to the maneuvers (indicated by letters A through S).

SUM OF SCORES

Basic instruments:	Radio instruments:
Page 2	
Page 3	Page 6 ....
Page 4	Page 7 ....
Page 5	
Total	Total ....×2=..........
	Grand total .... Letter grade ....

A	B	C	D	E
Superior (over 500)	Excellent (400-500)	Very satisfactory (300-399)	Satisfactory (200-299)	Fail (below 200 or fail one or more maneuvers)

Comments: ....
....
....

A. COCKPIT CHECK: 3. complete. 2. incomplete.
1. omitted.

B. INSTRUMENT TAKE-OFF:

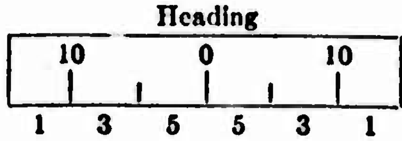
Directional control during roll:
5. Satisfactory.
3. Needed a little help.
1. Instructor took over.

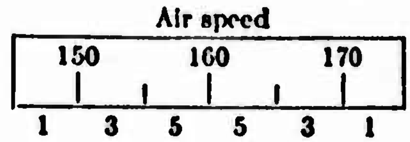
Take-off attitude:
3. Normal.
1. Nose high.
1. Nose low.

[1] The specific conditions used in the check represent those used at one Advanced two-engine school (Enid Army Air Field, Enid, Okla).

TABLE A8.8.—*TB-25 Instrument flight check (AAF 50-3)*—Continued

Deviations after 160 miles per hour first reached and until start of spiral climb:

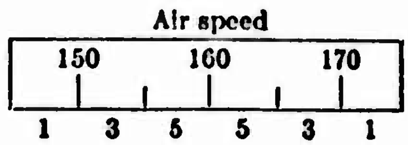


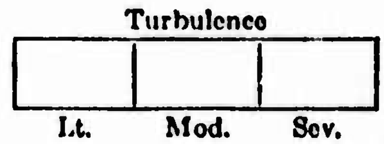


---- Pass.
---- Fail.

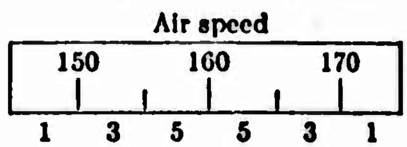
C. SPIRAL CLIMB: Reverse direction of turn every 500 feet during the first 2,000 feet of climb.

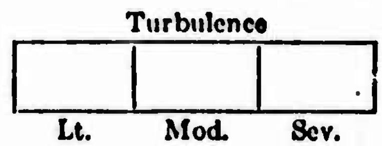
First 1,000 feet of climb:





Second 1,000 feet of climb:





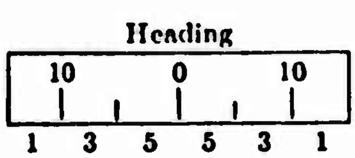
During entire climb to level out:

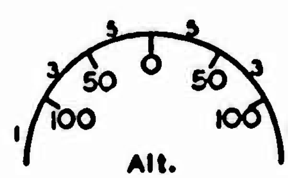
Trim control --------: 3. Satisfactory. 1. Unsatisfactory.
Control of power ----: 3. Satisfactory. 1. Unsatisfactory.
Technique at the controls:
5. Generally smooth.
3. Sometimes smooth, sometimes rough.
1. Generally rough.

---- Pass.
---- Fail.

Sum of scores on this page (2) ----------
Comments: --

D. LEVEL OUT: "Level out at ---- feet. Maintain level out altitude and heading of ----°". Grade for 2 minutes after start of level out.



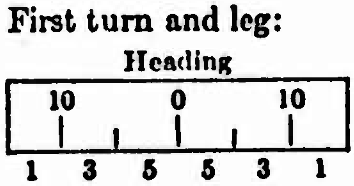


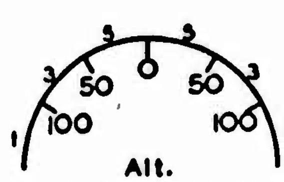
Trim control:
3. Satisfactory.
1. Unsatisfactory.

------ Pass.
------ Fail.

E. PATTERN A (restricted panel):

First turn and leg:



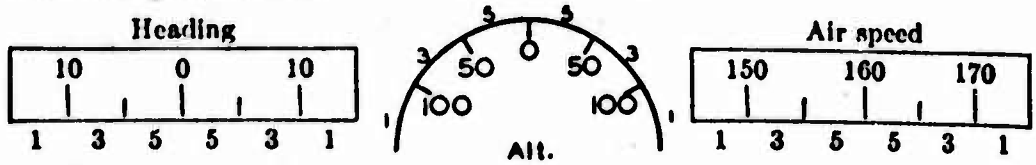


Score limits:
Time error in turn:
3 for: Within 3 seconds.
1 for: Over 3 seconds.
Roll out heading:
5 for: Within 5°.
3 for: Within 10°.
1 for: Over 10°.

Time error: 3 1 Roll out error: 5 3 1

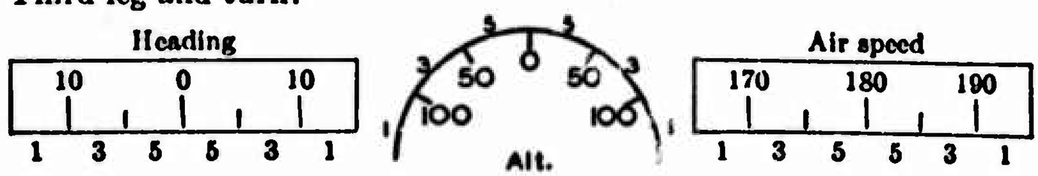
TABLE A8.8.—*TB-25 Instrument flight check (AAF 50-3)*—Continued

Second leg and turn:



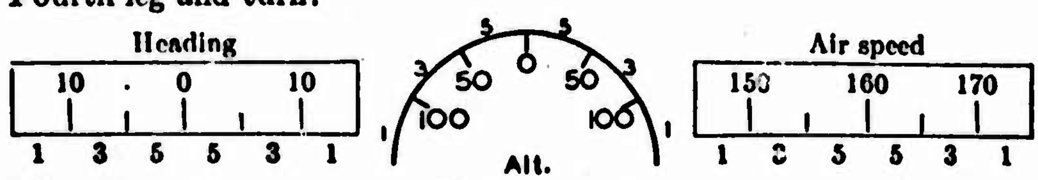
Time error: 3 1 Heading at roll out: 5 3 1

Third leg and turn:



Time error: 3 1 Roll out heading error: 5 3 1

Fourth leg and turn:



Time error: 3 1 Roll out heading error: 5 3 1

Throughout pattern: Technique — 3. Satisfactory. 1. Unsatisfactory.
Cross-checking: 3. Satisfactory. 1. Unsatisfactory

Sum of scores on this page (3) ------ ---- Pass. Fail ----

F. PATTERN B:

	Scores 5	3	1
Heading..........	5°.	10°.	Over 10°.
Altitude..........	50 feet.	100 feet.	Over 100 feet.
Airspeed..........	5 miles per hour.	10 miles per hour.	Over 10 miles per hour.
Time error........	5 seconds.	10 seconds.	Over 10 seconds.

First leg:	Heading 5 3 1	Altitude 5 3 1	Air speed 5 3 1	
First turn:	Altitude 5 3 1	Air speed 5 3 1	Roll out 5 3 1	Time error 5 3 1
Second leg:	Heading 5 3 1	Altitude 5 3 1	Air speed 5 3 1	
Second turn:	Altitude 5 3 1	Air speed 5 3 1	Roll out 5 3 1	Time error 5 3 1
Third leg:	Heading 5 3 1	Altitude 5 3 1	Air speed 5 3 1	
Third turn:	Altitude 5 3 1	Air speed 5 3 1	Roll out 5 3 1	Time error 5 3 1

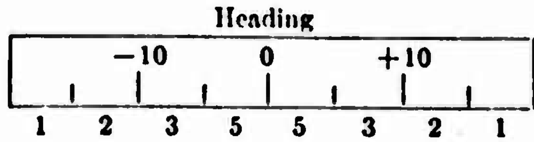
TABLE A8.8.—*TB-25 Instrument flight check (AAF 50-3)*—Continued

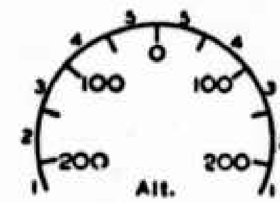
Fourth leg:	Heading	Altitude	Air speed	
	5 3 1	5 3 1	5 3 1	
Fourth turn:	Altitude	Air speed	Roll out	Time error
	5 3 1	5 3 1	5 3 1	5 3 1

Control technique:

3. Generally smooth.
2. Sometimes smooth, sometimes rough. .... Pass.
1. Generally rough. .... Fail.

G. SINGLE-ENGINE OPERATION:





Maintain safe single-engine air speed: 5. Yes. 1. No.

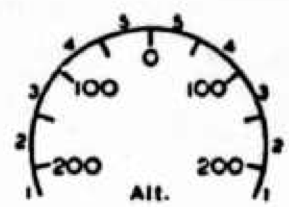
Procedure:

5. Correct and safe.
3. Correct but hesitant.
2. Incorrect but safe.
1. Incorrect and possibly dangerous.

.... Pass.

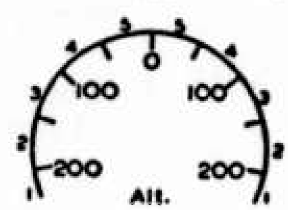
Sum of scores on this page (4) .... Fail.

H. STEEP TURNS:



Maintained constant angle of bank.......:	3. Yes.	1. No.
Held ball in center......:	2. Yes.	1. No.
Control technique......:	3. Smooth.	1. Rough.

Restricted panel:



Maintained constant angle of bank......:	3. Yes.	1. No.
Held ball in center...:	3. Yes.	1. No.
Control technique...:	3. Smooth.	1. Rough.

.... Pass.
.... Fail.

I. STALLS (write in type of stall ....):

Maintained constant approach..:	3. Yes.	1. No.	
Recognized stall condition.......:	3. Satisfactory.	1. Poor.	

RECOVERY:

Initial recovery attitude..:	3. Normal.	1. Nose low.	1. Nose high.
Applied power............:	3. At about right time.		1. Too soon.
Amount of altitude lost...:	3. Normal.	1. Excessive.	
Control of plane was......:	3. Smooth.	1. Rough.	

.... Pass.
.... Fail.

TABLE A8.8.—*TB-25 Instrument flight check (AAF 50-3)*—Continued

J. UNUSUAL POSITION RECOVERY:

Start of recovery was.....: 3. Immediate. 1. Hesitant.
First recovery action was..: 3. Correct. 1. Incorrect.
In general, recovery was...:
- 3. Smooth and safe.
- 2. Rough but safe.
- 1. Dangerous but recovered without help.
- 1. Instructor had to assist recovery.

---- Pass.
Sum of scores on this page (5) .... ---- Fail.

Comments: ....
....
....
....

Radio check

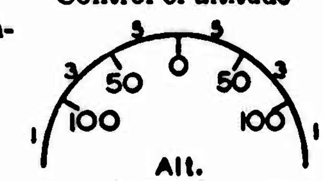
K. PROCEDURES:

Properly tuned in desired station................: 5. Yes. 1. No.
Check D/G against magnetic and reset...........: 5. Yes. 1. No.
Turn to correct bisector heading (10°)...........: 5. 5°.
3. 10°.
1. Over 10°.

Recognize fade or build:
- 5. Within 5 minutes.
- 3. Within 10 minutes.
- 1. Over 10 minutes or done incorrectly.

Call for orientation clearance, weather, and altimeter setting:
5. Yes. 1. No.

Control of altitude



Remarks: .... ---- Pass.
.... ---- Fail.

L. ORIENTATION:

Recognized change of signal...........: 5. Satisfactory.
1. Unsatisfactory.

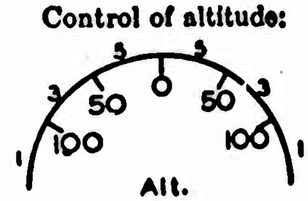
Procedure:
- 5. Selected best system and did it correctly.
- 3. Selected less desirable system but did it correctly.
- 1. Failed to properly orient himself.

Remarks: .... ---- Pass.
.... ---- Fail.

M. BEAM FOLLOWING:

Check D/G and reset........: 5. Yes. 1. No.
Make necessary corrections to follow beam.................: 5. Yes. 1. No.
Station recognition............: 5. Satisfactory.
1. Unsatisfactory.
---- Signal interference.

Control of altitude:



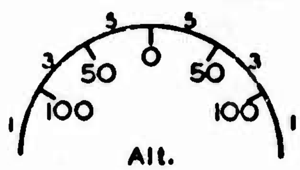
Remarks: .... ---- Pass.
.... ---- Fail.

TABLE A8.8.—*TB-25 Instrument flight check (AAF 50-3)*—Continued

N. HOLDING:

Make necessary corrections to follow beam..: 5. Yes. 1. No.
Direction of procedure turn..............: 5. Correct.
1. Incorrect.

Control of altitude:



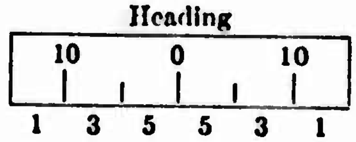
Remarks: .... Pass.
.... Fail.

Sum of scores on this page (6) ....

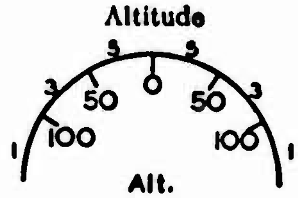
O. LET-DOWN: During change of air speed, did student

Change props to 2,200 r. p. m....................: 3. Yes. 1. No.
Lower gear (under 170 m. p. h.)..................: 3. Yes. 1. No.
Complete GUMP check..............................: 3. Yes. 1. No.
Make radio call..................................: 3. Yes. 1. No.

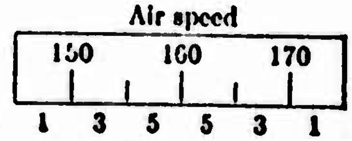
Heading



Altitude



Air speed

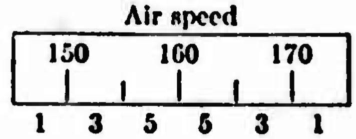


Remarks: . ... Pass.
.... Fail.

P. PROCEDURE TURN:

Time from station: 5. Correct. 1. Incorrect.
Was procedure turn done properly: 5. Yes. 1. No.

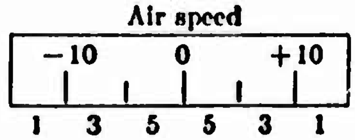
Air speed



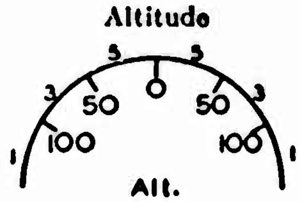
Remarks: .... Pass.
.... Fail.

Q. INBOUND TO LOW CONE:

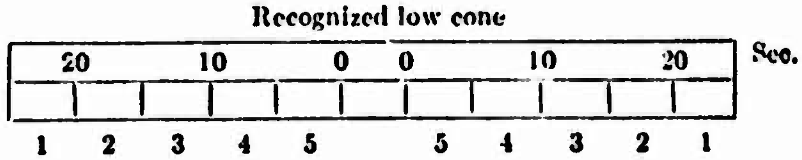
Air speed



Altitude



Recognized low cone



Remarks: .... Pass.
.... Fail.

TABLE A8.8.—*TB-25 Instrument flight check (AAF 50-3)*—Continued

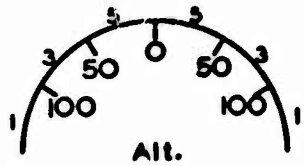
R. LOW APPROACH:

Air speed

−10		0		+10	
1	3	5	5	3	1

Time error:
5. Within 5 seconds.
3. Within 10 seconds.
1. Over 10 seconds.

Altitude when signaled over the field:



Position over the field:
5. Very accurate approach.
3. Safe distance.
1. Too far for safety.

Remarks: .... ---- Pass.
.... ---- Fail.

S. EMERGENCY PULL-UP:

Add power..... ......: 5. Yes. 1. No.
Retract gear...........: 5. Yes. 1. No.
Heading and A/S deviations for 1 minute after adding power:

Heading

10		0		10	
1	3	5	5	3	1

Air speed

150		160		170	
1	3	5	5	3	1

Sum of scores on this page (7) ....

