**Army Air Forces**
**Aviation Psychology Program**
**Research Reports**

# The Aviation Psychology Program in the Army Air Forces

**REPORT NO. 1**

*Edited by*
JOHN C. FLANAGAN
*Professor of Psychology*
University of Pittsburgh

1948

AD 655 590

# THE AVIATION PSYCHOLOGY PROGRAM IN THE ARMY AIR FORCES

John C. Flanagan

Army Air Forces
Washington, D.C.

1948

## NOTICE TO DEFENSE DOCUMENTATION CENTER USERS

This document is being distributed by the Clearinghouse for Federal Scientific and Technical Information, Department of Commerce, as a result of a recent agreement between the Department of Defense (DOD) and the Department of Commerce (DOC).

The Clearinghouse is distributing unclassified, unlimited documents which are or have been announced in the Technical Abstract Bulletin (TAB) of the Defense Documentation Center.

The price does not apply for registered users of the DDC services

# Preface

This series of research reports was planned to make available the major research findings and experience of the Aviation Psychology Program in the Army Air Forces during the period of World War II.

This Program owed its inception primarily to the vision and foresight of Gen. H. H. Arnold, Maj. Gen. David N. W. Grant, and Col. Loyd E. Griffis.

The preparation of these reports, and particularly of this introductory report, was carried on under the sympathetic leadership of Maj. Gen. Malcolm C. Grow and Col. Otis O. Benson, Jr., who succeeded General Grant and Colonel Griffis in the positions of The Air Surgeon and Chief of the Medical Research Division respectively.

The research findings briefly discussed in this report were the work of the officers and enlisted men who were affiliated with the Aviation Psychology Program during this period. Credit should go to these men for the quality and amount of this research.

For the source materials from which this report was prepared credit is due to the editors of the other reports in this series and the many officers and men who assisted them in compiling these reports.

In the immediate task of producing this report the writer was assisted especially by Dr. Tracy S. Kendler who, with the assistance of Miss Monnie Davis, collected the data for the tables and figures and supervised the preparation of the final forms of them. Miss Lucretia Bruni and Miss Ernestine Messick were responsible for typing the draft and preparing the final copy. All of these individuals worked many hours overtime in order that this report could be completed on schedule.

Parts of the first draft of the manuscript were read by Robert L. Thorndike, Merrill F. Roff, and Anthony C. Tucker and their comments were found very helpful in preparing this revision.

JOHN C. FLANAGAN,
*Colonel, Air Corps.*

WASHINGTON, D. C., *30 June 1946.*

# Contents

PART I

THE BACKGROUND AND DEVELOPMENT OF THE AVIATION PSYCHOLOGY PROGRAM

## PART II

### SPECIFIC SOLUTIONS TO PROBLEMS OF THE AVIATION PSYCHOLOGY PROGRAM

Chapter Page

## PART III

## GENERAL CONTRIBUTIONS TO AVIATION PSYCHOLOGY

# Part I

# The

# Background and Development

# of the

# Aviation Psychology Program

# CHAPTER ONE

# Introduction

## INTRODUCTION TO THE RESEARCH REPORT SERIES

This is the first of the series of Research Reports on the Aviation Psychology Program in the Army Air Forces. These nineteen reports were prepared through the cooperative efforts of a large proportion of the psychological personnel in the Aviation Psychology Program. In general, the editors of specific reports and the individuals assisting them in this task were selected as the individuals who during the period from 1941 to 1946 had worked most directly with the problems of the report. These editors have attempted to make full use of all original studies, analyses, and interpretations and have endeavored to integrate these specific findings into a unified and comprehensive report. These final reports therefore try to take account not only of the specific findings but also of the previously unrecorded experiences and insights which were gained by these psychologists through intensive study of the problems and the supervision of the design, execution, and interpretation of projects in this specific field.

The general objective of this series of reports is to present the complete story of procedures developed, predictive tests and evaluative techniques produced, and facts and relationships established. It was intended that through these reports a comprehensive statement of findings and results would be made available for planning future research and operating procedures in the Air Force.

In order that these reports accomplish this objective as effectively as possible, general plans for the preparation of the reports were developed in a series of three conferences attended by most of those individuals having editorial responsibility for each of the various reports. The first discussion of the reports took place at a conference in San Francisco early in August 1944. A memorandum sent out in October 1944 requested that preliminary outlines be submitted. At a conference at Lincoln Army Air Field in May 1945 tentative target dates were set for the completion of the various reports. On the 3d and 4th of August 1945 a special conference was held at Randolph Field for the purpose of developing final plans for the reports. At this conference each of the editors presented the detailed outline for his report. General matters of policy and various problems concerning the tentative style manual prepared by Headquarters AAF Training Command were discussed and agreed upon. The revised manual was reproduced and distributed as the "Style Book for Preparing Reports on the Aviation Psychology Program" by the Psychological Section, Headquarters AAF Training Command.

The editors were requested to submit copies of their manuscripts to Headquarters Army Air Forces, Headquarters AAF Training Command, and to other interested units. The Psychological Branch in Headquarters Army Air Forces examined all manuscripts with regard to considerations of policy, technical procedures, conflicts, and general contents. The officer in charge of psychological publications in Headquarters AAF Training Command examined the manuscripts primarily for style, organization, and expression. Other units examined specific portions of the manuscript primarily with regard to overlap and special knowledge concerning discussions of work for which these other units had been primarily responsible.

Although numerous suggestions for changes were made, the final responsibility for the content of the reports was in all cases that of the editor of the specific report. In only a few instances was it possible to discuss suggestions in person and in very few instances were revisions seen prior to the preparation of the mimeographed preliminary drafts.

The titles of the reports and the names of the officers in the Aviation Psychology Program primarily responsible for supervising their preparation and editing are listed below:

1. *The Aviation Psychology Program in the Army Air Forces*, John C. Flanagan, Professor of Psychology, University of Pittsburgh.
2. *The Classification Program*, Philip H. DuBois, Professor of Psychology, Washington University, St. Louis, Mo.
3. *Research Problems and Techniques*, Robert L. Thorndike, Associate Professor of Education, Teachers College, Columbia University.
4. *Apparatus Tests*, Arthur W. Melton, Professor of Psychology, Ohio State University.
5. *Printed Classification Tests, Parts I and II*, J. P. Guilford, Professor of Psychology, University of Southern California; assisted by John I. Lacey, Psychophysiologist, The Samuel S. Fels Research Institute, Yellow Springs, Ohio.
6. *The AAF Qualifying Examination*, Frederick B. Davis, Consultant, American Council on Education, Washington, D. C.
7. *Motion Picture Testing and Research*, James J. Gibson, Associate Professor of Psychology, Smith College.
8. *Psychological Research on Pilot Training*, Neal E. Miller, Research Associate, Yale University.
9. *Psychological Research on Bombardier Training*, Edward H. Kemp, Associate Professor of Psychology, University of Rochester; and A. Pemberton Johnson, Coordinator of Personnel Guidance, Purdue University.
10. *Psychological Research on Navigator Training*, Launor F. Carter, Assistant Professor of Psychology, University of Rochester.

11. *Psychological Research on Flexible Gunnery Training*, R. N. Hobbs, Assistant Professor of Education, Teachers College, Columbia University.
12. *Psychological Research on Radar Observer Training*, Stuart W. Cook, Research Director, Commission on Community Interrelations, New York, N. Y.
13. *Psychological Research on Flight Engineer Training*, John T. Dailey, Capt., AC, Psychological Research Project (Pilot)
14. *Psychological Research on Problems of Redistribution*, Frederic Wickert, Research Psychologist, Commonwealth Edison Company, Chicago.
15. *The Psychological Program in AAF Convalescent Hospitals*, Sidney W. Bijou, Assistant Professor of Psychology, Indiana University.
16. *Psychological Research on Operational Training in the Continental Air Forces*, Meredith P. Crawford, Professor of Psychology and Dean of the Junior College, Vanderbilt University; Richard T. Sollenberger, Associate Professor of Psychology, Mt. Holyoke College; Lewis B. Ward, Research Associate, School of Business Administration, Harvard University; Clarence W. Brown, Associate Professor of Psychology, University of California; Edwin E. Ghiselli, Associate Professor of Psychology, University of California.
17. *Psychological Research in the Theaters of War*, William M. Lepley, Associate Professor of Psychology, Pennsylvania State College.
18. *Records, Analysis, and Test Procedures*, Walter L. Deemer, Lt. Col., AC., Chief, Statistical Laboratory, AAF School of Aviation Medicine, Randolph Field, Tex.
19. *Psychological Research on Equipment Design*, Paul M. Fitts, Jr., Chief, Psychology Branch, Aero-Medical Laboratory, Wright Field, Ohio.

## THIS REPORT

This first report in the series is the over-all report on the Aviation Psychology Program in the Army Air Forces. It is not intended to be comprehensive or inclusive but rather to provide background information concerning the development of the program and a brief summary of the more important contributions to the solution of Air-Force problems and the major implications for future work. In presenting this report of findings and conclusions, no attempt has been made to insure that this summary provide interpretations representative of the ideas of all those concerned. Neither has any effort been made to have the conclusions conform to those given by the aviation psychologists who had immediate supervision over the specific projects reported. The views of these individuals have been recorded by them in the other reports of this series and the reader is referred to those reports for a more comprehensive and detailed discussion of the problems and findings. These other reports, which have provided the principal source

703323—47—2

of material for this report, are frequently q[illegible]ed in the pages that follow. To simplify presentation, quotation marks have not ordinarily been used.

This report consists of three main parts. The first part is devoted to (1) the planning and development of the program from a professional point of view; and (2) the history of the program in terms of military directives and official organization and structure. The second part presents briefly the important research findings and accomplishments of the program with respect to specific problems of the Army Air Forces during the war. The third part of the report contains a listing and brief discussion of the general contributions to theory, knowledge, and techniques in the general areas of activity of the personnel of the Aviation Psychology Program. These are presented as statements of tentative conclusions or implications for work in these fields and are accompanied by brief discussions of the evidence on which they are based. These are not necessarily the conclusions of all of the personnel concerned, but to a considerable extent are believed to represent generalizations accepted by most of those participating in this research program.

# CHAPTER TWO

# The Origins and Development of the Aviation Psychology Program

## INITIATION OF THE PROGRAM

In the spring of 1941 the great expansion in the training program for pilots and other air-crew specialties created problems which demanded the development of new procedures for selecting and classifying air-crew personnel. There were not enough trained flight surgeons to conduct the interviews on which Adaptability Ratings for Military Aeronautics had been based and the use of men selected solely on the basis of having been eliminated from pilot training had been found unsatisfactory for meeting the requirements of the bombardier and navigator training programs.

Besides this general situation there were a number of specific influences which caused the Chief of the Medical Division in May 1941 to recommend to the Chief of Air Corps the establishment of a Psychological Research Agency in the Medical Division. In recommending this research program on pilot selection, the desirability of selecting for both flying training and operational flying was pointed out by Maj. Gen. David N. W. Grant. He also cited selection research at the School of Aviation Medicine under the direction of the Medical Division, research in the United States Navy, and a project just initiated by the Training and Operations Division, Office, Chief of Air Corps, for research on the selection of bombardiers and navigators. The work of the Royal Air Force and the Civil Aeronautics Authority in this field was also known to the staff.

In commenting on the recommendation of the Medical Division, the Military Personnel Division opposed concurrence on the grounds that they were having enough difficulty filling quotas without adding additional requirements. They also stated that such research should be done at a time when the nation was not confronted with a national emergency. The comment of the Training and Operations Division recommended concurrence on the grounds that suitable tests could help to increase the production rate in training.

On 14 June 1941 the Chief of Air Corps, Gen. C. H. Brett, approved the recommendation of the Medical Division. The approval made possible the recruitment by the Medical Division of a "practical psychologist" and "suitable personnel" to develop and try out psychological tests in the Flying Cadet Replacement Centers.

# CHAPTER TWO

# The Origins and Development of the Aviation Psychology Program

## INITIATION OF THE PROGRAM

In the spring of 1941 the great expansion in the training program for pilots and other air-crew specialties created problems which demanded the development of new procedures for selecting and classifying air-crew personnel. There were not enough trained flight surgeons to conduct the interviews on which Adaptability Ratings for Military Aeronautics had been based and the use of men selected solely on the basis of having been eliminated from pilot training had been found unsatisfactory for meeting the requirements of the bombardier and navigator training programs.

Besides this general situation there were a number of specific influences which caused the Chief of the Medical Division in May 1941 to recommend to the Chief of Air Corps the establishment of a Psychological Research Agency in the Medical Division. In recommending this research program on pilot selection, the desirability of selecting for both flying training and operational flying was pointed out by Maj. Gen. David N. W. Grant. He also cited selection research at the School of Aviation Medicine under the direction of the Medical Division, research in the United States Navy, and a project just initiated by the Training and Operations Division, Office, Chief of Air Corps, for research on the selection of bombardiers and navigators. The work of the Royal Air Force and the Civil Aeronautics Authority in this field was also known to the staff.

In commenting on the recommendation of the Medical Division, the Military Personnel Division opposed concurrence on the grounds that they were having enough difficulty filling quotas without adding additional requirements. They also stated that such research should be done at a time when the nation was not confronted with a national emergency. The comment of the Training and Operations Division recommended concurrence on the grounds that suitable tests could help to increase the production rate in training.

On 14 June 1941 the Chief of Air Corps, Gen. C. H. Brett, approved the recommendation of the Medical Division. The approval made possible the recruitment by the Medical Division of a "practical psychologist" and "suitable personnel" to develop and try out psychological tests in the Flying Cadet Replacement Centers.