TABLE A8.9.—*Inventory of objective measures of TB-24 instrument flying skill*

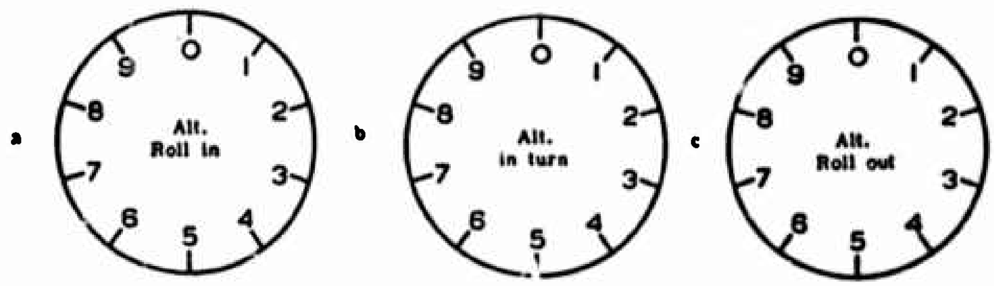
Inventory: TB-24 instrument scale	Number of measures	Inventory: TB-24 instrument scale	Number of measures
A. Instrument take-off	5	M. Radio compass (position report, tracking, boxing, intersecting bearing, time check)	15
B. Spiral climb	6	N. Close in procedure	5
C. Level out on cruising altitude	3	O. Outbound on final approach	7
D. Straight and level, partial panel	4	P. Procedure turn	5
E. Engine out, partial panel	10	Q. Inbound to low cone	4
F. Time turns: 360°	10	R. Low approach	4
G. Steep turns: 360°	12	S. Emergency pull-up	5
H. Partial panel turn: 270°	10	T. Unknown beam bracketing	9
I. Slow flying	14		
J. Power on stall	4		
K. Diving spiral	3		
L. True fade	4	Total	139

TABLE A8.10.—*Two sample instrument maneuvers for the TB-24*

H. PARTIAL PANEL TURN: 270°:

"Cage both gyros. Do a partial panel standard rate time turn of 270° to the (right) (left). Start the turn from any cardinal heading of the magnetic compass, and the time on any cardinal heading of the clock."

1. *Altitude deviations:* Grade from start of turn to level flight.



a b c

2. *Heading deviation:* At roll out.

Over 20	10	0	10	Over 20

3. *Did student:* Use correct calibrated time for this amount of turn? Yes ---- (3). No ---- (1).

M. RADIO COMPASS:

4. *Time check:* (Complete instructions before reaching wing tip position on R/C) "As soon as the R/C is exactly on wing tip position, turn inbound on the *bearing* of (33°) (123°) (213°) (303°).

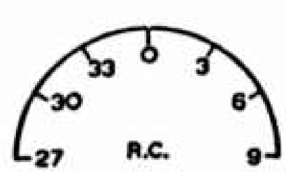
a. Did student signal time within ±15 seconds of instructor's time? Yes ---- (3). No ---- (1).

b. Instructor's time to station: ---- minutes ---- seconds.

5. *Heading and R/C deviations:* Grade after in-bound turn is completed.

Over 20	10	0	10	Over 20

Heading



a b

Objective Measures of Single-Engine Instrument Flying Skill at the Basic Level of Flying Training

BASIC INSTRUMENT FLIGHT CHECK (AT-6)

For Research Purposes Only

(Form #2)

Date .... Take-Off Time .... ______1
Student .... ASN .... ______2, 3 (1, 2)
Class .... Squadron .... Group .... ______4
Total AT-6 Time .... Inst. Dual .... Inst. Solo .... ______5, 6
Regular Instr. Instructor .... ______7, 8
 ______9
Check Pilot .... ______10, 11
Stanines: B .... N .... BP .... FP .... FE .... ______12, 13
Quartile Rating .... Rank Order .... ______14
 ______15
 ______16
 ______17
 ______18
 ______19
 ______20

This test is being administered for research purposes only. Scores on this booklet will not affect the final instrument grade nor will it influence future assignments. *All test scores will be handled as a group and no comparisons will be made between individual students or individual instructors.*

To THE STUDENT: You are familiar with all of the maneuvers in this booklet. Your score will depend upon how well you control heading, altitude, air speed, and time during each maneuver.

Point score	5	4	3	2	1
Heading range	5°	10	15	20	Over 20.
A/S range	5 m.p.h	10	15	20	Over 20.
Altitude range	50 ft	100	150	200	Over 200.
Time range	3 sec	6	9	12	Over 12.

Your point scores for altitude, heading, etc., in each maneuver will be combined as shown in the following table to give you a single score for each maneuver. In order to make your best score, you must carefully control all parts of each maneuver, as a single low score will give you a low score for the maneuver.

		Heading					
		5	*4*	*3*	*2*	*1*	
	5	5	5	4	3	2	
	4	5	4	3	3	2	*Maneuver*
Altitude	*3*	4	3	3	2	1	*Scores*
	2	3	3	2	2	1	
	1	2	2	1	1	1	

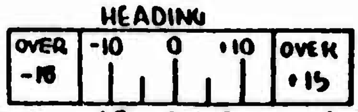
To the Check Pilot: The success of the experiment will depend upon how carefully the booklets are scored. Mark the student's deviations as carefully and immediately as you can. If you miss a measure, do not guess. Cross it out and go on to the next.

Keep a Sharp Lookout for Other Aircraft!

A. Instrument Take-Off: (Grade from start of roll to 700 feet indicated).

1. Heading Deviation During Roll. —1,2

HEADING

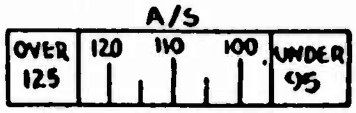


—3

—4

2. A/S Range. (Grade from time 110 mph first reached to 700 ft.) —5

A/S

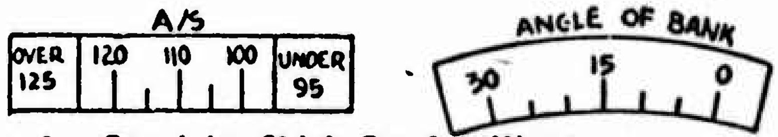


☐ 21

3. Did student make take-off without *ANY* help from instructor?
(3) ---- yes (1) ---- no.

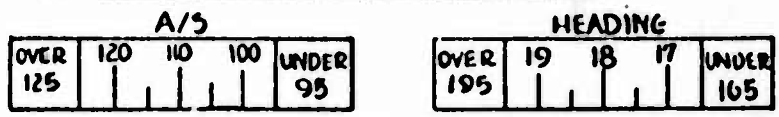
(At 700 feet have student break traffic and climb to 2,500 feet before continuing check.)

B. Climbing Turns: (At 2,500 feet read:) "Climb straight ahead and set the D/G on 0°. At 3,000 feet begin a climbing turn to the right holding 15° of bank and 110 m. p. h. Roll out on a heading of 180° and continue climb straight ahead for one minute."



____ 6
____ 7
☐ 22

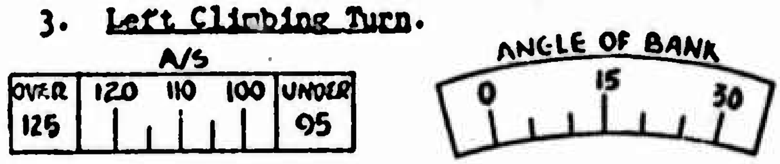
2. Straight Climb For One Minute.



____ 8
____ 9

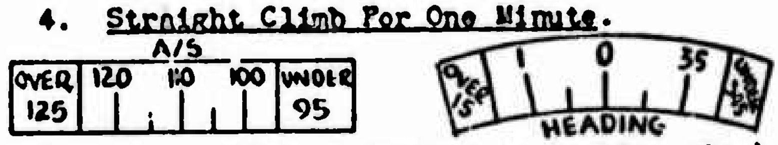
"NOW BEGIN A CLIMBING TURN TO THE LEFT HOLDING 15° OF BANK AND 110 MPH. ROLLOUT ON A HEADING OF 0° AND CONTINUE CLIMB STRAIGHT AHEAD FOR ONE MINUTE." ☐ 23

3. Left Climbing Turn.



____ 10
____ 11
☐ 24

4. Straight Climb For One Minute.

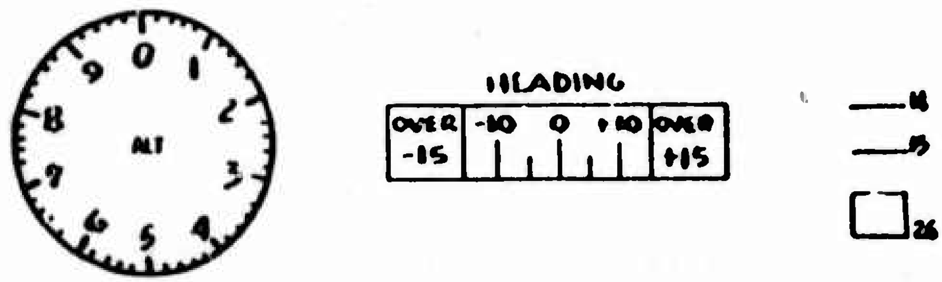


____ 12
____ 13

(continue climb as desired to working alt.)

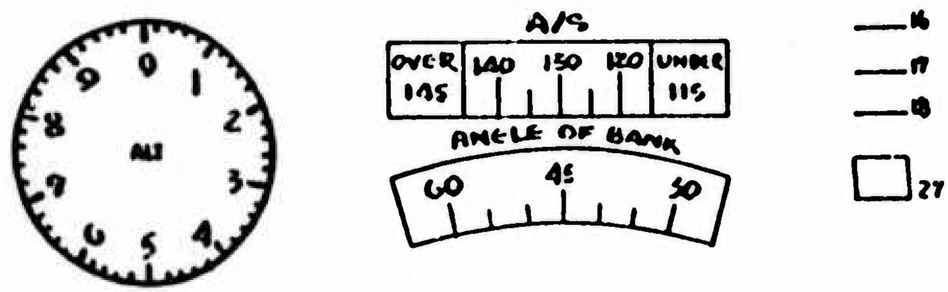
☐ 25

C. Level Out: "Level out at ---- ft. Establish 130 m. p. h. at 1,800 r. p. m." (Grade from start of level out for one minute.)



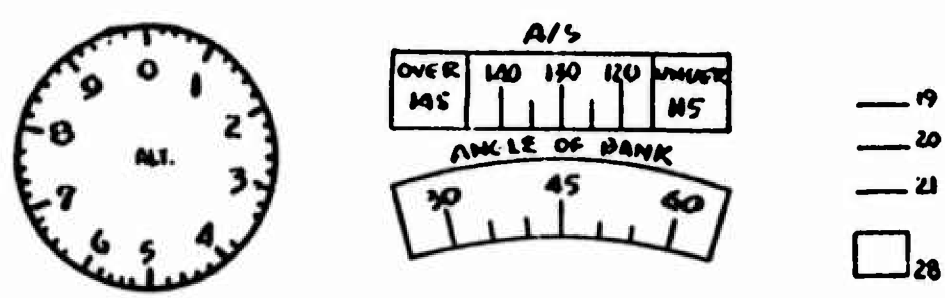
D. Steep Turns: "Set D/G on 0°. Now do a steep turn to the right with 45° of bank. Roll out on a heading of 180°. Use power if needed to hold your air speed." (Grade from start of turn to level flight.)

1. *Right Steep Turn.*



"Turn to a heading of 180°. Now do a steep turn to the left with 45° of bank. Roll out on a heading of 0°."

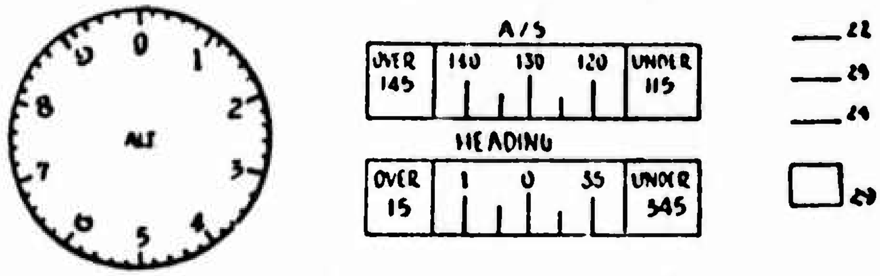
2. *Left Steep Turn.*



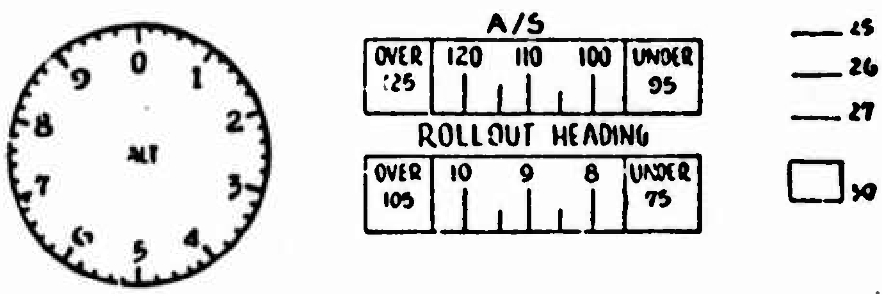
E. Pattern B: "Set D/G on 0°. Now do a pattern B. Shake the stick when you start." (Grade each leg and turn separately. Where change of air speed is required, continue grading altitude and heading, but allow 30 seconds before grading air speed. If student turns in wrong direction or fails to complete pattern, stop grading, and X out remaining measures.)

PATTERN 'B'