## PERSONNEL RECRUITMENT

On 15 July 1941 Dr. John C. Flanagan, Associate Director of the Cooperative Test Service of the American Council on Education, was commissioned as a major in the Specialist Reserve branch of the Officers' Reserve Corps and the following day reported for duty in the Medical Division, Office, Chief of Air Corps. Shortly thereafter he obtained approval to select an outstanding man with professional training and administrative experience to be commissioned as a major for the position of Director of Psychological Research Unit at each of the three Training Centers. After consultation with a number of psychologists, including personnel of the National Roster of Scientific and Specialized Personnel and the National Research Council, letters of invitation were written which resulted in the commissioning as majors on 28 August 1941 of Dr. Robert T. Rock, Head of the Department of Psychology and Director of the Psychological Clinic, Fordham University, and on 8 September 1941 of Dr. Laurance F. Shaffer, Professor of Psychology and Head of the Bureau of Measurement and Guidance, Carnegie Institute of Technology. Major Shaffer initiated the Psychological Research Unit in the Southeast Training Center in September and Major Rock, after spending nearly 3 months assisting in the Washington headquarters, activated the Psychological Research Unit in the Gulf Coast Training Center in November.

On 29 November 1941 the Air Staff approved an allotment of officer vacancies to be filled by direct commissioning from civilian life including 6 majors, 12 captains, 12 first lieutenants and 12 second lieutenants. On the basis of this authority 4 additional psychologists were commissioned at the rank of major. These included the commissioning on 14 January 1942 of Dr. Arthur W. Melton, Chairman of the Department of Psychology, University of Missouri; on 19 February 1942 of Dr. Frank A. Geldard, Professor of Psychology and Director of the Psychological Laboratory, University of Virginia; on 27 February 1942 of Dr. J. P. Guilford, Professor of Psychology, University of Southern California; and on 14 July 1942 of Dr. A. Paul Horst, Supervisor of Selection Research, Personnel Research Department, Proctor and Gamble Company. Thirty-five men were also commissioned as captains and lieutenants for this work under this allotment during the first 9 months of 1942. After that time the direct commissioning of civilians for this program was discontinued in accordance with War Department policy. During this period also a number of men with professional training in psychology and measurement who held commissions in the Officers' Reserve Corps were ordered to active duty or transferred from other assignments for work with this research program. During this recruiting period 17 of the group of outstanding professional men invited to apply for commissions declined, 9 others were disqualified for medical reasons, 3 were disqualified for other War Department policy reasons, and 6 applications were pending in September when direct commissioning was stopped.

During the fall of 1941 a search was made for enlisted men with training in psychology and measurement with the assistance of the specialist-assignment section in the Adjutant General's Office. A number of such men were located through this and other sources and sent to the Psychological Research Units. However, this source was inadequate for staffing the units and a new occupational specialty designated as Psychological Assistant (SSN 428) was established by obtaining approval for the publication of a supplement to AR-615-26 in Circular No. 14, 19 January 1942. Arrangements were made to have all men so classified sent directly to Psychological Research Units. To expedite this recruitment an announcement was made in the professional journals and applicants were requested to send a statement of qualifications to the Psychological Branch in Headquarters, Army Air Forces. Men who appeared to be qualified were sent a letter to this effect with instructions to show it to the Classification Officer at the time of induction or enlistment in order to facilitate their direct assignment to a Psychological Research Unit.

Authorization for 200 enlisted men at the Psychological Research Units was obtained on 15 January 1942. During 1942 and early 1943 the turnover in the units was quite rapid since practically all of these men were college graduates and almost half had master's or doctor's degrees. This procurement procedure made possible the direct recruiting of several hundred men with training in psychology for the Aviation Psychology Program. These men were well qualified for Officer Candidate Schools and were allowed to attend as rapidly as they could be replaced. Through this procedure and as a result of the program during the last year of the war for the direct commissioning of men with training in clinical psychology, more than 250 of the psychological assistants obtained commissions as second lieutenants. Some of these were subsequently promoted to ranks as high as lieutenant colonel. Rapid promotions were a little more likely in administrative work than in professional work in psychological units. Because of the limited number of vacancies available in the Aviation Psychology Program at the time these men received their commissions, it was possible to bring only a part of them back into psychological work in the AAF. Approximately 80 of them returned to the Aviation Psychology Program for service as officers.

In the fall of 1943 it was necessary to establish 7 new psychological examining units to process applicants before they entered the 5-month college training course which had been established. To staff these units a large number of additional officers and enlisted men was needed. Officers with appropriate professional training were obtained from all over the AAF. Quite a few of those transferred had been originally selected for duty as instructors in the training schools. To fill the need for a [illegible] men, 200 men just completing courses in Personnel Psych[illegible] under the auspices of the Army Specialized Trainin[illegible] [illegible]. Additional men were obtained by selec[illegible] [illegible] young men from among the applicants for flying training who were disqualified for medical reasons at the time of classification testing. During a later period qualified candidates who had to wait several months to obtain flying training were used while awaiting orders to flying schools.

By the end of June 1945 the Aviation Psychology Program included about 200 officers, 750 enlisted men, and 500 civilians. Most of this latter group were stenographers, typists, and clerks. About 30 civilians were employed in professional psychological and statistical positions under Civil Service. An additional group of 10 professional civilians was employed as special consultants for short periods.

## STATUS OF THE PROBLEM OF PILOT SELECTION IN 1941

The first project of the new Psychological Research Agency in the Office, Chief of Air Corps was to survey available information regarding the selection and classification of airplane pilots.

The first intensive research on problems of selecting personnel for aviation was done in connection with World War I. This research is reported in Air Service Medical[1] and the Report of the Psychology Committee of the National Research Council.[2] Much of the British experience appears in *Medical Problems of Flying.*[3]

In addition to these general references, a large number of publications by individuals relating to this topic were examined. Many tests were developed during the First World War but very few were evaluated adequately. The reports stressed tests of general intelligence, tremor and various types of reaction time.

During the twenties and early thirties, research and developmental work continued in some foreign countries and at the School of Aviation Medicine at Randolph Field in this country. Reports of this work were examined in Armstrong's *Principles and Practice of Aviation Medicine*[4], original articles in the *Journal of Aviation Medicine,* and especially in *Notes on Psychology and Personality Studies in Aviation Medicine.*[5] These references described the various complex reaction time and coordination tests developed during this period. For only four of the tests developed during this period are validity data reported for a substantial number of cases. On the basis of a follow-up of 1,274 students tested on the Thorne Reaction Time Test between 1926 and 1930 a small but significant relation was found between the test scores and graduation or elimination for failure to make satisfactory progress in flying training. The other tests all involved the use of a stick and rudder. In this country, the Complex Coordinator and the Automatic Serial Action Apparatus were developed and studied by O'Rourke,

[1] Air Service Medical, War Department, U.S. Government Printing Office, Washington, 1919.
[2] Report of the Psychology Committee of the National Research Council, *Psychol. Rev.*, 26, 1919, 83–149.
[3] Medical Research Council, London; *Medical Problems of Flying.* His Majesty's Stationery Office, 1920.
[4] Armstrong, Harry G. *Principles and Practice of Aviation Medicine,* Baltimore; Williams and Wilkins, 1939.
[5] *Notes on Psychology and Personality Studies in Aviation Medicine,* Technical Manual 8-320, War Department, Washington: Government Printing Office, January 1941.

Mashburn, Constable, and Glenn. The British test which was similar to the Complex Coordinator was developed by Reid and Flack. On the basis of follow-up of samples of approximately a thousand cases each in the late twenties and early thirties, it was found that each of these tests showed a biserial correlation coefficient with graduation or elimination in pilot training around 0.35.

By the end of the thirties, the international emergency had stimulated new research and development activities in Great Britain and the United States. There were also reports of considerable activity in Germany but few details regarding this latter research could be obtained.

It was interesting that studies conducted by the RAF, by the RCAF and by the National Research Council for the Civil Aeronautics Administration and the United States Navy all agreed in finding coordination tests involving the use of a stick and rudder, lathe type two-hand coordination tests, and pencil and paper intelligence tests predictive of success in pilot training. The latter type of test did not appear to be as stable from one sample to another as the other tests. One study, that done for the RAF by G. O. Williams and H. E. Whittingham[6] reported that in addition to the Sensory-Motor Apparatus No. 3 and the lathe-type, two-hand tests, a pencil-and-paper test of mechanical comprehension had predictive value. Several tests were being given routinely in the RAF in 1941 for the purpose of classifying men as either fighter pilots or bomber pilots. Unfortunately the research data regarding these tests were ambiguous and this classification testing was later abandoned. In the summer of 1941 the National Research Council was experimenting with a number of personality, interest, and biographical inventories for the CAA and the Navy. The Royal Canadian Air Force was making extensive use of the Link Trainer as a procedure for classifying men for pilot, observer, or gunner training.

On the basis of personal observation and study of the procedures in use at the Initial Training Wing in Toronto, it was decided that the use of the Link Trainer as a preselection device was too expensive, time-consuming, and difficult to standardize to warrant serious consideration in view of the promising results from simpler devices.

All of the groups mentioned above which were actively engaged in research on problems of the selection of aircraft pilots in the summer and fall of 1941, were very cooperative in making available all of the available data. Of special assistance were Dr. John G. Jenkins, Director of Research for the Committee on Selection and Training of Aircraft Pilots of the National Research Council (later head of the Aviation Psychology Division, United States Navy Department); Dr. Dean Brimhall, Director of Research for the Civil Aeronautics Administration; Commander E. Liljencrantz, Chief, Research Section, Division of Aviation Medicine, United States Navy Department; Dr. E. A. Bott and Dr. C. R. Myers, Canadian

[6] Whittingham, H. E. *J. R. Naval Med. Serv.*, Jan. 1940, 15-24; and Williams, G. O. Ph.D. Thesis London University, 1939.

National Research Council (later in Training Research, Air Ministry, Air Member for Training, Royal Air Force); and Dr. Ross McFarland of the Harvard Fatigue Laboratory and Consultant for Pan American Airways.

In summary it can be said that in the summer of 1941, there was evidence from a number of samples that certain apparatus tests and possibly one or two pencil-and-paper tests had predictive value for success in pilot training. However, the samples for the recently tested populations tended to be small and the results not entirely consistent. Much additional research seemed necessary before a satisfactory procedure for selecting pilots could be based on established relationships.

## INITIAL PLANS AND DECISIONS

In order to provide a general perspective from a professional point of view of the development of the Aviation Psychology Program in terms of problems and plans, the major decisions and broad policies which provided the basis for the organization and functions of the Program will be briefly sketched in this chapter.

The first important decision to be made was the decision to have a coordinated research program rather than a single, strong centralized group or a number of relatively independent units. The principal reasons were (1) the Army's natural framework which emphasizes the separation of policy-making and operation and (2) the decision that the research program should be intimately associated with the service or operating activities. It was believed that research personnel should avoid participating in activities not essential to research, but should be close enough to field activities to obtain a clear picture of operational problems.

A corollary of this decision was that the research program should be carefully planned and articulated. Responsibilities for specified areas of activity were clearly defined and delegated but an attempt was made to retain sufficient flexibility to make maximum use of special proficiencies, interests, and opportunities. The broad plan for research was in terms of areas of activity rather than specific projects since the process of developing procedures was regarded as a continuous evolution and final solutions were not expected. It was believed that the planning and the delegation of responsibility for specific areas of research should be based on comprehensive knowledge of the total situation. This should include such information as is usually available only in the highest headquarters concerning general plans, requirements, and problems, and the special abilities of the personnel available to do this work. At the same time it was important that full responsibility for details of execution be delegated to those doing the work.

A system was established for circulating plans for new tests or projects and reports, including progress reports, of research accomplished. The reasons for circulating these materials to all units of the Program were to avoid duplication of effort, to permit experts in other units to contribute to projects in which they were interested, and to keep personnel in all units informed concerning all pertinent research findings. These plans and reports were reviewed in the headquarters branches and priorities and deadlines were established. The aim of this supervision of the projects was to make sure that the work of the units was so planned that procedures and research findings could be expected which could be immediately translated into action. In addition to utilizing the best talent within the program by establishing procedures for coordination, plans were made to make use of the experience of outside groups and the advice of experts not in the Program.

Another early decision was that simple objective tests which could be administered to groups would be used wherever possible. This decision seemed necessary if such a large-scale testing program were to be kept within practical limits. Although the advantages in this situation of printed tests of the multiple-choice type were recognized, it was anticipated from the outset that these procedures would have to be supplemented by tests involving apparatus, motion pictures, and more subjective procedures, including an interview.

After a careful survey of available analyses of the pilot's job and facts and ideas about the importance of various traits for this job, it was decided to plan the test-development program around a group of 20 categories obtained from an analysis of reasons for elimination in primary flying training as reported by the academic boards. These categories were selected because they were comprehensive and included all traits obtained from surveys of other sources. These traits appeared especially relevant since they were described in terms of the language actually used by the flying instructors and check pilots. There was also some evidence that they were real requirements in that they were given to students as the reasons for their having been eliminated from flying training. For purposes of assigning responsibility for test development, the 20 categories were grouped into 4 principal areas, (1) tests of information, judgment, and intellectual ability, (2) tests of alertness, observation, and speed of perception, (3) tests of coordination and visual-motor skill, and (4) tests of personality, temperament, and interest. Personnel whose training made them especially suitable for test development in each of these four fields were sent to the appropriate psychological research unit to conduct such research.

Another decision was to try out tests as soon as they became available rather than to delay until all of the promising tests were ready and could be tried out in one group. This led to the assignment of Lt. Col. Laurance F. Shaffer to Maxwell Field on 21 September 1941. On 13 October 1941, examining was begun at the first psychological examining unit, which he established at Maxwell Field, Ala. On 17 November 1941 Lt. Col. Robert T. Rock reported to Kelly Field to establish Psychological Research Unit No. 2 and the following month research testing was initiated in that unit.