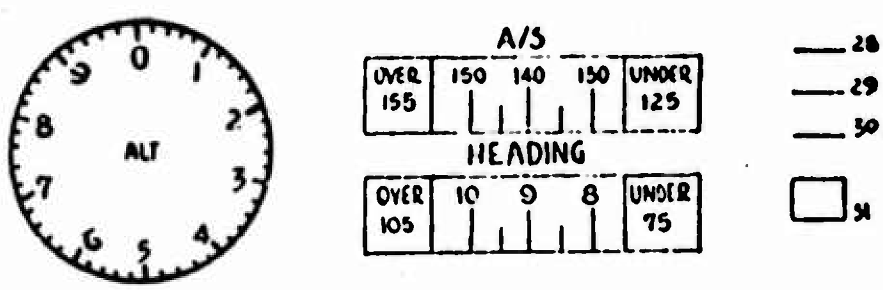
1st Leg



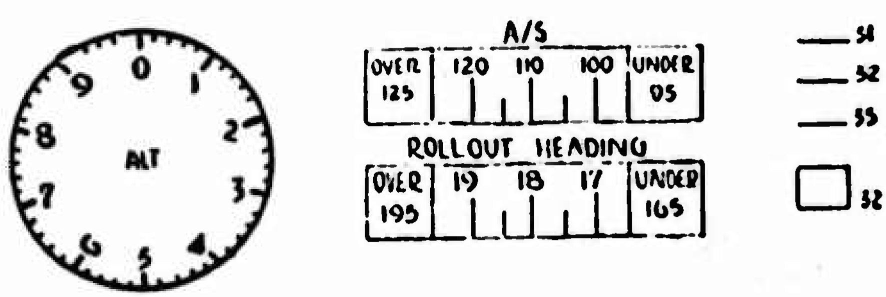
1st Turn



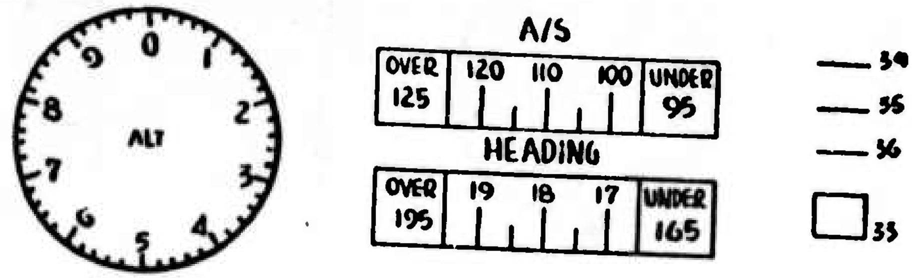
2nd Leg



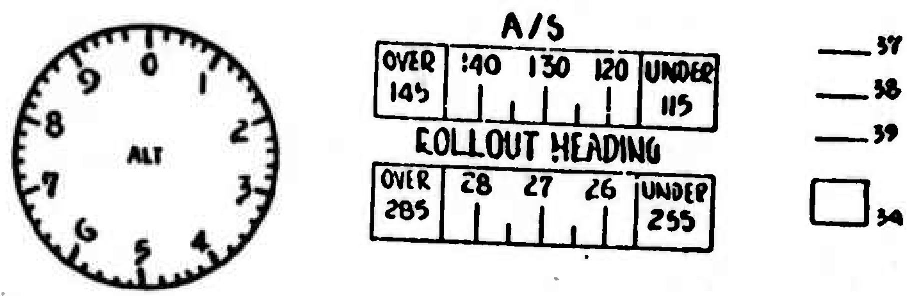
2nd Turn



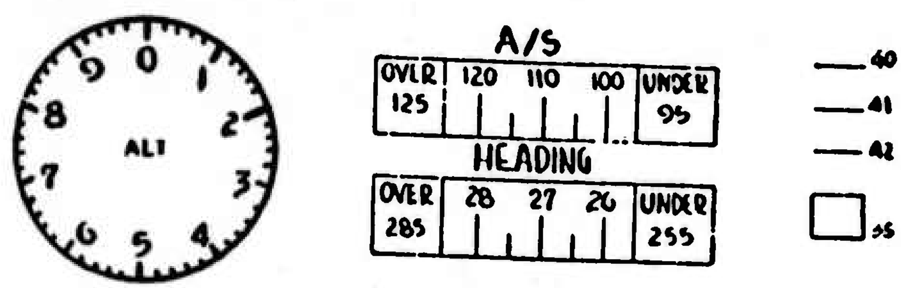
3rd Leg



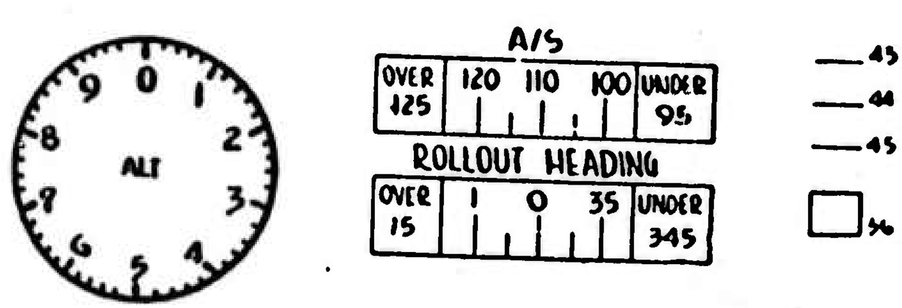
3rd Turn



4th Leg

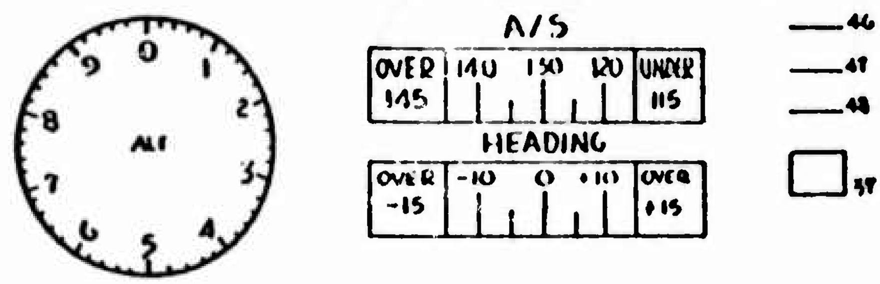


4th Turn

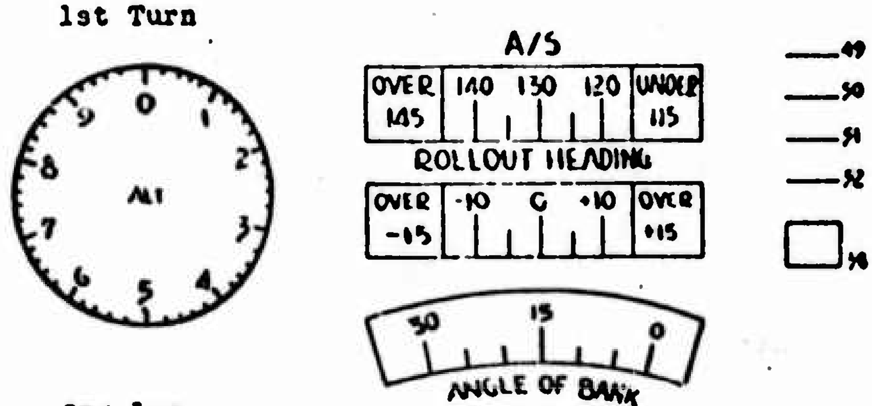


F. Pattern A: "Cage both your gyros. Now do pattern A. Make the first two turns to the right and the next two turns to the left. Fly one minute legs. Signal when you start."

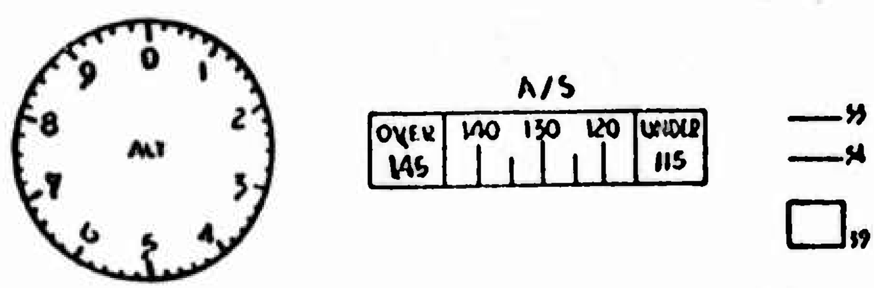
1st Leg



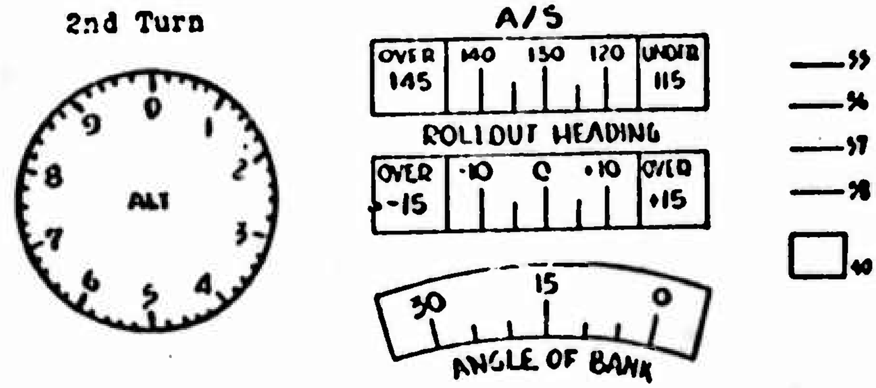
1st Turn



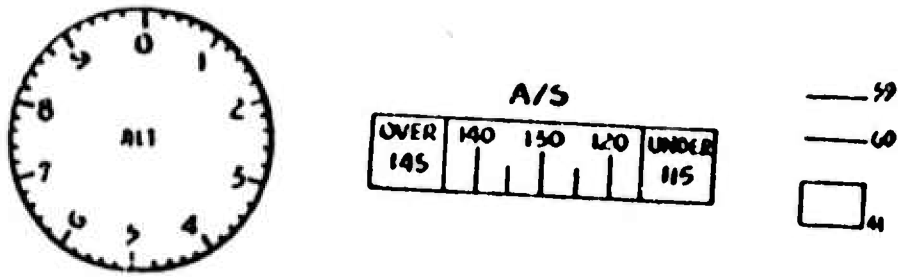
2nd Leg



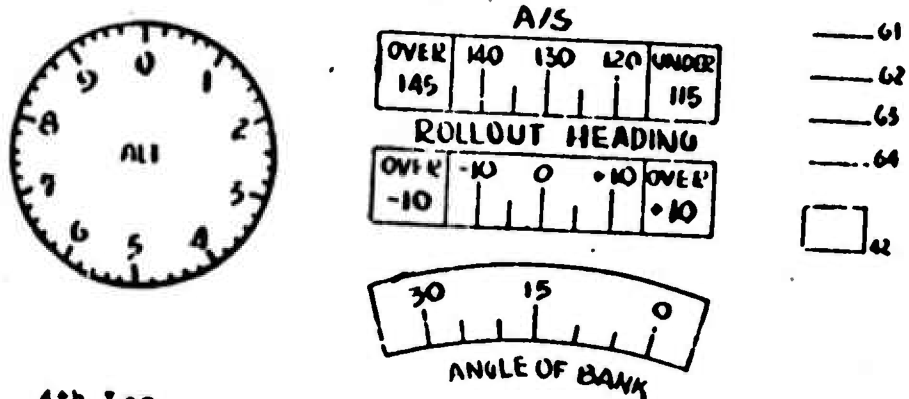
2nd Turn



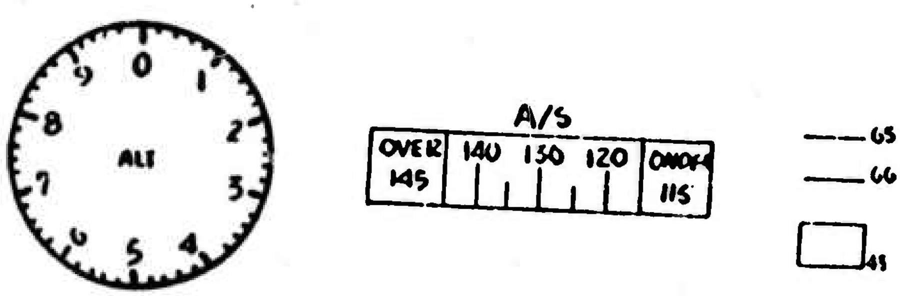
3rd Leg



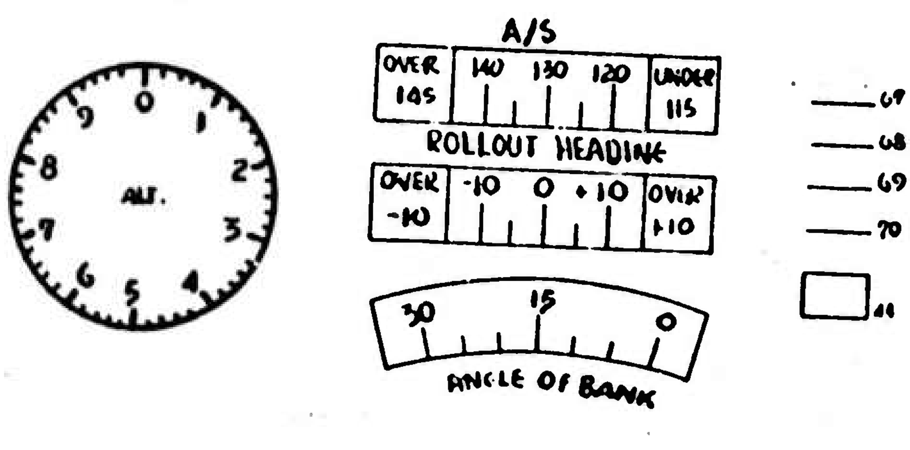
3rd Turn



4th Leg



4th Turn

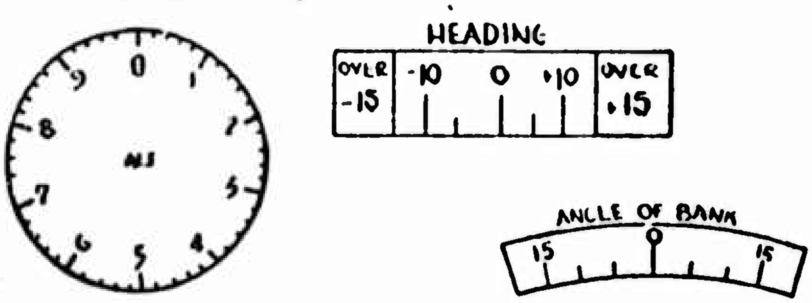


G. Unusual Position Recoveries: "I am going to give you unusual position recoveries. You are to take over and recover to straight and level flight when I shake the stick. (Give the following in any order.)

Grade by Letter

1. Top of wing over............	*a.* Smooth and safe.
2. Power-on stall...............	*b.* Rough but safe. ☐45
3. Power-off stall...............	*c.* Dangerous, but recovered without help.
4. Diving spiral................	*d.* Instructor had to take over.

H. Icing Stall: "I will reduce power and simulate an icing stall. Hold your altitude and heading until the stall and then make a normal recovery."

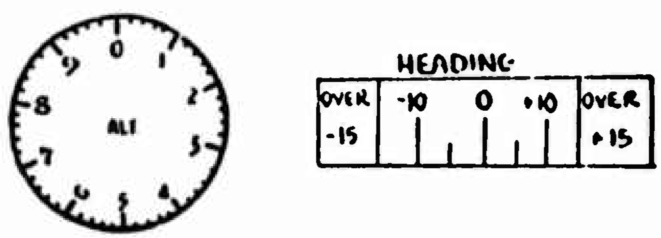


___71
___72
___73
___74
☐46

1. Highest air speed in recovery ....

I. Power Let-down: "Uncage the D/G only and set it on 0°. Establish a power let-down. Hold your altitude and heading until you lower the gear. Then set up 500 feet/minute descent."

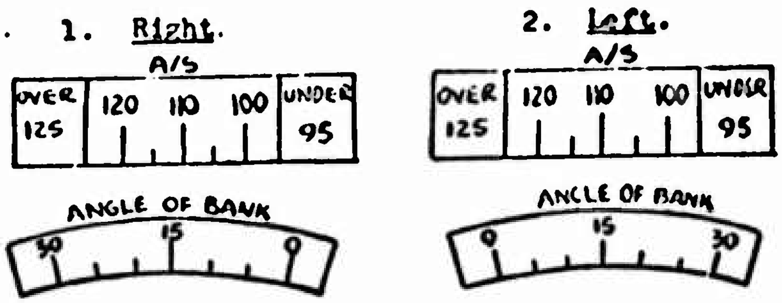
1. *While establishing let-down.*



___75
___76
☐47

J. Descending Turns: "Now do a 180° let-down turn to the right. Upon completion of turn do a 180° turn to the left."

___77
___78



☐48
___79
___80
☐49

Table A9.1.—*Reliabilities of coded item scores in AT-6 instrument scale*

Based on test and retest administered to 60 students on successive days in different planes by different instructors

Item	r_{II}	M_1	M_2	*SD 1*	*SD 2*
INSTRUMENT TAKE-OFF:					
Heading control during roll	0.18	3.91	4.03	0.99	0.96
A/S control initial climb	.11	2.97	3.88	1.21	.92
Inst. help during take-off	[2] −.03	------	------	------	------
RIGHT CLIMBING TURN:					
Air speed control	.22	3.49	4.17	1.10	.72
Angle of bank control	.02	3.66	4.42	.93	.79
STRAIGHT CLIMB:					
Air speed control	[1] .29	3.87	4.40	.67	.74
Heading control	[1] .31	4.17	4.32	.83	.72
LEFT CLIMBING TURN:					
Air speed control	[1] .35	3.92	4.40	.78	.61
Angle of bank control	.04	4.03	4.40	.84	.64
STRAIGHT CLIMB:					
Air speed control	[1] .25	4.22	4.44	.63	.66
Heading control	.00	4.61	4.49	.65	.68
LEVEL OUT:					
Altitude control	[1] .25	3.91	3.91	.95	.84
Heading control	.19	3.90	4.12	.92	.97
RIGHT STEEP TURN:					
Altitude control	.23	3.33	3.43	1.08	1.10
Air speed control	.06	3.71	3.93	.98	.89
Angle of bank control	.00	3.86	4.43	.94	.79
LEFT STEEP TURN:					
Altitude control	[1] .28	3.73	3.63	.87	.82
Air speed control	.00	3.56	3.54	.85	.93
Angle of bank control	[1] .29	3.93	4.29	1.04	.80
PATTERN "A"					
First leg:					
Altitude	.24	3.97	4.38	.96	.67
Air speed	−.05	4.18	4.65	.58	.66
Heading	.10	3.55	3.91	1.28	1.17
Second leg:					
Altitude	.15	3.75	4.30	1.17	.82
Air speed	[1] .27	3.93	4.41	1.08	.65
Third leg:					
Altitude	.22	3.47	4.12	1.32	.89
Air speed	[1] .29	3.67	4.42	1.34	.87
Fourth leg:					
Altitude	[1] .33	3.44	4.11	1.36	.91
Air speed	.20	3.76	4.33	1.22	.92
First turn:					
Altitude	[1] .36	3.48	4.00	1.12	.89
Air speed	.02	3.74	4.21	1.17	.78
Heading	−.07	3.31	3.49	1.58	1.52
Angle of bank	.01	3.62	4.18	1.22	.91
Second turn:					
Altitude	[1] .34	3.31	4.02	1.33	.98
Air speed	.00	3.78	4.22	1.26	.74
Heading	−.19	3.20	3.61	1.14	1.41
Angle of bank	−.02	3.86	4.23	1.87	.91
Third turn:					
Altitude	.19	3.35	4.12	1.36	.89
Air speed	.09	3.80	4.32	1.16	.78
Heading	.07	3.11	3.32	1.62	1.49
Angle of bank	.14	3.72	4.16	1.20	.89
Fourth turn:					
Altitude	.09	3.35	4.14	1.32	.91
Air speed	.22	3.70	4.25	1.36	.91
Heading	−.06	3.02	3.47	1.67	1.43
Angle of bank	.07	3.61	4.17	1.24	.92
PATTERN "B"					
First leg:					
Altitude	[1] .34	4.22	4.72	.78	.49
Air speed	[1] .40	4.31	4.61	.87	.49
Heading	−.05	4.53	4.75	.79	.47
Second leg:					
Altitude	−.09	3.63	3.83	.95	.90
Air speed	[1] −.32	3.69	4.10	.92	.89
Heading	.20	4.50	4.38	.58	.70
Third leg:					
Altitude	.05	3.88	4.15	1.04	.92
Air speed	.05	4.10	4.05	.92	.96
Heading	[1] .25	4.23	4.33	1.01	.89
Fourth leg:					
Altitude	.13	3.30	3.77	1.32	1.09
Air speed	[1] .27	3.87	3.91	1.21	1.12
Heading	[1] .31	3.69	4.07	1.19	1.08

See footnotes at end of table.

TABLE A9.1.—*Reliabilities of coded item scores in AT-6 instrument scale*—Con.

Based on tests and retest administered to 60 students on successive days in different planes by different instructors

Item	r_{12}	M_1	M_2	*SD 1*	*SD 2*
PATTERN "B"—Continued					
First turn:					
Altitude	0.13	3.38	3.65	0.99	0.77
Air speed	.13	3.74	3.93	.99	.85
Heading	.05	4.55	4.62	.89	.59
Second turn:					
Altitude	[1].31	3.19	3.39	1.10	.94
Air speed	[1].29	3.53	3.78	.90	.87
Heading	.08	4.46	4.23	.73	1.23
Third turn:					
Altitude	.07	3.32	3.49	1.00	1.18
Air speed	.01	3.61	3.92	1.11	1.18
Heading	.16	3.98	4.26	1.37	1.14
Fourth turn:					
Altitude	.19	3.87	3.88	1.07	1.04
Air speed	.21	3.74	3.93	1.11	1.03
Heading	.10	4.27	4.36	1.16	1.16
UNUSUAL POSITION RECOVERIES:					
Wing over	.23	3.14	3.31	.92	.72
Power-on stall	−.06	3.26	3.42	.77	.70
Power-off stall	−.22	3.37	3.35	.72	.69
Diving spiral	−.13	2.84	3.08	.73	.72
ICING STALL:					
Altitude	.05	2.70	3.17	1.30	1.29
Heading	.09	2.87	3.34	1.40	1.48
Angle of bank	.00	3.16	3.25	1.24	1.23
POWER LET-DOWN:					
Altitude	−.24	3.58	4.11	1.24	.86
Angle of bank	.05	3.38	3.83	1.19	1.04
RIGHT DESCENDING TURN:					
Air speed	.14	3.38	4.06	1.19	.79
Angle of bank	.05	3.38	3.83	1.19	1.14
LEFT DESCENDING TURN:					
Air speed	[2].34	3.61	3.95	1.18	1.05
Angle of bank	.15	3.47	3.74	1.04	1.13

[1] 5 or less chances in 100 of obtaining a positive correlation of this size by chance alone.
[2] 1 or less chances in 100 of obtaining a positive correlation of this size by chance alone.
[3] Tetrachoric *r*.

TABLE A9.2.—*Validities of coded item scores in AT-6 instrument scale*

Based on scores for second day

Item	N_t	P_s	M_s	M_f	SD_t	r_{bis}
INSTRUMENT TAKE-OFF:						
Heading control during roll	94	0.532	4.18	3.73	1.98	[1]0.29
Air speed control initial climb	93	.538	3.80	3.44	1.13	.20
Instructor help during take-off	94	.521	2.67	2.20	.90	[2].33
RIGHT CLIMBING TURN:						
Air speed control	97	.536	3.90	3.62	.97	.18
Angle of bank	97	.536	4.12	4.09	1.14	.00
STRAIGHT CLIMB:						
Air speed control	97	.536	4.29	4.00	.96	.19
Heading control	96	.531	4.37	4.33	.72	.03
LEFT CLIMBING TURN:						
Air speed control	96	.531	4.29	4.20	.72	.08
Angle of bank control	97	.536	4.27	4.27	.79	.00
STRAIGHT CLIMB:						
Air speed control	94	.532	4.46	4.32	.61	.18
Heading control	94	.532	4.58	4.30	.68	[1].26
LEVEL OUT:						
Altitude control	97	.526	4.16	3.70	.80	[2].36
Heading control	97	.526	4.35	3.74	.95	[2].41
RIGHT STEEP TURN:						
Altitude control	98	.531	3.52	2.96	1.20	[1].29
Air speed control	97	.526	3.98	3.70	.96	.19
Angle of bank control	97	.536	4.43	4.30	.84	.09
LEFT STEEP TURN:						
Altitude control	98	.531	3.63	3.50	.85	.04
Air speed control	96	.542	3.67	3.55	.95	.08
Angle of bank control	97	.536	4.19	4.27	.96	−.05
PATTERN "A"						
First leg:						
Altitude	98	.531	4.31	4.09	.86	.16
Air speed	95	.526	4.54	4.33	.78	.19
Heading	95	.526	3.74	3.51	1.28	.11

See footnotes at end of table.

Table A9.2.—*Validities of coded item scores in AT-6 instrument scale*—Continued

Based on scores for second day

Item	N_t	P_s	M_s	M_t	SD_t	r_{bis}
PATTERN "A"—Continued						
Second leg:						
Altitude	98	0.531	4.38	3.87	0.89	[2]0.36
Air speed	96	.531	4.37	3.96	.80	[2].33
Third leg:						
Altitude	98	.531	4.00	3.89	.99	.07
Air speed	98	.531	4.42	4.04	.88	[1].27
Fourth leg:						
Altitude	97	.536	4.21	3.87	1.05	.21
Air speed	94	.521	4.37	3.93	1.02	[1].27
First turn:						
Altitude	98	.531	3.96	3.50	1.08	[1].27
Air speed	96	.531	4.12	4.04	.87	.05
Heading	98	.531	3.46	3.30	1.53	.06
Angle of bank	96	.531	4.12	4.04	.91	.05
Second turn:						
Altitude	96	.531	4.12	3.49	1.02	[2].39
Air speed	97	.526	4.27	4.00	.79	.22
Heading	94	.543	3.69	3.28	1.36	.19
Angle of bank	96	.531	4.08	4.24	.91	−.12
Third turn:						
Altitude	97	.536	4.21	3.80	1.00	[1].26
Air speed	97	.536	4.35	3.89	.95	[1].30
Heading	93	.527	3.43	3.34	1.51	.04
Angle of bank	96	.521	4.30	3.89	.90	[1].29
Fourth turn:						
Altitude	97	.536	4.12	3.82	1.00	.18
Air speed	97	.536	4.33	3.93	.91	[1].27
Heading	97	.536	3.38	3.09	1.54	.12
Angle of bank	96	.531	4.20	3.91	.98	.18
PATTERN "B"						
First leg:						
Altitude	98	.531	4.73	4.41	.67	[1].30
Air speed	97	.526	4.53	4.33	.69	.19
Heading	97	.526	4.78	4.39	.70	[2].33
Second leg:						
Altitude	98	.531	3.92	3.76	.89	.12
Air speed	95	.547	3.98	3.91	1.06	.04
Heading	95	.537	4.41	4.16	.84	.19
Third leg:						
Altitude	98	.531	4.35	3.80	.92	[2].37
Air speed	97	.526	4.24	3.91	.90	.22
Heading	96	.521	4.36	4.09	1.03	.17
Fourth leg:						
Altitude	96	.521	3.74	3.50	1.12	.13
Air speed	92	.511	4.15	3.64	1.14	[1].29
Heading	96	.521	4.12	3.87	1.21	.13
First turn:						
Altitude	98	.531	3.67	3.50	.82	.13
Air speed	96	.542	4.04	3.70	.86	.24
Heading	95	.537	4.57	4.16	1.05	.24
Second turn:						
Altitude	98	.531	3.62	3.15	1.07	[1].27
Air speed	97	.526	3.94	3.57	.97	.24
Heading	95	.516	4.41	4.09	1.22	.16
Third turn:						
Altitude	98	.531	3.71	3.17	1.21	[1].28
Air speed	97	.526	4.04	3.59	1.15	[1].25
Heading	96	.521	4.26	4.13	1.25	.07
Fourth turn:						
Altitude	97	.526	3.98	3.54	1.14	.24
Air speed	96	.531	4.12	3.56	1.13	[1].31
Heading	90	.522	4.47	3.98	1.25	[1].25
UNUSUAL POSITION RECOVERIES:						
Wingover	95	.537	3.18	3.39	.73	−.18
Power-on stall	94	.543	3.31	3.56	.64	−.24
Power-off stall	92	.543	3.32	3.36	.63	.00
Diving spiral	95	.537	2.96	2.91	.78	.04
ICING STALL:						
Altitude	95	.526	2.96	3.29	1.36	−.15
Heading	94	.532	3.14	3.09	1.36	.02
Angle of bank	87	.540	3.11	3.05	1.33	.03
POWER LET-DOWN:						
Altitude	95	.526	3.96	3.98	.96	.01
Heading	97	.526	3.90	3.61	1.01	.18
RIGHT DESCENDING TURN:						
Air speed	97	.526	4.14	3.72	.95	[1].28
Angle of bank	97	.526	3.96	3.48	.95	[1].29
LEFT DESCENDING TURN:						
Air speed	96	.521	4.16	3.72	1.02	[1].27
Angle of bank	96	.521	3.90	3.52	1.08	.22

[1] 5 or less chances in 100 of obtaining a positive correlation of this size by chance alone.
[2] 1 or less chances in 100 of obtaining a positive correlation of this size by chance alone.

Copies of the Check Ride Booklets for Objective Measures of Flying Skill During Temporary Training Freezes

PRIMARY

A. ______________________________ B. ______________________
Name: (Last) (First) (Middle) (Avn. Cadet ASN)

C. Date: ______________________ Class ______________________
Day Month Year

School ______________________ Don't write here: Secretary will supply pilot stanine (with credit added).

Place ______________________

P+Cr: ______________________

D. Did you have any flying experience, handling the controls of a plane in the air, *before entering primary flying school?* (Circle the number following the answer that applies to you.)

Soloed before entering primary school 1
More than 4 hours stick time but never soloed 2
Less than 4 hours of stick time 3

E. Circle the appropriate number to indicate the kind of plane or planes on which you had the 10 weeks of primary training *before the freeze.*

Fairchild followed by Stearman 1
Only Stearman 2
Only Fairchild 3
Other 4

F. Circle the appropriate number to indicate the number of classes you have been held over *after graduating from preflight.*

None 0
Once 1
Twice 2
Three or more 3

G. Reject:

No 1
Yes 2

Check Pilot: ______________________

703322—47—30

H. Number of Hours Student Has Had on Stearman Plane.

Circle proper number below:

0–10 hours	1
11–20 hours	2
21–30 hours	3
31–40 hours	4
41–50 hours	5
51–60 hours	6
61–70 hours	7
71–80 hours	8
81 or more hours	9

Immelmann

I. Check pilot lines plane up with road. He says, "Do an Immelmann. *Start with an air speed no greater than 150 m. p. h.* Use throttle any way you want to. You will get a good score if your starting speed is not above 150 m. p. h. and if you come out lined up within 45 degrees of the road without stalling."

Circle proper number below:

COMES OUT WITHIN ±45° OF ROAD WITHOUT STALLING	3
COMES OUT BEYOND ±45° OF ROAD WITHOUT STALLING	2
STALLS (OR STARTS FASTER THAN 150 M. P. H.)	1

Loop

J. Check pilot lines plane up with road. He says, "Do a loop. *Start with an air speed no greater than 110 m. p. h.* Use throttle any way you want to. You will get a good score if your starting speed is not above 110 m. p. h. and if you come out lined up within 45 degrees of the road without stalling."

COMES OUT WITHIN ±45° OF ROAD WITHOUT STALLING	3
COMES OUT BEYOND ±45° OF ROAD WITHOUT STALLING	2
STALLS (OR STARTS FASTER THAN 110 M. P. H.)	1

Two 360° Turns (Left and Right)

Check pilot lines plane up with road or section line at safe altitude. He says, "Do a 360° turn to the left with a bank of at least 60°. When the bank is 60°, the cabane strut will be parallel with the horizon. Immediately after the 360° turn to the left, do a 360° turn to the right with a bank of at least 60°. You will get a good score if your altitude varies no more than 20 feet above or below your starting altitude, if your bank stays above 60° after you roll into a turn, and if you roll out of the second turn lined up within 45° of your starting heading."

Scoring Directions:	*Circle* proper number below:
K. Note the starting altitude and record the maximum deviation, either above or below, from this starting altitude.	K. Altitude Deviation: 0 to ±20 ft. . . . . 3 ±21 to ±60 ft. . . . . 2 ±61 ft. and over 1
L. Record if student falls below 60° bank between roll-in and roll-out on each turn.	L. Maintained 60° bank or more: In each turn . . . . 3 In one turn only . . . . 2 In neither turn 1
M. Record if student rolls out of second turn within 45° of heading at beginning of first turn.	M. Final roll-out: Within 45° . . . . 3 Outside 45° 1

Simulated Forced Landing

Check pilot chooses a field about three-quarters of a mile square, disregarding suitability of terrain for actual landing. He says, "See that field? (Describes field and makes sure student identifies it.) When we are directly over that field flying downwind at *1,200 feet above the ground*, I'll cut the throttle. You will simulate a forced landing in the first two-thirds of that field. You will get a good score if you make a base leg which would get into the first two-thirds of the field without slipping."

	Circle proper number below:
N. Student Simulates a Base Leg:	N. Yes . . . . 3 No 1
O. Student Would Have Landed:	O. In first 2/3 . . . . 3 Outside first 2/3 1

Two Accuracy Landings

Check pilot puts plane on 45° entry to rectangular pattern. He says, "Make a three-point landing in the middle zone marked off by the lines (flags). *After cutting the throttle at the key position, do not use any throttle unless you are going to miss the field entirely.* Don't make a wheel landing. Make a three-point landing. You will be graded on where you land. A landing with tail as high as horizontal or a bounce of more than 3 feet will give you the lowest score. After you finish the first landing, take off and do another one just the same.

"You will get only one try at each landing. If you miss the field and have to go around for safety reasons, that counts as one landing."*

*If bad landing necessitates go-around, circle the No. 1 for P, Q, and R or for S, T, and U, as the case may be.

Scoring Directions:

P and S. Record in which zone the wheels of the plane *first* touched the ground. (Plane must land; no second try or go-around allowed.)

Q and T. Record if tail in horizontal position or higher on landing.

R and U. Record if plane bounced 3 feet or more.

Circle proper number below:

P. Area where plane first touches:

1
| 2 |
| 3 |
| 2 |
1

Q. **Landing Position:**

3-point or slightly tail first..	3
Wheels first—tail below horizontal..................	2
Tail horizontal or higher...	1

R. **Bounced or Dropped in:**

Neither..................	3
Less than 3 feet............	2
More than 3 feet...........	1

S. Area where plane first touches:

1
| 2 |
| 3 |
| 2 |
1

T. **Landing Position:**

3-point or slightly tail first..	3
Wheels first—tail below horizontal..................	2
Tail horizontal or higher...	1

U. **Bounced or Dropped in:**

Neither..................	3
Less than 3 feet............	2
More than 3 feet...........	1

BASIC

A. ....
Name: (Last) (First) (Middle)

B. ....
(Avn. Cadet ASN)

C. Date: ....
Day Month Year

School ....

Place ....

Class ....

Don't write here: Secretary will supply pilot stanine (with credit added)

P + Cr: .....