The final decision in developing the original plans for the Program was that all selection tests and procedures developed would be evaluated in the first instance against success or failure in primary flying training schools. It was believed that the decision to pass or fail a student was the most carefully made decision in the flying training schools. A student was failed only after several pilots independently observed his flying and agreed that his deficiencies were such that he could not be expected to become a satisfactory military pilot. There seemed little doubt that whether the particular student passed or failed was definitely influenced by the specific school, instructor, plane, equipment, weather conditions, and other factors in the local situation. However, it also seemed clear that there were stable and identifiable differences in individual students which were very important factors in determining whether they passed or failed. Observation of instruction and evaluation procedures in primary flying schools tended to confirm this decision.

## SELECTION AND CLASSIFICATION OF AIR-CREW

For some months it had been becoming increasingly obvious that it would be impossible to meet the quotas for men for pilot, bombardier, and navigator training if the requirement of 2 years college was retained. Certain groups in the Office, Chief of Air Corps, believed that it was very desirable, instead of trying to maintain an educational examination which established a "2-year college equivalent," to substitute a simple intelligence test which would enable the candidate to demonstrate his capacity to accomplish the work in the training school. As previously mentioned, the Training Division had initiated a research project on procedures for selecting bombardiers and navigators and in November on the basis of this research it was proposed that men with only high school graduation be admitted to bombardier and navigator training if they were able to pass certain tests of knowledge and ability.

It was obvious that this procurement problem and the pilot procurement problem were interrelated and could not long remain independent. Therefore, the Medical Division, which had been requested to develop an intelligence test for the selection of pilots, proposed that there was need for coordination of the various procedures being developed for different air-crew specialties. This division recommended that all research and test development relating to tests for the selection and classification of flying personnel be made the responsibility of the Medical Division of the Office, Chief of Air Corps. There was much discussion of this proposal.

As a result of the conferences which ensued, it was agreed that the Medical Division should prepare a selection and classification system consisting of two steps. First, the 2-year college or "equivalent" requirement would be discontinued and a general examination designed to reveal general potentialities, practical judgment, and capacity for absorbing instruction substituted for it. Advantages suggested for this proposal were: (a) a more uniformly proficient and able group of cadets would be assured by substituting demonstrated performance in handling reading material, diagrams, problems, and practical situations which were drawn directly from the tasks with which the cadet would later be confronted in his training for the more remotely related requirement of completion of 2 years college or success on academic-type examinations; (b) a larger number of applicants could be secured by using examinations for which specific training was unnecessary; (c) the length of examining time and also the time required for scoring and reporting the results could be reduced. After the initial testing program, it should be possible for the boards to administer and score tests immediately so that the applicant could be accepted, rejected, or conditionally accepted the same day that he made application.

It was also proposed that all applicants be enrolled as air crew without designation as to specific training assignments, and that classification for specific training assignment be made on the basis of more extensive tests administered in the Air Corps Replacement Training Centers.

These proposals were accepted by the Air Staff, and the Medical Division was made responsible for developing the necessary tests. In making these proposals, the psychological research group had clearly decided to abandon its purely research role and to recommend practical procedures for immediate use even though these procedures would necessarily be initially without adequate research foundation. This decision was made with little hesitation since by the time that the final decision had to be made this country was at war, and there seemed little justification for remaining aloof from the urgent and important practical problem of developing procedures to procure high-quality men for air-crew training immediately.

The initial screening test was to be given by untrained examiners in several hundred boards and the classification test battery was to be given by the three Psychological Research Units in classification centers to be established for that purpose. It was believed that because of the novelty of the classification program it would be necessary to combine the service testing aspects with the research activities at least during the stage of their development.

Since practically all applicants desired pilot training, a real problem existed in deciding which of the men should be sent into bombardier or navigator training. It was decided that aptitude scores would be calculated for each man from the scores obtained on a battery of air-crew classification tests. In obtaining the three aptitude scores the score on each test was weighted in accordance with the estimated relative importance and uniqueness of the factors measured by that test for the particular type of air-crew duty being considered. In this way an oustanding score on one important trait was allowed to compensate for a weak score on another important test or on each of several less important tests. The weighted averages were converted to scores on a scale running from 1 low to 9 high for which the intervals were half standard deviation units on a normalized scale. The average candidate obtained a score of five. For these aptitude scores the name "stanines" was coined to provide a ready and convenient form of reference to them.

In making recommendations for assignment, preferences and strengths of preferences as well as aptitudes were used. The relative strengths of preferences for the various types of air-crew training and the degree of willingness to be assigned on the basis of aptitude scores rather than preference were obtained from a specially prepared blank.

Although classification testing was initiated in all three of the flying training centers within 2 months after it was decided to establish such a program, nearly a year elapsed before the apparatus tests were available to establish a completely standardized program in all centers utilizing only tests which had been tried out and proved to be valid. During this initial period and for several months thereafter it was believed desirable to concentrate the available personnel in the Aviation Psychology Program almost exclusively on selection and classification problems for air-crew personnel. It was believed that a fairly adequate solution to this problem could and should be obtained before the personnel of the Program turned to other problems.

## PLANS FOR BROADENED SCOPE

By May of 1943, air-crew selection and classification was well established in the Army Air Forces. Although there was still a need for continuing research in the improvement of these procedures, work in this area necessarily led psychologists to an examination of criteria of success and objectives of training and operations. Exploratory studies had shown that an urgent need for psychological research existed in many of these other areas in the Army Air Forces. One of the most obvious of these was in relation to the evaluation of proficiency of flying personnel. For example, the grading procedures in primary flying training were such that an instructor after having only about six students to train for 2 months was unable at the end of this period to predict these students' success in basic flying training significantly more accurately than could be done on the basis of the tests taken by the students before entering flying training.

By this time also, the tremendous expansion in training facilities and staff was completed and training personnel were able to devote a little more time to trying to improve procedures in use. Thus, curriculum, methods, and devices and equipment were coming in for closer scrutiny than it had been previously possible to give them.

By the summer of 1943 also, there had been a sufficient amount of combat operation so that substantial numbers of combat-experienced personnel were being returned to the Zone of the Interior for duty assignment. Numerous problems were arising in connection with decisions as to the assignment of personnel to convalescent hospitals, the selection of some of them for duty as flying instructors, and general problems of morale and attitude created by the close association of these individuals with supervisory personnel who had not had combat experience. It was also believed very important to collect and utilize as efficiently as possible the knowledge and experience that these men had gained during their combat tours for the improvement of selection, training, and other administrative procedures.

To utilize the available psychological research personnel to the maximum extent in assisting with some of these important problems, a study was prepared and submitted to the Air Staff proposing a coordinated program of psychological research on problems of personnel procurement, classification, training, and redistribution. It was also proposed that an analysis of combat duties of air-crew personnel based on observation in theaters of operation was urgently needed. It was suggested that such a study would lead to the development and refinement of selection and classification tests designed for the prediction of success in combat as well as in training. This plan was approved by the Air Staff.