D. Did you have any flying experience, handling the controls of a plane in the air, *before entering primary flying school?* (*Circle the number following the answer that applies to you.*)

Soloed before entering primary school	1
More than 4 hours stick time but never soloed	2
Less than 4 hours of stick time	3

E. Circle the number indicating the kind of plane or planes on which you had your primary training.

Fairchild, then Stearman	1
Ryan, then Stearman	2
Only Stearman	3
Only Fairchild	4
Only Ryan	5

F. Circle the number to indicate what plane or planes you used during the 10 weeks of basic training *before the freeze.*

BT (13, 14, 15) and 8 or +hrs. AT-6	1
BT (13, 14, 15) and 8 or +hrs. AT-17 or UC-78	2
BT (13, 14, 15) and 8 or +hrs. AT-10	3
*Only BT 13, 14, or 15	4
*Only AT-6	5
*Only AT-17 or UC-78	6
*Only AT-10	7
Other	8
If other, specify	

*Less than 8 hrs. on other plane.

G. Circle the appropriate number to indicate the number of classes you have been held over in all phases of training after *graduating from pre-flight.*

None	0
Once	1
Twice	2
Three or more	3

H. Reject:

No	1
Yes	2

Check Pilot: Contact .... Instrument ....

Two 180° Steep Turns (Left and Right)

Directions to Student: "You are lined up with a road (or section line). I want you to do a 180° steep turn to the *left* and *immediately* roll from that into a 180° steep turn to the *right*. You *must* reach a bank of at least 60° in *each* turn. You will be graded on your greatest deviation from the starting altitude and on whether or not you reach a 60° bank in *each* turn."

Scoring Directions: *Circle* proper number below:

I. Note the starting altitude and record the maximum deviation, either above or below, from this starting altitude.

I. Maximum deviation in feet from starting altitude.

0 to ± 50 feet	3
± 51 to ± 100 feet	2
± 101 feet and over	1

J. Record whether or not student reaches a 60° bank at least once in *each* 180° turn.

J. Reached 60° bank:

In each turn	3
In one turn only	2
In neither turn	1

Accuracy Landing (Throttle and Flaps As Desired)

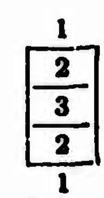
DIRECTIONS TO STUDENT: "I'll put the plane on 45° entry to rectangular pattern. I want you to make a three-point landing in the middle zone marked off by the lines (flags). *Use throttle and flaps as you desire.*

"Don't make a wheel landing. Make a three-point landing. You will be graded on *where* you land. A landing with tail as high as the horizontal position *or* bounce of more than 3 feet will give the lowest score. Do not go around unless you have to for safety reasons."*

SCORING DIRECTIONS:

K. Record in which zone the wheels of the plane first touched the ground.

K. Area where plane first touches.



L. Record attitude on landing.

L. LANDING POSITION:

Three-point or slightly tail first	3
Wheels first—tail below horizontal	2
Tail horizontal or higher	1

M. Record whether plane bounced or dropped in.

M. BOUNCED OR DROPPED IN:

Neither	3
Less than 3 feet	2
More than 3 feet	1

*If bad landing necessitates go-around, circle the No. 1 for K, L, and M.

Accuracy Landings (Flaps As Desired)

DIRECTIONS TO STUDENT: "Continue in traffic pattern. Make a three-point landing in the middle zone marked off by the lines (flags). *After cutting the throttle at the key position, do not use any throttle unless you are going to miss the field entirely. Use flaps as desired.*

"Don't make a wheel landing. Make a three-point landing. You will be graded on *where* you land. A landing with tail as high as the horizontal position or a bounce of more than 3 feet will give you the lowest score. Do not go around unless you have to for safety reasons".*

* If bad landing necessitates go-around, circle the No. 1 for N, O, and P.

SCORING DIRECTIONS:

N. Record in which zone the wheels of the plane *first* touched the ground.

O. Record if tail in horizontal position or higher on landing.

P. Record if plane bounced 3 feet or more.

Circle proper number below:

N. Area where plane first touches.

1
2
3
2
1

O. LANDING POSITION:

- Three-point or slightly tail first........ 3
- Wheels first—tail below horizontal........ 2
- Tail horizontal or higher.. 1

P. BOUNCED OR DROPPED IN:

- Neither........ 3
- Less than 3 feet........ 2
- More than 3 feet........ 1

Straight And Level Flight (Instrument)

DIRECTIONS TO STUDENT: "Signal when you are flying level at 90° (or 270°) heading at 120 m. p. h. at —— feet altitude. (Select altitude for smoothest air possible.) (Wait for signal.) Cage the directional gyro and artificial horizon. Indicate when this has been done. (Wait for caging of instruments.) Now fly straight and level at 120 m. p. h. for 3 minutes. You will be graded on variations in heading, air speed, and altitude."

SCORING DIRECTIONS:

Q. Note starting heading and record maximum deviation, plus or minus.

Q. HEADING DEVIATION:

- 0° to ±5°........ 3
- ±6° to ±10°........ 2
- ±11° and over........ 1

R. Note starting air speed and record maximum deviations, plus or minus.

R. AIR SPEED DEVIATION:

- 0 to ±5 m. p. h........ 3
- ±5 to ±10 m. p. h........ 2
- ±11 m. p. h. and over... 1

S. Note starting altitude and record maximum deviation, plus or minus.

S. ALTITUDE DEVIATION:

- 0 to ±50 feet........ 3
- ±51 to ±100 feet........ 2
- ±101 feet and over........ 1

360° Steep Right Turn (Instrument)

DIRECTIONS TO STUDENT: "Make a 360° *right* turn with a bank of over 30° at an air speed of 120 m. p. h. and come out on a heading. Use full instrument panel. You will be graded on variations in altitude during turn and deviation from correct heading at end of turn. *If your bank falls to 30° or below during turn, you will receive the lowest score.*"

SCORING DIRECTIONS:	*Circle* proper number below:	
T. Note heading at beginning and end of turn. Record difference, plus or minus.	T. HEADING DEVIATION:	
	0° to ±5°	3
	±6° to ±10°	2
	±11° and over	1
U. Note starting altitude and record maximum deviation, plus or minus.	U. ALTITUDE DEVIATION:	
	0 to ±50 feet	3
	±51 to ±100 feet	2
	±101 feet and over	1
V. Note bank on artificial horizon and record if bank falls to 30° or below at any time during turn except on roll-in and roll-out.	V. OVER 30° BANK:	
	Maintained	3
	NOT maintained	1

Constant Speed Climb and Descent (Instrument)

DIRECTIONS TO STUDENT: "Fly straight and level at a 90° or (270°) heading at *110 m. p. h.*, repeat *110 m. p. h.* (Wait) Now climb 1,000 feet at 500 feet per minute and *immediately* start descent. Level off at starting altitude, holding *110 m. p. h. During this maneuver hold the heading and air speed constant.* You will be graded on variations in heading and air speed during maneuver and on deviations from correct altitude at top and bottom."

SCORING DIRECTIONS:		
W. Note starting heading and record maximum deviation, plus or minus.	W. HEADING DEVIATION:	
	0° to ±5°	3
	±6° to ±10°	2
	±11° and over	1
X. Note starting air speed and record maximum deviation, plus or minus.	X. AIR SPEED DEVIATION:	
	0 to ±5 m.p.h	3
	±6 to ±10 m.p.h	2
	±11 m.p.h. and over	1
Y. Note deviations from the correct altitudes at top of climb and at end of descent and check the deviation that was greatest.	Y. ALTITUDE DEVIATION:	
	0 to ±50 feet	3
	±51 to ±100 feet	2
	±101 feet and over	1

ADVANCED SINGLE ENGINE

A. .... B. ....

Name: (Last) (First) (Middle) (Avn. Cadet ASN)

C. Date: .... Class ....

Day Month Year

School .... Don't write here: Secretary will supply pilot stanine (with credit added).

Place .... P+Cr: ....

D. Did you have any flying experience, handling the controls of a plane in the air, *before entering primary flying school?* (*Circle the number following the answer that applies to you*).

Soloed before entering primary school	1
More than 4 hours' stick time but never soloed	2
Less than 4 hours of stick time	3

Circle the number indicating the kind of plane or planes on which you had your *primary* and *basic* training.

E. At Primary:

Fairchild and Stearman	1
Ryan and Stearman	2
Only Stearman	3
Only Fairchild	4
Only Ryan	5
Other	6

F. At Basic:

BT (13, 14, 15) and 8 or + hours AT-6	1
BT (13, 14, 15) and 8 or + hours AT-17 or UC-78	2
BT (13, 14, 15) and 8 or + hours AT-10	3
*Only BT 13, 14, or 15	4
*Only AT-6	5
*Only AT-17 or UC-78	6
*Only AT-10	7
*Other	8

*Less than 8 hours on other planes.

G. Circle the number to indicate what plane you used during the 10 weeks of ADVANCED training *before the freeze*.

AT-6	1
AT-9	2
AT-10	3
AT-17 or UC-78	4
B-25	5
Other	6

H. Circle the appropriate number to indicate the number of classes you have been held over in all phases of training *after graduating from preflight*.

None	0
Once	1
Twice	2
Three or more	3

I. Reject:

No	1
Yes	2

Check Pilot: Contact .... Instrument ....

Two Maximum Performance Loops

DIRECTIONS TO STUDENT: "Do two separate maximum performance loops with a starting A/S no greater than *160 m. p. h.*, tachometer reading of *1,925 r. p. m.*, and manifold pressure of *25" Hg*. Do each loop without using more throttle. After finishing the first loop, clear yourself and then do the second. Do each loop without stalling out or letting the plane shudder."

SCORING DIRECTIONS: For each loop:

Circle 3.—If the plane does not shudder even slightly or stall or fall out.

Circle 2.—If the plane shudders even slightly but does not stall or fall out.

Circle 1.—If starting A/S is more than 160 m. p. h., if throttle is used in loop, or if plane stalls or falls out.

Circle proper number below:

J. FIRST LOOP:

No stall or shudder	3
Shudder	2
Stalled, fell out, used throttle, or over 160 m. p. h.	1

K. SECOND LOOP:

No stall or shudder	3
Shudder	2
Stalled, fell out, used throttle, or over 160 m. p. h.	1

Two 180° Steep Turns (Left and Right)

DIRECTIONS TO STUDENT: "You are lined up with a road (or section line). I want you to do a 180° steep turn to the *left* and *immediately* roll from that into a 180° steep turn to the *right*. You must reach a bank of at least 60° in *each* turn. You will be graded on your greatest deviation from the starting altitude on whether or not you reach a 60° bank in *each* turn."

SCORING DIRECTIONS:

L. Note the starting altitude and record the maximum deviation, either above or below, from this starting altitude.

M. Record whether or not student reaches a 60° bank at least once in each 180° turn.

Circle proper number below:

L. Maximum deviation in feet from starting altitude:

0 to ±50 feet	3
±51 to ±100 feet	2
±101 feet and over	1

M. Reached 60° bank:

In each turn	3
In one turn only	2
In neither turn	1

Accuracy Landing (Throttle and Flaps as Desired)

DIRECTIONS TO STUDENT: "I'll put the plane on 45° entry to rectangular pattern. I want you to make a 3-point landing in the middle zone marked off by the lines (flags). *Use throttle and flaps as you desire.*

"Don't make a wheel landing. Make a 3-point landing. You will be graded on *where* you land. A landing with tail as high as the horizontal position *or* bounce of more than 3 feet will give the lowest score. Do not go around unless you have to for safety reasons."*

SCORING DIRECTIONS: *Circle* proper number below:

N. Record in which zone the wheels of the plane *first* touched the ground.

N. Area where plane first touches.

1
2
3
2
1

O. Record attitude on landing.

O. LANDING POSITION:

3-pt. or slightly tail first.	3
Wheels first—tail below horizontal..........	2
Tail horizontal or higher.	1

P. Record whether plane bounced or dropped in.

P. BOUNCED OR DROPPED IN:

Neither................	3
Less than 3 feet.........	2
More than 3 feet.......	1

Straight and Level Flight (Instrument)

DIRECTIONS TO STUDENT: "Signal when you are flying level at 90° (or 270°) heading at 130 m. p. h. at —— feet altitude. (Select altitude for smoothest air possible.) (Wait for signal.) Cage the directional gyro and artificial horizon. Indicate when this has been done. (Wait for caging of instruments.) Now fly straight and level at 130 m. p. h. for 3 minutes. You will be graded on variations in heading, air speed, and altitude."

SCORING DIRECTIONS:

Q. Note starting heading and record maximum deviation, plus or minus.

Q. HEADING DEVIATION:

0° to ±5°.............	3
±6° to ±10°..........	2
±11° and over........	1

R. Note starting air speed and record maximum deviation, plus or minus.

R. AIR SPEED DEVIATION:

0 to ±5 m. p. h..........	3
±6 to ±10 m. p. h.......	2
±11 m. p. h., and over....	1

S. Note starting altitude and record maximum deviation, plus or minus.

S. ALTITUDE DEVIATION:

0 to ±50 feet..........	3
±51 to ±100 feet......	2
±101 feet and over.....	1

*If bad landing necessitates go around, circle the No. 1 for N, O, and P.

360° Steep Right Turn (Instrument)

DIRECTIONS TO STUDENT: "Make a 360° *right* turn with a bank of *over* 30° at an air speed of 130 m. p. h. and come out on a heading. Use full instrument panel. You will be graded on variations in altitude during turn and deviation from correct heading at end of turn. If *your bank falls to 30° or below during turn, you will receive the lowest score.*"

SCORING DIRECTIONS:	Circle proper number below:	
T. Note heading at begining and end of turn. Record difference, plus or minus.	T. HEADING DEVIATION:	
	0° to ± 5°	3
	±6° to ±10°	2
	±11° and over	1
U. Note starting altitude and record maximum deviation, plus or minus.	U. ALTITUDE DEVIATION:	
	0 to ±50 feet	3
	±51 to ±100 feet	2
	±101 feet and over	1
V. Note bank on artificial horizon and record if bank falls to 30° or below at any time during turn except on roll-in and roll-out.	V. OVER 30° BANK:	
	Maintained	3
	NOT maintained	1

Instrument Let-Down and Low Approach

DIRECTIONS TO STUDENT: "Make a regular let-down and low approach according to the instructions for this test. Shake the stick at the proper number of seconds after the low cone. When you shake the stick, you will be graded on whether you are on the beam, at the correct altitude, at the correct time after the low cone. Reset your altimeter to —— feet. You are now on the beam at proper altitude about one minute before the high cone and headed toward it."

SCORING DIRECTIONS:

WHEN STUDENT SHAKES STICK, CHECK PILOT RECORDS:

W. POSITION.—Plane's track in relation to ground markers.	W. POSITION:	
	Inside center markers	3
	Outside center markers but inside outer markers	2
	Outside outer markers	1
X. TIME.—Difference between prescribed interval (—— seconds) and student's elapsed time past low cone.	X. TIME DIFFERENCE:	
	0–5 seconds	3
	6–15 seconds	2
	16 or more	1

Y. ALTITUDE.—Number of feet difference between prescribed indicated altitude (—— ft.) and student's indicated altitude.

Y. ALTITUDE DIFFERENCE:

0–20 feet	3
21–60 feet	2
61 or more feet	1

ADVANCED TWO-ENGINE

A. .... Name: (Last) (First) (Middle)

B. .... (Avn. Cadet ASN)

C. Date: .... Day Month Year

Class........................

School........................

Don't write here: Secretary will supply pilot stanine (with credit added).

Place........................

P+Cr: ....

D. Did you have any flying experience, handling the controls of a plane in the air, *before entering primary flying school?* (*Circle the number following the answer that applies to you.*)

Soloed before entering primary school	1
More than 4 hours stick time but never soloed	2
Less than 4 hours of stick time	3

Circle the number indicating the kind of plane or planes on which you had your *primary* and *basic* training.

E. AT PRIMARY:

Fairchild and Stearman	1
Ryan and Stearman	2
Only Stearman	3
Only Fairchild	4
Only Ryan	5
Other	6

F. AT BASIC:

BT (13, 14, 15) and 8 or +hours AT–6	1
BT (13, 14, 15) and 8 or +hours AT–17 or UC-78	2
BT (13, 14, 15) and 8 or +hours AT–10	3
*Only BT 13, 14, or 15	4
*Only AT–6	5
*Only AT–17 or UC–78	6
*Only AT–10	7
Other	8

*Less than 8 hours on other planes.

G. Circle the number to indicate what plane you used during the 10 weeks of ADVANCED training *before the freeze.*

AT–6	1
AT–9	2
AT–10	3
AT–17 or UC–78	4
B–25	5
Other	6

H. Circle the appropriate number to indicate the number of classes you have been held over in all phases of training *after graduating from preflight.*

None	0
Once	1
Twice	2
Three or more	3

I. REJECT:

No	1
Yes	2

CHECK PILOT: Contact.................. Instrument..............

Single-Engine Procedure

(At 2,000 ft. or more above ground)

Check pilot sets up the following conditions:

1. Wheels down.
2. Cardinal heading.
3. A/S exactly 160 m. p. h.
4. 2,200 r. p. m. and enough manifold pressure to maintain 160 m. p. h.

"I've set up the plane for single-engine procedure. You will be graded on how well you maintain heading, air speed, and altitude."

When airplane and student are ready, write down the following *before* cutting engine:

A/S .... m. p. h. Altitude .... ft. Heading ....°

NOW, CHECK PILOT CUTS EITHER THROTTLE COMPLETELY

During the next *2-minute* period, circle the following deviations:

J. HEADING (maximum deviation):

Stayed within ±5° limits	3
Stayed within ±10° limits	2
Exceeded ±10° limits	1

K. AIR SPEED (minimum A/S):

150 m. p. h. and over	3
140 m. p. h. and over	2
Below 140 m. p. h	1

L. ALTITUDE (maximum lost):

Less than 100 feet	3
Less than 200 feet	2
More than 200 feet	1

M. Procedure check list: Did student do ALL of the following correctly?

Add power immediately Retard prop control Retard correct prop Raise wheels	Yes	3
	No	1

Single-Engine Landing

On the downwind leg after completion of the GUMP check, the check pilot will cut the inside engine, e. g., right engine for right-hand traffic.

"Fly the approach at 140 m. p. h. and make your first contact in the second zone."

CHECK EACH ITEM AS IT OCCURS:

N. Turn onto approach: Lined up with runway.	Yes	3
	No	1
O. A/S on approach: (Maximum deviation from 140 m.p.h.)	±5 m. p. h	3
	±10 m. p. h	2
	Over 10 m. p. h	1
P. Was gear checked after lowering on approach?	Yes	3
	No	1
Q. Where was power cut completely? (Check nearest answer.)	Roll onto approach	1
	Half-way down	1
	Just prior to round-out	3
	At landing	2
R. Where was landing contact first made? (Goal is Zone 2.)	Zone 1 (first 500 ft. of runway)	2
	Zone 2 (second 500 ft.)	3
	Zone 3 (third 500 ft.)	2
	Overshot or went around	1
S. Landed in skid:	Yes	1
	No	3

Straight and Level Flight (Instrument)

"Select a cardinal heading and set up a straight and level course at normal cruising A/S. When you are all set, cage the directional gyro and artificial horizon and continue on course for 3 minutes. To get the highest score, stay within the usual 5-5-50 limits."

T. HEADING DEVIATIONS:	Stayed within ±5°	3
	Stayed within ±10°	2
	Went beyond 10°	1
U. A/S VARIATIONS:	Stayed within ±5 m. p. h	3
	Stayed within ±10 m. p. h	2
	Went beyond 10 m. p. h	1
V. ALTITUDE VARIATIONS:	Stayed within ±50 feet	3
	Stayed within ±100 feet	2
	Went beyond 100 feet	1

Instrument Let-Down and Low Approach

"Make a regular let-down and low approach according to the instructions for this test. Give me a signal at the proper number of seconds after the low cone. When you give me the signal you will be graded on whether you are on the beam, at the correct altitude, at the correct time after the low cone. . . . You are now on the beam at proper altitude about one minute before the high cone and headed toward it."

WHEN STUDENT GIVES THE SIGNAL, CHECK PILOT RECORDS:

W. POSITION.—Plane's track in relation to ground markers:

Inside center	3
Outside center markers *but* inside outer markers	2
Outside outer markers	1

X. TIME.—Difference between prescribed time (______ seconds) and student's elapsed time past low cone:

0-5 seconds	3
6-15 seconds	2
16 or more	1

Y. ALTITUDE.—Number of feet difference between prescribed indicated altitude (______ ft.) and student's indicated altitude:

0-20 feet	3
21-40 feet	2
Over 40 feet	1

Power-Off Approach and Landing

"Make a power-off approach and landing. Cut the power completely at traffic altitude when you think you can land in zone 2 with power off all the way down."

Z. Altitude at roll into the approach varied above or below the altitude (______ ft.) by:

Less than 100 feet	3
Between 100 and 150 feet	2
More than 150 feet	1

AA. Was power added on approach?	Yes	1
	No	3
BB. When was contact first made?	Zone 1	2
	Zone 2	3
	Zone 3	2
	Over shot or went around	1

THE RELIABILITY OF THE OBJECTIVE MEASURES USED[1]

The fact that a fair proportion of the measures used in this study show reliable differences between the 10 and 15 week groups, the purpose for which they were designed, shows that they are reliable

[1] The design and statistical analysis of this study were supervised by Lt. J. K. Hemphill and T/Sgt. W. O. Matheny.

enough for measuring group differences. In experiments which were auxiliary to the main study it was possible to investigate further some of the measures in order to determine whether or not they were reliable enough for use in measuring differences between individuals.

Observer Reliability

Several of these studies have already been described in chapters 6 and 7 of this report. They deal with the ability of the check rider to make an accurate observation of the performance being measured. In these experiments the scores of two observers measuring the same thing at the same time were compared. In general, fairly high correlations were found between the scores recorded by the two independent observers. This means that each one of them is able to observe accurately enough so that his scores agree with those of the other observer. Since a number of different random pairings of observers were involved in each of these studies, the fairly high correlation indicates absolute as well as relative agreement.

Test-Retest Reliabilities

Studies of test-retest reliability give an estimate of the sum of the errors introduced by (1) the observer and (2) the variability of the performance itself, for example, the day-to-day differences caused by changes in conditions such as the weather.

Studies dealing with test-retest reliability of objective measures of flying skill at the primary level and in the basic phase of instrument flying have already been described in chapters 6 and 9 of this volume.

The test-retest reliability of each of the measures in the check ride administered to basic students during the training freezes was determined by administering two independent check rides to each student. The two check rides were given by different check riders, in different days, and in different planes. A total of 102 basic students were used in this study. They were tested at Goodfellow Field, San Angelo, Tex., with a 2-day interval between test and retest. Thirty-seven of these students were tested in the middle of November 1944 after having had 10 weeks of training at the basic level and 65 in the middle of December after having had 15 weeks of training at the basic level.

The test-retest reliability of the measures administered on different days was determined by computing product-moment correlation coefficients between the scores on the first and second administration. In order to get a more stable figure, the two different samples were combined by use of the Fisher z-transformation.

Since the measures were scored in three broad categories, a correction for the effect of this grouping was calculated. This correction is an estimate of the coefficient of the reliability which would be obtained if the measures were scored in the finer categories of feet of altitude, miles per hour, etc. The uncorrected coefficients give the best estimate of the reliability of the scale scored in the three broad categories used in this study. The corrected coefficients must be used in comparing the results of this study with those of others, such as the one in chapter 8 of this volume on the comparison of three different methods of scoring instrument measures, in which the measures were not grouped in broad categories.

The reliabilities of the measures in the basic check ride when test and retest were given on different days, with different check riders, and in different planes, are presented in table A10.1.

TABLE A10.1—*Test-retest reliability of objective measures of flying skill in the basic check ride*

Item measure	Goodfellow—10 wk (N=37)					Goodfellow—15 wk (N=65)					(N=102)	
	Test		Retest		*r*	Test		Retest		*r*	*z*[1] comb.	*r*[2] corr.
	Mean	*SD*	Mean	*SD*		Mean	*SD*	Mean	*SD*			
I Altitude deviation 180° steep turns.	2.03	0.68	2.27	0.72	0.26	1.66	0.64	1.77	0.70	0.07	0.14	0.19
J Reached 60° bank 180° steep turns.	2.73	.55	2.78	.47	.09	2.85	.40	2.85	.44	.13	.12	.17
K Zone plane lands—accuracy landing, training and flaps.	1.76	.75	1.95	.73	.12	2.00	.89	2.05	.90	.08	.09	.13
L Landing attitude—accuracy landing, training and flaps.	2.62	.63	2.70	.46	−.02	2.75	.50	2.60	.65	−.26	−.18	−.24
M Bounced or dropped—accuracy landing, training and flaps.	2.11	.65	2.41	.68	−.16	2.28	.76	2.25	.77	.12	.02	.03
N Zone plane lands—accuracy landing, flaps.	1.70	.77	1.81	.73	.00	1.63	.78	1.78	.73	.02	.01	.02
O Landing attitude—accuracy landing, flaps.	2.68	.52	2.65	.67	−.17	2.74	.50	2.68	.61	.13	.02	.03
P Bounced or dropped—accuracy landing, flaps.	2.27	.68	2.49	.76	.22	2.33	.71	2.31	.80	−.08	.03	.04
Q Head deviation—straight and level flight—instrument.	1.92	.78	2.19	.69	.18	2.14	.65	2.00	.78	.15	.16	.22
R Air speed deviation—straight and level flight—instrument.	2.51	.68	2.65	.53	−.07	2.55	.53	2.25	.68	.01	.00	.00
S Altitude deviation—straight and level flight—instrument.	2.43	.64	2.35	.67	.15	2.4[illegible]	.66	2.20	.66	−.09	.01	−.01
T Heading deviation—360° R. turn—instrument.	2.49	.64	2.32	.87	.01	2.74	.51	2.69	.58	−.17	−.11	−.15
U Altitude deviation—360° R. turn—instrument.	1.70	.67	1.76	.71	.16	2.37	.65	2.23	.72	.28	.24	.33
V Over 30° bank—360° R. turn—instrument.	2.78	.62	2.95	.32	−.06	2.88	.48	2.78	.62	−.09	−.08	−.11
W Heading deviation—constant speed, instrument.	2.27	.72	2.38	.63	.31	2.46	.68	2.28	.67	.10	.18	.25
X Air speed deviation—constant speed, instrument.	2.14	.62	2.14	.47	−.06	2.29	.65	1.97	.58	−.02	−.03	−.05
Y Altitude deviation—constant speed climb and descent—instrument.	2.51	.60	2.51	.55	.27	2.75	.47	2.68	.47	.06	.14	.19

[1] Combined *r* by Fisher's *z* transformation. The two groups being combined (10 and 15 week) are known to be different but the combined *r* is presented as a convenient summary on the assumption that the true correlations in the two groups are not markedly different.

[2] Corrected for grouping in broad categories; Peters and Van Voorhis, *Statistical Procedures and Their Mathematical Bases*, p. 395, Formula 211.

It can be seen that the test-retest reliabilities of most of the measures in this check ride were low. The two measures showing the best reliability coefficients were: (1) Altitude deviation in the instrument turn, and (2) heading deviation in the instrument constant speed climb and descent. The reliabilities computed from data grouped into three broad categories are 0.24 and 0.18, respectively, for these measures. Corrected for grouping into broad categories, these reliabilities become 0.33 and 0.25.

These results were in line with those which have been reported in chapter 6 of this volume for reliabilities of the measures administered to primary students during the training freezes.

When a single test and retest was given on different days in different planes and by different check riders, the reliability of most of the measures administered to primary and basic students was relatively low. A few of the measures were reliable enough so that, if they were combined with other good ones, the total score of a single check ride might show some promise of being stable enough for use in measuring differences between individuals.

APPENDIX TO CHAPTER ELEVEN

Fixed Gunnery as an Objective Measure of Flying Skill

Variation in Training Procedures

During the Advanced training period the students received approximately the following pregunnery instruction in the Eastern, Central, and Western Flying Training Commands for Classes 44-1-G and 44-1-H.

	Eastern	Central	Western
Lectures on gunnery principles	4-5 hours	8-10 hours	10 hours.
Armament	4 hours	10 hours	
Range estimation trainer	2-3 hours	1-2 hours	
Gunairstructor	½ hour	½ hour	None.
BB link	600 rounds [1]	200-800 rounds [1]	None.
Skeet	100-300 rounds [1]	350-500 rounds [1]	None.
Deflection training	None	½-2 hours	None.
Ground and aerial camera missions	None	2 hours	None.

[1] Rounds fired.

After classes 44-1-G and 44-1-H, the BB Link was discarded as a training aid. The Central Flying Training Command substituted a hand-held flexible gun in its place. In addition, one school in this same command added an optical sight to the skeet gun used to teach the concept of leading the target.

In the transition phase each class was divided into two groups: one was sent directly to the gunnery training while the other took P-40 transition training. After the completion of these phases the two sections reversed the type of training.

The ground school instruction given at the gunnery schools may be summarized as follows:

	Eastern	Western	Central
Lectures on gunnery principles	12 hours	None	12 hours.
Armament	10 hours	6 hours	12 hours.
Harmonization	1-2 hours	8 hours	Demonstration only.
Film assessing methods	1-2 hours	None	None.

During classes 44-1-G and 44-1-H, the average student in Eastern fired 1100 rounds in air-to-ground gunnery, while the average student in Central fired 800 rounds and the average student in Western fired 1300 rounds. The ranges, targets, and flight patterns used in the three commands were essentially the same. The gun was mounted in the wing in Central, while both Eastern and Western had the gun installed in the engine cowl. Later the cowl gun was introduced in Central.

A "foul" in air-to-ground gunnery was declared when the student continued to fire after crossing the foul line, 600 feet from the target. In Eastern one "free" foul was allowed in each practice mission. Otherwise, the rule was a penalty of 5 percent of the score for each of the first three fouls and a zero for the mission for the fourth foul. In Western, fouls were charged and penalized in exactly the same way. In Central, lack of supervisory personnel made the charging of fouls impossible.

Lack of assessing equipment resulted in Eastern students flying only one-half of their air-to-ground missions with a camera. In Central, no cameras were used in connection with air-to-ground gunnery. In Western, every mission was a camera mission.

The average student in Eastern flew 15 air-to-air missions, firing 2,100 rounds. The average student in Central and Western flew 24 air-to-air missions and fired 2,300 and 2,000 rounds, respectively.

The standard number of rounds loaded for a mission in Central and Western was 100. In Eastern, however, the first mission was 100 rounds and all succeeding missions were 200 rounds.

The standard flag-type target was used in all three commands; however, the size and design varied to some extent.

The student usually flew the same plane on all missions in Central and Western. The opposite held true in Eastern, where the students generally flew different planes on successive missions.

"Fouls" in air-to-air gunnery consisted of (1) exceeding the allowed number of passes at the target and (2) firing with less than a 30° angle between the bullets' path and the target path. The same penalties were called for each foul in air-to-air gunnery as in air-to-ground fixed gunnery. Fouls were regularly charged in Western. In Eastern, fouls were rarely charged and in Central were never charged.

Generally, weather conditions were similar at the gunnery ranges in Central and Eastern. Both were located in sea-level country near the Gulf of Mexico. There was considerable haze, frequent fog and rain during some period of the year. In contrast, weather conditions throughout the year were nearly ideal at the gunnery range in Western which was located in semi-arid land. Visibility was excellent and there was very little rain at any time of the year. Only from two to seven days per year were lost from gunnery firing due to weather conditions.

Effect of Ten Hours P-40 Training

TABLE A11.1.—*Means, sigmas and critical ratios of sections with and without prior P-40 training in air-to-ground fixed gunnery scores*

Transition school	Prior P-40 training	*N*	*M*	*D*	*SD*	*CR*	Level of sig.
							Percent
Williams	Without	87	88.94	+2.63	32.05	0.60	60
	With	68	91.57		30.00		
Craig	Without	60	87.57	-5.06	29.12	1.00	32
	With	45	82.51		22.50		
Spence	Without	66	127.52	-29.83	41.49	3.83	<1
	With	47	97.69		38.27		
Napier	Without	90	84.03	+14.32	34.94	2.64	<1
	With	101	98.35		38.02		
Foster	Without	65	130.05	+1.11	48.29	.12	90
	With	61	131.16		50.85		
Aloe	Without	74	121.66	+16.56	44.04	2.00	5
	With	52	138.22		47.06		
Moore	Without	84	123.18	+22.95	42.52	2.95	<1
	With	52	146.13		45.03		

TABLE A11.2.—*Means, sigmas, and critical ratios, with and without prior P-40 training sections in air-to-air fixed gunnery scores*

Transition school	Prior P-40 training	*N*	*M*	*D*	*SD*	*CR*	Level of sig.
							Percent
Williams	Without	87	166.93	-29.48	87.05	3.17	<1
	With	68	137.45		53.35		
Craig	Without	69	143.05	-38.47	62.95	4.23	<1
	With	59	104.58		38.72		
Spence	Without	68	144.72	-38.95	61.73	3.90	<1
	With	47	105.77		45.28		
Napier	Without	93	107.90	+18.80	44.38	2.34	2
	With	103	126.70		66.65		
Foster	Without	65	121.42	-1.84	49.26	.22	80
	With	90	119.58		47.19		
Aloe	Without	85	114.93	+10.43	51.30	1.29	20
	With	80	125.36		52.47		
Moore	Without	84	120.51	-.82	56.00	.11	90
	With	86	119.69		45.30		

Turbulence as Measured by Instructor Ratings

TABLE A11.3.—*N's, means, sigmas, and critical ratios by missions and total for instructor ratings of turbulence*

Mission	Turbulence	*N*	*M*	*D*	*SD*	*CR*	Level of sig.
							Percent
First practice	Smooth	200	19.38	....	14.47	....	....
	Rough	....	....	+5.36	14.03	2.34	2
Second practice	Smooth	176	20.36	....	11.85	....	....
	Rough	34	15.00	+1.69	12.21	.73	40
Third practice	Smooth	151	20.70	....	14.35	....	....
	Rough	54	19.11	+.12	15.37	.06	90
Fourth practice	Smooth	123	22.70	....	13.14	....	....
	Rough	48	22.64	+3.60	15.95	1.33	20
First record	Smooth	131	26.66	....	17.60	....	....
	Rough	58	23.16	-2.65	16.71	.88	38
Second record	Smooth	168	31.21	....	15.92	....	....
	Rough	35	33.86	+.88	15.43	.73	46
Total	Smooth	969	23.37	....	16.01	....	....
	Rough	229	22.49				

Relation Between Time of Day and Air-to-Air and Air-to-Ground Fixed Gunnery Scores

Mean scores by time of day were computed using the time of take-off as the time of the mission and the percent hits on a single mission for an individual student as the unit datum for both air-to-ground and air-to-air gunnery. For tabulation purposes the time of day was divided into six periods of two hours each.