Accordingly, in August 1943 Lt. Col. Laurance F. Shaffer was assigned to the headquarters of the AAF Personnel Distribution Command for the purpose of initiating a program of psychological research on problems related to the reassignment of personnel returned from combat.

In September 1943, a preliminary conference was held at Headquarters, AAF Training Command to assign responsibility for research on training for each of the air-crew specialties. This led to the activation in January 1944 of Psychological Research Projects in each of the three instructor-training schools for pilots, bombardiers, and navigators.

In September 1943, following his return from the conference on the training research program, the Chief of the Psychological Branch recommended to the Air Surgeon that he be sent to study operations in the European and Mediterranean theaters.

## IDENTIFYING PROBLEMS REGARDING COMBAT OPERATIONS

Suitable arrangements were made with the theater commanders and on 22 November 1943 the Chief of the Psychological Branch of Headquarters Army Air Forces left Washington, arriving in Scotland the following day. The time from that date until 6 March 1944, which was the date of departure from the Mediterranean Theater, was spent visiting stations of the Eighth, Ninth, Twelfth, and Fifteenth Air Forces. All of the heavy bombardment groups which had been operating for more than 2 months in the Eighth Air Force were included in the survey. Air-crew personnel in fighter groups, medium bombardment groups, and ferry squadrons were also studied in these air forces. The survey was primarily concerned with the activities of air-crew personnel and units engaged in combat operations.

The principal purpose was to discover the extent to which known differences in aptitudes and other traits are reflected in the combat effectiveness of air-crew personnel. In addition to certain necessary headquarters stations, 37 operating groups were visited. During the major part of this period, the observer lived with the personnel in the groups and squadrons. Most of the time was spent with squadron commanders, squadron operations officers, squadron navigators, squadron bombardiers, and the corresponding officers in the groups. A large proportion of these officers had earned their advancement to these positions while engaged in combat operations and all of them were still active participants in actual combat operations over enemy territory. The information obtained from these officers who were immediately responsible for current operations was supplemented by observation and by discussion with the air-crew personnel under their supervision.

The following specific projects were included in this survey:

1. A follow-up of 1,616 officers who had been classified for pilot, bombardier, and navigator duty on the basis of aptitude tests given in Air Forces Classification Centers before assignment to flying training schools. The Squadron Commander and Squadron Operations Officer provided a brief description of the quality of the officers' work and also rated them on certain specific characteristics.

2. More than 300 of the officers in immediate supervision of air-crew personnel indicated on a check list of 20 aptitudes, abilities, and other personal characteristics, their judgments concerning the relative importance of these requirements for an individual capable of doing superior work of a specific type in combat operations.

3. The specific tasks performed by the personnel mentioned above and the skills and abilities necessary for obtaining good results were observed on various flights and operational missions, including a high-altitude group-formation training mission over England with the 91st Bombardment Group (B-17), an operational mission from England to Ludwigshafen, Germany, with the 100th Bombardment Group (B-17), a close-support fragmentation bombing mission in Italy on materiel and troop concentrations back of the northern enemy lines of the Nettuno beachhead with the 47th Bombardment Group (A-20), routine flights in B-25 and B-26 type planes, and a half-hour pickaback ride including simulated combat maneuvers in a P-38 with the Commanding Officer of the 1st Fighter Group.

4. The mission folders for groups in the Eighth Air Force on each visual bombing mission in which the particular group participated during the period 1 September 1943 to 31 December 1943 were examined and the results analyzed. The analysis was primarily concerned with determining the reasons why the bombs fell where they did with respect to aiming point. In establishing where the bombs fell, the strike photographs in the group mission folders were supplemented by the bomb plots prepared by the Operational Research Section of Headquarters, Eighth Air Force. In assigning the principal factor responsible for poor bombing accuracy the narrative reports of the lead bombardier and the lead navigator were supplemented with informal discussions with the group and squadron officers. Similar studies were made of bombing accuracy for groups in the Ninth and Twelfth Air Forces.

In the report[7] which was submitted to the Air Staff in April 1944, it was recommended that (1) psychological personnel be sent to England immediately to carry on combat validation studies and develop procedures for the selection of lead crews; (2) psychological personnel be assigned to the Continental Air Forces for the purpose of assembling superior individuals in each

[7] Flanagan, John C. *Report on Survey of Air Crew Personnel in the Eighth, Ninth, Twelfth, and Fifteenth Air Forces*, Washington: Headquarters, Army Air Forces, 1944.

of the air-crew specialties to be trained together as potential lead crews; (3) existing procedures for the classification of air-crew personnel be modified by substituting special aptitude scores (stanines) for fighter and bomber pilots for the single pilot stanine in use at that time and also obtaining special stanines for the various enlisted air-crew duty assignments; (4) certain new tests designed to measure combat requirements not adequately covered be developed for inclusion in the Air-Crew Classification Test Battery; and (5) psychological research in the AAF Personnel Distribution Command be focussed on the problem of determining the type of duty in which the returning air-crew personnel with combat experience can make the most effective contribution to the general war effort.

## AN INTEGRATED RESEARCH PROGRAM ON AIR-CREW PROBLEMS

In May 1944, the Air-Crew Evaluation and Research Detachment recommended was assembled and flown to the European theater to carry out research projects in the Eighth Air Force. At a conference in the AAF Training Command Headquarters in May 1944 procedures were developed for obtaining separate fighter-pilot and bomber-pilot stanines. Changes were also made in the tests and weights to be used in computing other stanines on the basis of the report of combat requirements.

Early in August 1944, a conference was held in conjunction with an Air Forces training conference at San Francisco, Calif. At this conference plans were developed for the work of the psychological research units which had been authorized for the Continental Air Forces by the Air Staff. Plans were also discussed for psychological research in the combat air forces.

One of the central problems for the Continental Air Forces units was to develop procedures for the selection of personnel for lead crews and other special types of operational training. It was also planned that additional data regarding special requirements of fighter pilots and bomber pilots, including necessary qualifications for airplane commanders, would be studied by these groups. Another important problem was the adjustment of combat returnees and their attitude toward a second tour of combat duty.

By the spring of 1945, reports on studies of combat operations from the three Air-Crew Evaluation and Research Detachments were available and rosters on microfilm of the aptitude-test scores of men assigned to air-crew training had been prepared. Proficiency tests had also been developed by the Training-Command research projects for each of the air-crew specialties and a center for the formation of combat crews was activated at Lincoln Army Air Field in Lincoln, Nebr. With the availability of these data and materials it was possible to develop the final procedures to be used in the formation of bomber crews. These were agreed upon at a conference of aviation psychologists at Lincoln Army Air Field in May 1945.