There was one factor which might have given a false appearance to such data. There was the tendency of the schools to fly the later missions in each student's series at the time of day judged to be most favorable for obtaining high scores. This tendency was particularly prevalent for the record missions which were the last two flown. It was evident from the learning curve that if the later missions in the practice series were flown predominantly at one period of the day, that time of day would be expected to yield higher mean scores. There was evidence that this sort of scheduling did sometimes occur. Therefore, the percent hits for each time of day was corrected for learning.

This was accomplished in the following manner: Each student's missions were numbered consecutively. The numbers of all missions flown in each two-hour period were tallied and the mean mission number determined. The mean percent hits to be normally expected for each mission was determined from the combined learning curves (see figures 11.1 and 11.2). The difference between this figure and the percent hits expected for the mission was then added to (or subtracted from) the actual percent hits. That yielded the mean percent hits corrected for learning.

Tables A11.4 and A11.5 present the mean scores by time of day for both raw scores and scores corrected for learning. It must be remembered that the times of day Eastern and Central are not directly comparable since the gunnery school in Eastern was in the extreme eastern portion of the Central War Time zone and the school in Central was in the western portion of the same zone. The difference in solar time between the two places was approximately 44 minutes.

TABLE A11.4.—*Air-to-ground fixed-gunnery scores as a function of time of day and mean mission by time of day*

Time of day	6-7		8-9		10-11		12-13		14-15		16-17	
	N	*M*	*N*	*M*	*N*	*M*	*N*	*M*	*N*	*M*	*N*	*M*
Class 44-1-G (EFTO):												
Raw scores	130	33.85	169	29.97	169	26.76	72	25.25	116	28.09	232	34.94
Corrected scores [1]	130	34.78	169	30.87	169	27.61	72	27.05	116	29.29	232	34.34
Class 44-1-H (EFTO):												
Raw scores	526	30.63	560	30.75	371	26.81	259	22.62	449	24.24	464	30.77
Corrected scores [1]	526	23.58	560	23.35	371	25.31	259	25.02	449	25.74	464	29.87
Class 44-1-J (1) (CFTO):												
Raw scores	.....	.....	441	25.70	370	26.15	373	22.62	436	24.34	180	23.32
Corrected scores [1]	.....	.....	441	24.80	370	24.95	373	21.72	436	22.84	180	21.22

[1] Corrected to level of the third mission.

TABLE A11.5.—*Mean scores by time of day for air-to-air fixed gunnery*

Time of day	6-7		8-9		10-11		12-13		14-15		16-17	
	N	*M*	*N*	*M*	*N*	*M*	*N*	*M*	*N*	*M*	*N*	*M*
Class 44-1-G (EFTC):												
Raw scores	178	10.0	247	11.1	262	12.8	217	11.9	216	12.5	233	10.8
Corrected scores [1]	178	10.7	247	11.7	262	13.2	217	12.5	216	13.1	233	10.8
Class 44-1-H (EFTC):												
Raw scores	540	13.2	786	13.8	607	14.0	695	14.2	631	14.6	539	13.4
Corrected scores [2]	540	13.8	786	14.2	607	14.1	695	14.5	631	15.1	539	14.3

[1] Corrected to fifth mission level.
[2] Corrected to sixth mission level.

Effect of Wind Direction and Velocity on Air-to-Ground Fixed Gunnery Score.

A record of velocity and direction of the surface wind for each mission was entered on each student's individual record sheet along with the percent hits obtained for the mission. Since the wind data was obtained from instruments located in the gunnery field control tower, the system assumed that the tower wind was the same as that over a gunnery range several miles distant and at altitudes from 200 to 800 feet. Without doubt there was some error in that assumption; however, the wind conditions in the two locations should have been sufficiently similar to justify the study. No method of determining the surface winds was available at the ranges themselves.

Wind direction relative to the direction of dive on the target was divided into head, tail, right cross, and left cross wind. It was determined that a 30° variation in wind direction from the central line of each direction would not greatly affect the wind vector acting on the plane. This division of wind directions accounted for four arcs of 60° each, and left four intermediate arcs of 30° each for which the wind direction was judged to be ambiguous and so was not considered in this analysis.

After inspection of the data, wind velocities were divided into 6 categories of miles per hour: 1–5, 6–10, 11–15, 16–20, 21–25, 26–30. Left and right cross winds were combined after a preliminary analysis showed no difference in their effect. The mean percent hits for each velocity of each wind direction was then determined. Since some wind directions and/or velocities might have been more prevalent at one stage of training than at another, the scores were corrected for learning. This correction was a linear one since the assumption of a linear learning relationship was the only one which did not reduce the reliability of the corrected score.

Table A11.6 presents the mean percent hits by wind direction and velocity for the Pre and Post P–40 sections, separately and combined, for Class 44-1-J in the Central Flying Training Command.

TABLE A11.6.—*Mean percent hits by wind direction and velocity for two sections with and without prior P-40 training, separately and combined for class 44-1-J in CFTC*

Wind direction	Without P-40 training			With P-40 training			Combined		
Velocity	*N*	Uncorrected *M*	Corrected *M* [1]	*N*	Uncorrected *M*	Corrected *M* [1]	*N*	Uncorrected *M*	Corrected *M* [1]
Right and left crosswind:									
1-5	......	......	......	124	29.0	23.6	124	29.0	23.6
6-10	107	23.8	24.9	166	36.6	34.6	273	31.6	30.8
11-15	301	23.9	24.3	31	37.6	33.2	332	25.2	26.6
16-20	307	22.5	20.0	7	30.4	33.5	314	22.7	20.4
21-25	124	15.7	14.6	26	21.2	18.3	150	16.6	15.3
26-30	43	12.6	11.7	......	......	......	43	12.6	11.7
Head Wind:									
6-10	......	......	......	104	44.5	42.7	104	44.5	42.7
11-15	62	41.5	43.6	274	42.5	42.5	336	42.3	41.9
16-20	63	36.3	34.9	......	......	......	63	36.3	34.9
21-25	114	30.0	30.4	6	21.8	23.0	120	29.6	30.0
Tail Wind:									
1-5	......	......	......	41	21.0	27.2	41	21.0	27.2
6-10	103	26.5	30.8	187	22.3	29.0	200	23.8	28.6
11-15	37	22.2	26.3	......	......	......	37	22.2	26.3
16-20	101	18.7	18.0	......	......	......	101	18.7	18.0
21-25	......	......	......	28	20.8	19.0	28	20.8	19.0

[1] Corrected for learning.

Correction of Air-to-Ground Fixed-Gunnery Scores for Variation in Wind Direction and Velocity

The data used were those introduced in the preceding section, i. e., Class 44-1-J in Central. Using the data for the Pre P-40 section of that class, a set of curves of mean scores for wind direction and velocity were made. It was assumed from inspection of the data that cross and tail winds (as defined in the preceding section) had the same net effect on gunnery scores. Therefore, the two curves were combined to determine a new curve, which was projected to zero wind velocity.

Since no missions were flown in head winds of less than 11-15 miles per hour, it was necessary to estimate part of the head wind curve.

Two conversion tables were made on the basis of the above curves to correct individual scores obtained: (1) while flying in any given cross wind or tail wind and (2) while flying in any given head wind, to those that would be expected had the mission been flown in the optimum head wind. These conversions were equal to the actual difference in percent hits between the expected score for any given velocity and direction of wind and the optimum head wind. The assumption of a linear-type relationship was based on experimentation with several methods of correction. The assumption of linear relationship was the only one which kept converted scores within possible limits, i. e., did not predict scores of over 100 percent hits.

With these two conversion tables, the individual scores obtained by the students in the Post P-40 section of class 44-1-J were corrected for the wind direction and velocity in which they were flown. The odd-even reliability of these converted scores and also of the raw scores on which they were based was determined. Table A11.7 presents these data.

TABLE A11.7.—*Means, standard deviations, and reliabilities for the raw scores, scores corrected for wind direction and velocity, for post P-40 section, class 44-1-J, CFTC*

Type score	*N*	*M*	*SD*	r'_{II}	r_{II}
Raw scores:					
Odd	265	48.22	31.14	0.53	0.69
Even	....	60.33	31.99	....	....
Corrected scores:					
Odd	265	76.52	29.33	.47	.64
Even	....	88.31	29.94	....	....

TABLE A11.8.—*Correlation of instructor ratings of pilot proficiency and fixed-gunnery scores*

Type of gunnery	Gunnery scores			Instructor ratings		Correlation with rating
	N	*M*	*SD*	*M*	*SD*	
Air-to-ground	88	31.07	12.30	1.62	0.67	[1] 0.35
Air-to-air	88	....	4.70	1.62	.67	[1] .35

[1] Sign changed to indicate association of "good" performance.

Glossary[1]

SYMBOLS

$B_{01.234 \ldots n}$ = Beta weight.
C = Coefficient of contingency.
CR = Critical ratio.
F-ratio = Ratio of the larger variance to the smaller variance. See Analysis of Variance.
i = Class interval.
M = Mean (average).
N = Number of cases.
N_e = Number of cases eliminated.
N_g = Number of cases graduated.
N_l = Number of cases in lower group.
N_u = Number of cases in upper group.
P = Significance level.
P_e = Percentage eliminated.
ϕ = Phi coefficient (Pearson r from a 2 x 2 table).
P_g = Percentage graduated.
R = Multiple correlation coefficient, also number of right answers in a scoring formula.
r = Pearson product-moment coefficient of correlation.
r_{bis} = Biserial correlation coefficient.
ρ = Correlation coefficient corrected for restriction of range.
r_{pbis} = Point biserial correlation coefficient.
r_{tet} = Tetrachoric correlation coefficient.
r_z = Correlations combined by Fisher's z-transformation which is the correct method for averaging correlations.
r_{ptt} = Correlation between odd and even scores or two halves of a test without Spearman-Brown correction for full length.
r_{tt} = Reliability coefficient with Spearman-Brown correction or where correction is not necessary.
SD = Standard deviation.
σ = Standard deviation.
SE = Standard error.
W = Number of wrong answers in a scoring formula.
X = Raw score.

TERMS

AAF Regulation 50-3.—An AAF regulation which governs the issuance of instrument certificates to rated pilots. It specifies the requirements of the written and flight checks which all pilots are required to pass annually to maintain the instrument ratings which allow them to fly in bad weather.

Aerobatics.—Maneuvers such as loops, rolls, spins, etc.

Advanced Training.—The final 10-week phase of pilot training before the student receives his wings and commission. He receives instruction in heavier and more powerful aircraft and completes the formation, navigation, instrument, etc., phases which he began in basic training. (See Primary, basic.)

[1] Capt. William V. Hagin was primarily responsible for this glossary.

Aileron.—A control surface, a hinged segment of the trailing edge of the wing, which maintains the lateral stability of the aircraft. Moving the control stick raises one aileron and lowers the other, thus banking the plane to right or left.

Air Speed.—The velocity of an aircraft through the air mass.

Air-Speed Indicator.—An instrument which measures the speed of an aircraft through the air mass.

Altimeters.—Instruments which can be set to indicate the altitude, in feet, above sea level or specified ground check points.

Analysis of Variance.—Statistical methods for segregating from comparable groups of data the variance in the dependent variable traceable to specified sources.

Angle of Attack.—The angle at which the wing meets the air. Increasing the angle of attack, up to the stalling point, increases the lift produced by the wing.

Angle of Bank.—The angle between the horizon and a line passing through the wing tips. In level flight it is 0°, in a gentle bank it is approximately 15°, and in a vertical bank it is 90°.

Approach, Instrument.—An instrument flight procedure for descending through an overcast in order to make visual contact with the ground over a specified location.

Approach, Landing.—A straight glide toward the near end of the landing field or runway.

Artificial Horizon.—A gyroscopic instrument which shows the position of the aircraft with relation to the horizon.

A/S.—An abbreviation for air speed.

AT-6.—A single-engine, low-wing monoplane with tandem dual controls, retractable landing gear, power-operated flaps, and constant-speed propeller. Cruising speed 150–160 m. p. h. Used extensively as an advanced and basic trainer.

AT-9, AT-10, AT-17.—Two-engine, low-wing monoplanes with side-by-side dual controls, retractable landing gear, power-operated flaps, and constant-speed propellers. Cruising speed approximately 150–160 m. p. h. Used as two-engine advanced trainers until replaced by the TB-25.

Attenuation, Correction for.—A correction applied to estimate what the correlation between two variables would be if chance errors in the measurement of the two variables could be eliminated.

Attitude, Flight.—The position of the aircraft in relation to the horizon or ground. For example, when the nose of an aircraft is well above the horizon, it is said to be in a nose-high attitude.

Ball-Bank Indicator.—A dual-purpose instrument which has a free-moving ball for indicating slipping or skidding and a needle which shows rate of turn.

Bank, of a Plane.—See Angle of bank.

Base Leg.—See Pattern, traffic.

Basic Instruments.—Those instruments which are fundamental to maintaining safe flight when reference to the horizon and other normal cues is prevented by clouds, darkness, or the cloth hood.

Basic.—The intermediate 10-week phase of flying training which includes transition to heavier aircraft (BT-13, AT-6) beginning instrument flying, formation, and cross-country navigation. (See Primary, advanced.)

Battery, Test.—A group of tests covering a variety of subjects, or aspects, of a given field.

Beta Weight.—A numerical value indicating the relative importance of a single variable with respect to other variables when the several variables are combined to predict a criterion most effectively.

BT-13, BT-15.—Single-engine, low-wing monoplanes with tandem-dual controls, fixed landing gear, manually operated flaps, and two-speed propellers. Cruising speed 130–140 m. p. h. Used as the standard basic trainer until replaced by the AT–6.

CADETS, AVIATION.—Students in preflight or actual flying training leading to commission as rated air-crew officers. Commissioned officers taking the same training are designated "Student officers."

CENTRAL FLYING TRAINING COMMAND.—A regional branch of the AAF Training Command. (See chap. 2.)

CENTRAL INSTRUCTORS SCHOOLS (C. I. S.).—Schools which trained instructors for all the schools in a command and thus helped to achieve standardization of procedures and methods of instruction. Contrasted with local instructor squadrons which trained instructors only for their own school.

CHANDELLE.—A climbing turn in which the student was required to gain the most altitude possible without stalling while turning 180°.

CHECK LIST, PROCEDURE.—A list of specific operating procedures, carried in every plane, to be used by the pilot (1) before starting the engines, (2) during warm-up, (3) before take-off, (4) during flight, (5) before landing, and (6) after landing. All pilots were required to use the check list, regardless of how familiar they were with the equipment, to avoid errors of omission.

CLASS INTERVAL.—One of the groups into which the individuals measured on a given variable may be divided for convenience of tabulation and analysis. For example, in tabulating the heights of pilots in a school, they might be grouped in class intervals of ½ inch.

CLASSIFICATION TESTING PROGRAM.—The AAF aptitude testing program in which potential air-crew members were tested and ranked in nine categories, stanines, on their aptitude for pilot, bombardier, and navigator training.

COEFFICIENT OF CONTINGENCY, MEAN SQUARE.—A measure of degree of association between two variables, when each is expressed in several categories.

CONFIDENCE LEVEL.—See Significance level.

COMPASS, GYRO.—A gyroscope-stabilized compass which, when set with the magnetic compass, gives an accurate reading of heading which is not subject to the acceleration and turning errors of the magnetic compass.

COMPASS, MAGNETIC.—An instrument which indicates magnetic direction, or bearing.

CONTACT FLIGHT.—Flight conducted in such a manner that ground and water can at all times be used for visual reference.

COORDINATION.—The correct, simultaneous operation of the flight controls. The term is most often applied to the proper use of the aileron and rudder in rolling into or out of turns without slipping or skidding. It may, however, include the simultaneous use of additional controls such as throttle and elevator.

CO-PILOT.—The pilot, usually seated on the right side near the engine instruments, whose chief responsibility is serving as an assistant to the first pilot.

CORRELATION COEFFICIENT (r).—A numerical index of the degree of relationship between two variables. The correlation coefficient may vary from 0.00, indicating no relationship, to +1.00, indicating perfect agreement, and −1.00, indicating perfect inverse agreement. The term "correlation" or "correlation coefficient," used without qualification, implies the Pearson product-moment coefficient.

CORRELATION COEFFICIENT, BISERIAL (r_{bis}).—An expression of relationship between two variables, one of which is measured on a continuum (e. g. test scores) while the other is a dichotomy (e. g. success or failure on a task). If the distributions of scores are normal, the biserial coefficient is equivalent to the product-moment.

CORRELATION COEFFICIENT, MULTIPLE (R).—An expression of the relationship between one variable and the combined effect of several others so weighted as to maximize the resulting coefficient.

Correlation Coefficient, Pearsonian or Product-Moment (r).—An expression of relationship between two variables which are expressed in quantitative terms, i. e., a set of continuous measurements of the function or attribute concerned.

Correlation Coefficient, Phi (Φ).—An expression of relationship between two variables, neither of which are assumed to be continuous and both of which are expressed in dichotomies.

Correlation Coefficient, Point-Biserial ($r_{p.bis.}$).—An expression of the relationship between two variables, one of which is continuous and the other of which is expressed as a dichotomy and is not assumed to be continuous.

Correlation Coefficient, Tetrachoric (r_{tet}).—An expression of relationship between two continuous and normally distributed variables which are expressed in dichotomies. If the distributions are normal, the tetrachoric is equivalent to the product-moment.

Correlation Matrix.—A tabular arrangement of correlations of several variables with each other.

Crab.—See Drift Correction.

Criterion.—The standard by which tests or measures are evaluated.

Critical Ratio.—The difference between two comparable statistics (correlation coefficients, means, etc.) divided by the standard error of that difference. An index of the reliability of the difference.

Cross Wind Landing.—Landing of an aircraft other than directly into the wind. The drift caused by the side wind necessitates corrective action by the pilot in order to maintain a straight-in approach.

Curve of Pursuit.—The flight path of an aircraft attacking an aerial target which enables the attacking airplane to keep the target constantly under fire.

Directional Gyro (D/G).—See Compass, gyro.

Director of Flying.—The officer responsible for station flying training activities.

Director of Training.—The officer responsible for station training activities, including both ground school and flying training.

Distribution.—A systematic arrangement or tally of the range of scores based on magnitude. For example, a distribution of height measures on a population would show the number of cases falling in each unit interval ranging from the lowest to the highest.

Downwind.—Flight of an aircraft in the same direction as that of the wind.

Drift.—The angular difference between the aircraft heading, or direction it is pointing, and its track, or the path it is traveling over the earth's surface. This difference is due to the effect of the wind and its magnitude depends upon the direction and force of the wind.

Drift Correction.—To counteract the effect of wind drift (see Drift) the aircraft must be turned into the wind a sufficient number of degrees so that it will follow the desired path over the earth's surface. The number of degrees turned is called the drift correction, or crab.

Dual Flight.—Flight by a student pilot and instructor in aircraft equipped with dual controls.

Eastern Flying Training Command.—A regional branch of the AAF Training Command. (See chap. 2.)

Elevator.—The hinged portion of the horizontal stabilizer which controls climb or dive of the aircraft.

Elimination.—As used in this volume, relief from flying training for flying deficiency. Generally used to include relief from training for any cause.

Factor.—A basic trait or ability, the presence of which in combination with other traits or abilities, accounts for the relationship between two variables.

Factor Analysis.—A statistical technique which resolves a set of descriptive variables into a smaller number of fundamental categories or factors by an algebraic analysis of the intercorrelations of the variables.

Feel.—See Sustentation.

First Pilot.—The pilot, usually seated on the left side of the aircraft near the flight instruments, who is in command of and chiefly responsible for the operation of the aircraft.

Flaps.—A section of the trailing edge of the wing which may be raised or lowered to decrease or increase the lift of the airfoil. Flaps are most frequently used to assist in take-off, steepen the glide angle during landing, lower the landing speed of the aircraft, and reduce length of roll after landing.

Forced Landing.—An emergency landing caused by engine or structural failure, etc.

Formation Flying.—The flying of two or more aircraft in close proximity so that the several aircraft are flown, in effect, as a single airplane. Formation serves to concentrate offensive and defensive firepower. (See figs. 3.3, 3.4, and 3.5.)

Frequency Distribution.—See Distribution.

Ground Loop.—An aircraft with conventional landing gear is in relatively unstable equilibrium while traveling along the ground. Once it starts swerving to either side, it tends to continue turning in an increasingly tight turn, called a ground loop. The severe side thrust developed often causes the landing gear to collapse.

Gunnery, Fixed.—The firing of machine guns or cannon, rigidly fastened to the wing or airframe, at air or ground targets. Thus, the aircraft must be aimed to aim the guns.

Gunnery, Flexible.—The firing of movable machine guns or cannon at air or ground targets. These guns may be aimed independently of the direction of the aircraft.

Halo Effect.—The tendency of a rater to allow general characteristics to influence ratings of specific traits. For example, there is a tendency to give people one likes too high a rating on their weak points and people one dislikes too low a rating on their strong ones.

Hooded Flight.—Flight by reference to cockpit instruments using a cloth hood or similar device to confine the pilot's vision to the instruments.

Immelmann.—An acrobatic maneuver in which the aircraft is flown through the first half of a loop. Then, at the top of the loop, the plane is half-rolled from the inverted position back to level flight.

Instrument Flight.—Flight which cannot be made by visual reference to the ground or water and which requires the pilot to fly by reference to the cockpit instruments and other aids.

Instrument Let-Down.—See Approach, instrument.

Instrument Panel.—A panel directly in front of the pilot which contains all of the flight and engine instruments.

Item.—Individual test question or measure.

Item Analysis.—The statistical examination of responses to individual questions to determine their difficulty, their reliability, their discriminative value, or their relationships with other items or with the total score on the test.

Job Analysis.—A break-down of a task into its component parts or required abilities.

Landing Gear, Conventional.—The type of landing wheel arrangement found on most trainers and civilian aircraft. The two main wheels are ahead of the center of gravity of the aircraft and the third, or tail-wheel, is located toward the rear.

703322—47——32

Landing Gear, Tricycle.—The type of landing wheel arrangement found on most large aircraft. The two main wheels are behind the center of gravity and the third, or nose-wheel, is ahead of the center of gravity.

Landing, 3-Point.—A landing in which all three wheels touch the ground approximately simultaneously.

Landing, Wheel.—A landing in which the two main wheels touch the ground before the nose or tail wheel.

Lazy Eights.—A training maneuver in which the nose of the aircraft traces a horizontal figure eight, half of which is above the horizon, and the other half below.

Level of Significance.—See Significance level.

Link Trainer.—A training apparatus which consists of a mock-up aircraft equipped with a complete set of operating flight controls and instruments. Thus it is possible to simulate some of the conditions of instrument flight without leaving the ground.

Maneuver.—A finite unit, or portion, of the pilot's task such as a steep turn, a lazy eight, a loop, etc.

Mean.—The arithmetic average of a set of values.

Needle Ball.—See Ball-bank indicator.

Normalization.—The fitting of a given group of scores to a normal distribution.

Objective Measures.—Measures which are defined in terms of a permanent standard, such as a reference point on the airplane or an instrument reading, which is the same for everyone and thus produces absolute (in addition to relative) agreement among observers. (See chap. 4.)

Operational Training.—The pre-combat training given assembled air crews after graduation from the Training Command.

Partial Panel.—Instrument flight in which the flight indicator and directional gyro, or other instruments are covered so that they cannot be used for reference. It is more difficult than "full panel" because the instruments remaining do not give the instantaneous picture of flight attitude that is given by the gyro instruments.

Pattern, Traffic.—A rectangular pattern flown by aircraft landing or taking off from an airport. It is composed of four legs: take-off leg, cross-wind, down-wind, base leg, and final approach (both take-off and final approach are up-wind legs). The pattern is used to obtain uniformity of flight path and to alleviate traffic congestion. (See fig. 3.1, chap. 3.)

P. I. F.—Pilot's Information File.—A manual of general instructions and information for all pilots kept up to date by periodic revisions.

Primary Training.—The first 10-week phase of military flight training. Emphasis was placed on elementary flying abilities and judgment.

Procedure Checklist.—See Checklist.

PT-13, PT-17.—Single-engine biplane with tandem dual controls. Fixed landing gear, no flaps, and fixed-pitch propeller. Cruising speed approximately 90 m. p. h. Used extensively as primary trainers.

PT-19.—A single-engine low-wing monoplane with tandem dual controls, fixed landing gear, and fixed pitch propeller. Cruising speed approximately 90 m. p. h. Used as primary trainer.

Random Sample.—Sampling of a population on a chance basis as in Selective Service drawing for draft numbers.

Range.—The difference between the highest and lowest score for a group.

Rated Work Sample.—The performance is divided into a number of specific aspects or subtasks, each one of which is rated separately, for example, the student might be rated Excellent, Good, or Fair, on his use of throttle-to-check approaching stall in slow flying. (See chap. 4.)

Raw Scores.—The score as originally obtained from the test or other measuring device.

Reliable Difference.—A difference of such a magnitude that it is unlikely that it could have been produced by chance.

Reliability (r_{11}).—The consistency of a measure as determined in one of the following ways:

- **Test-Retest.**—determined by correlating two administrations of the test.
- **Odd-Even.**—determined by correlating the average of the odd-numbered items, trials, or measures with the average of the even-numbered ones.
- **Split-Half.**—determined by correlating first half of test with second half of test.
- **Split-Group.**—obtained by correlating ratings on the same individual made by two independent groups of raters.
- **Hoyt Method.**—determined by analysis of variance; does not require Spearman-Brown correction.

Reliability, Correction for Double Length.—The Spearman-Brown correction to estimate the reliability of the total group of items or judges from the correlation between two equal subdivisions of the group.

Restriction of Range, Correction for.—When the range of a population is restricted by selection, the correlation between two variables may be corrected for restriction of range in order to estimate what the correlation would have been if the range had not been restricted.

Rudder.—The hinged portion of the vertical stabilizer which maintains directional control of the aircraft.

Significance Level.—A statement of the statistical reliability of a finding. For example, the 5 percent significance level means that the result obtained would be expected to occur by chance only 5 (or fewer) times in 100; the 1 percent level means that result would be expected by chance not more than 1 time in 100.

Significant Difference.—A difference which is of such magnitude that it is unlikely that it could have been produced by chance.

Single-Engine Operation (2-Engine Aircraft).—The emergency operation of two-engine aircraft on only one of the engines.

Slow Flying.—A condition of flight in which the aircraft is flown between the power-on and the power-off stalling speed by the use of flaps and power. Useful in short-field landings, etc., where a slow-speed ground contact is desired.

Solo.—Flight by a student or students not accompanied by an instructor.

Spin.—A "top-like" rotation of an aircraft about its vertical axis which occurs when one wing is stalled and the other only partially stalled. The drag on the stalled wing is greater than that on the other, creating a strong turning tendency.

Stall.—A condition in which the smooth flow of air over the wing is disrupted so that the lift of the wing is decreased and its drag increased. Stalls are caused by too low an air speed in relation to the wing loading.

Standard Deviation (S. D.).—A measure of the degree of dispersion of a group of measures around the mean. One S. D. in both directions from the mean includes approximately 68 percent of the population in a normal distribution.

Standard Score.—A raw score converted to units of one standard deviation or some constant fraction of a standard deviation. By using standard scores, measures of dissimilar characteristics such as height and weight may be compared.

Standardization Board.—A Board of officers the purpose of which is to maintain standardized training methods.

Stanine.—Standard scores on a normalized 9-point scale. The highest 4 percent of the scores receive a stanine of 9, the next highest 7 percent a stanine of 8, etc., down through 12 percent, 17 percent, 20 percent, 17 percent, 12 percent, 7 percent, to the lowest 4 percent which receive a stanine of 1. Two of the common forms of stanine are:

PILOT.—A 9-point scale (based on the weighted scores of a battery of performance tests) for predicting the aptitude of students for pilot training.

INSTRUCTOR.—A score to predict aptitude for teaching flying.

STEARMAN.—See PT-13, PT-17.

STICK.—The cockpit control lever which operates the ailerons and elevators.

SUBJECTIVE MEASURES.—Measures which are made in terms of judgments such as "poor" or "too slow" which are dependent on the standards of the individual making the judgments. (See chap. 4.)

SUSTENTATION.—A feeling of support such as that experienced when seated in a chair. This feeling, although not reliable, helps the pilot maintain flight attitudes since changes in attitude are indicated by changes in feeling on the "seat of the pants." For example, in entering a climb, the pilot feels pushed down in the seat, while in leveling out or in beginning a dive, he feels lighter.

TACHOMETER.—An instrument which indicates the speed in r. p. m. Since engine power is a function of r. p. m. (with fixed-pitch propellers), the tachometer also indicates power output of the engine.

TAXI; TAXIING.—Ground operation of an aircraft as in going from the parking mat to take-off position.

TB-25.—The trainer version of the B-25 (Mitchell) Bomber. Lightened by being stripped of all combat equipment, the TB-25 was used as the standard 2-engine Advanced trainer during the last years of the war.

TEST BATTERY.—See Battery.

TEST-RETEST RELIABILITY.—The correlation between one administration and a second administration of the same test.

TORQUE.—To the pilot, the sum-total of all forces which tend to make the aircraft yaw to the right or left. These forces are of concern to the pilot because he must apply corrective rudder pressures in order to maintain desired flight attitudes.

TRAINING COMMAND.—An Air Forces Unit responsible for all Army Air Forces Training activities. (See chap. 2.)

TRAINING FREEZE.—In this report, the term "Training Freeze" refers to an occasion on which students were held at a given level for 5 additional weeks of training. (See chap. 10.)

TURBULENCE.—Roughness in the air due to vertical currents. Severe turbulence is found in thunderstorms and is frequently violent enough to damage aircraft.

UC-78.—Synonymous with AT-17.

UP WIND.—Into or against the wind.

VALIDITY.—The extent to which a test measures what it purports to measure.

VALIDITY COEFFICIENT.—A correlation coefficient (q. v.) which indicates how successfully a test or other selection device will predict performance with respect to a criterion. For example, the correlation of a score on a measure of flying proficiency with whether or not a student graduates from flying training will give a validity coefficient indicating how well that measure will predict graduation-elimination.

VERTIGO.—A "dizziness" or disorientation experienced by pilots flying at night or on instruments which causes them to feel that they are still turning after they have returned to straight and level flight.

WEIGHT.—The relative importance or value of items entering into statistical computations. (See Weighting.)

WEIGHTING.—The multiplying of different scores or measures by different constants so that the contribution of each measure to the total score will be proportional to its importance and the total score will have maximum predictive power.

Western Flying Training Command.—A regional branch of the Training Command. (See chap. 2.)

Wind Sock.—A cloth bag mounted so as always to be free to swing with the wind and indicate the direction from which the wind is coming.

Wind Tee.—A T-shaped marker at airfields which indicates the direction of landing traffic. Landings are made parallel to the leg and toward the arms of the T.

Wing Loading.—A measure of the amount of weight supported by each square foot of the wing's surface. For example, a wing of 100 square feet on an airplane grossing 1,000 pounds would have a wing loading of 10 pounds. For a given design of wing, the higher the loading, the faster the stalling (and hence landing) speed.

Work Adder.—A mechanical device for summating the number or extensiveness of movements of a control such as throttle, stick, or rudder.

Subject Index

U. S. GOVERNMENT PRINTING OFFICE: 1947