In the spring of 1945, research on classification-test development was concentrated on the preparation of an alternate battery of tests which it was planned to try out along with tests in current use. Substantial progress had been made by the Psychological Research Projects in the Training Command and the Psychological Branches in the Personnel Distribution Command on developing procedures for selecting returned personnel for instructor duty, and policies regarding a second tour of duty were established. Research in the Personnel Distribution Command was centered around the study of combat leadership and the identification and prediction of men developing anxiety reaction as a result of combat stress.

Also, in the late spring of 1945 plans were developed for establishing a Psychological Branch in the Aero Medical Laboratory at Wright Field to conduct research regarding the application of fundamental psychological principles to the design of equipment. Plans included the study of the individual's activities in the collecting and integrating of informaton from the instrument panel and other sources, the types of controls and movements which will insure maximum speed and precision in making necessary adjustments, and the basic facts concerning other human capacities affecting the design of equipment.

At a conference in early August 1945 at the School of Aviation Medicine, Randolph Field, plans were developed for the activation and staffing of the Psychological Branches in the Central Medical Establishments in combat theaters which had finally been authorized by the War Department. The personnel were selected and in many cases orders issued, but with the close of the war in Japan, the men were returned to their stations before being shipped overseas.

Following the close of the war, activities were largely confined to the preparation of comprehensive reports regarding the research findings and procedures developed during the war period. A substantial amount of time was also spent in planning the peacetime Aviation Psychology Program for the Army Air Forces. As recently adopted by the Air Staff, this plan delegates responsibility for all psychological research in the Air Forces to an integrated Aviation Psychology Program.

## CHAPTER THREE

# Official Directives on the Organization and Functions of the Aviation Psychology Program

### THE INITIAL PSYCHOLOGICAL RESEARCH AGENCY

The first official directive with regard to the Psychological Research Agency was the approval in June 1941 by the Chief of the Air Corps (1)[1] of the recommendation by the Chief of the Medical Division that a Psychological Research Agency be established in the Medical Division "for the purpose of conducting intensive research in flying-cadet selection." The next official act was the activation of the Psychological Research Agency in the Medical Division by the commissioning of Dr. John C. Flanagan, Associate Director of the Cooperative Test Service of the American Council on Education, and the ordering of him to active duty (2) on 16 July 1941 to develop the program of this organization. On 29 July 1941, approval (3) was granted to the director to procure the services of three outstanding professional men in testing work at the rank of major to serve as directors of the Psychological Research Units in each of the three training centers. The request by the director that the program of instruction in aviation cadet replacement centers be modified by setting aside 6 hours for psychological examinations in the processing period was approved by the Training and Operations Division (4) on 15 August 1941.

In August 1941 the Chief of the Air Corps, in commenting on a report by Maj. Harry Armstrong, regarding the selection of candidates for flying training directed (5) that an intelligence test be introduced into the original examination for aviation-cadet appointment. This directive was referred by the Military Personnel Division to the Medical Division on 29 August 1941, and on that date the Medical Division replied, "This Division is at the present time preparing an intelligence test to be applied for test purposes immediately in the reception centers when established."

An official directive on "Selection of Aviation Cadets" was sent from the Office, Chief of Air Corps (6) to the Commanding General, Southeast Training Center, Maxwell Field, Ala. on 2 October 1941 informing the latter of the initiation of a research project for the purpose of improving the methods of selecting and classifying aviation cadets. The administration of tests was begun at Maxwell Field on 13 October 1941. On the basis of a research con-

[1] The numbers in parentheses in this chapter refer to the numbered list in Appendix A of this report in which the precise dates, sources, and titles of all directives mentioned in this chapter are recorded.

703323—47—3

ference of aviation psychologists at Maxwell Field during the last week in October, an official directive (7) and plan for the research program on selection of aviation cadets was sent to the Commanding Generals of the Southeast and Gulf Coast Air Corps Training Centers on 26 and 27 November 1941 respectively. As stated in this directive "the immediate objective of the Program will be to determine by methods of psychological research the special abilities and psychological characteristics of aviation cadets that are associated with the subsequent success or failure in flight training."

The plan stated that each cadet would be given 4 hours of group tests and 2 hours of individual tests. The tests were to be selected by the Medical Division, Office, Chief of Air Corps in consultation with the directors of the Psychological Research Units. All test papers were to be returned to the Medical Division, Office, Chief of Air Corps for scoring and correlation with subsequent training records. A wide variety of tests was administered to several thousand aviation cadets at the Southeast and Gulf Coast Air Corps Training Centers under this plan during November and December 1941 and January 1942.

## THE AIR-CREW SELECTION AND CLASSIFICATION PROGRAM

### Authorizing and Initiating the Program

On 23 October 1941 the Medical Division proposed (8) that, "all research and test development relating to tests for the selection and classification of flying personnel be made the responsibility of the Medical Division, Office, Chief of Air Corps." On 19 November 1941 this division also proposed (9) that a general examination be substituted for the 2-year college educational requirement as soon as practicable. As the result of comments and conferences which followed these proposals the Chief of the Air Corps issued a directive (10) on 20 December 1941 assigning the following responsibilities to the various staff divisions. The text of this directive is quoted in full below:

1. In connection with the classification and selection of air-crew personnel, the following assignment of functions to divisions is made:

a. The Medical Division:

(1) The preparation of classification tests and all research connected therewith.

(2) The submission of recommendations regarding the adoption and use of such tests to the Military Personnel Division through the Training Division.

(3) The handling of matters pertaining to the supervision of the conduct of such approved classification tests as may be given to Army personnel under the jurisdiction of Air Corps training center commanders.

(4) The reprodurtion and supply of classification tests.

(5) The preparation of fiscal estimates and the procurement of personnel required for the execution of the foregoing functions.

(6) The preparation of statistical data showing the results obtained by the utilization of classification tests.

b. The Military Personnel Division:

(1) Approve with the concurrence of the Training Division all classification tests prior to utilization for selection purposes.

(2) Establish standards for qualifications with the concurrence of the Training Division.

c. The Training Division:

(1) Will be responsible for the allocation and scheduling of time to be devoted to the taking of both approved classification tests and research tests in Air Corps training center establishments.

(2) Initiate the action to secure data on student training performance as requested by the Medical Division.

2. In accordance with the contents of the preceding paragraph, it is recommended that the Medical Division immediately take over responsibility for research in connection with the project for the selection of bombardier and navigator trainees now being carried on by the Personnel Classification Section of the Air Corps Technical Training Command. This program is being carried out in conjunction with the Personnel Procedures Section of The Adjutant General's Office.

In recommending the use of a general examination in the memorandum (9) mentioned above, the director of the psychological research group stated, "The substitution of this general test for the present educational requirement for aviation cadets seems to be clearly justified. However, it is recognized that the setting up of requirements based chiefly on logical analyses of the duties of personnel has often led to the establishment of restrictions in the selection of incoming personnel which, however reasonable in appearance, are not later found to be justified when an actual comparison test is made. It is therefore believed that special tests for the selection of aviation cadets for specific types of training should be recommended only after these tests have proved effective in classifying the cadets of superior promise in successive classes containing several thousand individuals."

The final draft of the general examination was completed shortly after this directive was received and a general directive was prepared (11) on instructions for Aviation Cadet Examining Boards and requirements for appointment as aviation cadets. This was signed for the War Department by the Chief of Staff on 7 January 1942 and published in printed form and distributed to the Aviation Cadet Examining Boards on 10 January. This form of the Aviation Cadet Qualifying Examination was approved (12) by a special board appointed by the Commanding General of the Army Air Forces and put into official use in the Aviation Cadet Examining Boards on 15 January 1942.

The directive of 7 January 1942 changed the age limits to 18 to 26 inclusive, removed the restriction against married men, and substituted the Aviation Cadet Qualifying Examination for the 2-year college requirement. This directive also provided that all aviation cadets (air-crew) report to one of the three replacement centers where they would be given a complete physical examination for flying duty. The directive further stated, "Those cadets who pass the physical examination for flying duty, WD, AGO Form 64, will then be given further mental and psychological examinations for the purpose of determining for which type of air-crew training the cadet is best qualified. Upon the basis of these further tests, the aviation cadet (air-crew) is classified and specifically selected for either bombardier, navigator, or pilot training."

To accomplish this classification, the War Department approved and published a directive (13) on 4 February 1942. This directive provided the authority for activating a Psychological Research Unit at each of the Air Corps Replacement Training Centers. Psychological Research Units Nos. 1, 2, and 3 were activated at Maxwell Field, Kelly Field, and Santa Ana Army Air Base, respectively. Lt. Col. L. F. Shaffer, Lt. Col. R. T. Rock, and Col. J. P. Guilford were assigned as Directors of Psychological Research Units Nos. 1, 2, and 3, respectively. This directive also provided: "all enlisted men who are classified as psychological assistants at the time of induction will be assigned to Psychological Units in the Air Corps until the allotments are filled. All enlisted men so classified and assigned will be sent immediately to one of the Air Corps Replacement Training Centers. The Commander at the Replacement Training Center will be responsible for the enlisted man's basic training and assignment to the Psychological Unit at that station."

The School of Aviation Medicine was given the responsibility along with Psychological Research Unit No. 2 for developing and procuring apparatus tests of coordination, reaction time, and other psychological functions. Col. Arthur W. Melton was made Chief, Department of Psychology, Research Section, School of Aviation Medicine, Randolph Field. While awaiting approval of his application for a commission, Dr. Melton had served for several weeks as an expert consultant in the Psychological Branch in Washington.

On 2 February 1942, following a conference of the three Directors of the Psychological Research Units and personnel of the Psychological Branch in the Medical Division, Office, Chief of Air Corps, the first of a series of directives and plans (14) for the classification of aviation cadets was sent to the Commanding Generals of the Southeast, Gulf Coast, and West Coast Air Corps Training Centers. The directive stated that the objectives of the procedures established were: (1) to classify newly enlisted aviation cadets (air crew) for training as bombardiers, navigators, or pilots on the basis of psychological aptitude tests, measures of skill, knowledge, interests, physical qualifications, and other pertinent considerations; and (2) to collect data and carry out research relating to the construction, improvement, and development of psychological aptitude tests for this purpose. The directive also provided that "direct communication regarding technical matters between the heads of Psychological Units and the Medical Division, Office, Chief of the Air Corps through the Surgeon is authorized."

The plan provided that 2 days would be allotted for obtaining the information concerning aptitudes, preferences, skills, knowledge, and interests to be used in the classification of aviation cadets. Six hours were provided for the administration of printed tests and 2 hours were provided for the administration of apparatus tests requiring one set of apparatus for each cadet tested. The plan specified that the classification procedures would consist of five steps. These steps were as follows:

The first step in the classification procedure is the presentation of materials describing the respective duties of bombardiers, navigators, and pilots and giving specific information concerning the aptitudes and other characteristics necessary for successful performance of these functions. After reading these descriptions the cadet will be required to answer a number of questions about their contents to insure that he reads and understands them fully.

The second step for the cadet is to indicate his preference for one of the specific types of training based on his own analysis of his interests and aptitudes and his judgment as to the type of service in which he could render the greatest contribution to the Air Forces' effort.

The next step is the taking of individual and group tests including tests of graph reading, dial reading, table reading, line length, point distance, path distance, numerical operations, speed of identification, spatial orientation, mathematics, following directions, feel of controls, coordination, serial reaction time, and finger dexterity as supplied by the Office, Chief of the Air Corps.

On the basis of the results obtained as outlined in the preceding paragraph, predictive scores (ranging from 1 to 9) are assigned each cadet for each of the three types of aviation cadet (air-crew) training. In this process, previous training will be credited as follows:

(1) Solo Certificate; 2 points to be added to pilot score.

(2) Solo Certificate and thirty or more log hours; 3 points to be added to pilot score.

A specific recommendation as to type of training for each aviation cadet is made on the following basis:

(1) Cadets stating that they prefer assignment to the type of aviation-cadet training in which their aptitude is most outstanding will be so recommended.

(2) Those cadets whose predicted performance is above average (a grade of 6 or better) with respect to the type of training listed as their first preference will be recommended for their first preference.

(3) Of the cadets remaining unassigned after these steps, those whose predicted performance is above average with respect to their second preference will be recommended for that type of training.

(4) The third preference will be treated in the same manner for the remaining group and recommendations made.

(5) The remainder of the aviation cadets will be recommended for the type of training in which they make the highest score without regard for preference, except in the following two situations: first, if the aviation cadet indicates a higher preference for training as aviation cadet in one of the ground-crew categories and appears to have the necessary qualifications for such training, the necessary credentials for him should be submitted to the Office, Chief of the Air Corps for consideration for assignment to that specialty; second, if the aviation cadet indicates a higher preference for air-crew duty in an enlisted status than for the type of aviation-cadet training for which his predicted performance is highest, this preference should be confirmed by an interview and the recommendation should be made that he revert to enlisted status and be assigned to the preferred type of training.

(6) Those cadets assigned for a type of training for which their predicted performance is low (a grade of 4 or below) will be recommended for training probationally with the request that their performance be observed closely in the initial stages of training with a view to early elimination if they prove unsatisfactory.

At the same time a list of the tests to be used and a statement (15) of procedures for combining and weighting these tests in obtaining the three aptitude scores (stanines) were transmitted to each Psychological Research Unit.

### Defining Responsibility and Functions

From July 1941 through February 1942 the psychological research and development work carried on under the auspices of the Office of the Air Surgeon operated on the basis of directives previously referred to, prepared in the Office, Chief of Air Corps. After the War Department reorganization described in War Department Circular No. 59 (16), published 2 March 1942, the functions of that office were assumed by the Commanding General, Army Air Forces, and the Psychological Section, Research Branch, Medical Division, Office, Chief of Air Corps, became the Psychological Branch, Research Division, Office of the Air Surgeon, Headquarters, Army Air Forces.

It had been assumed that the Commanding General, Army Air Forces, was responsible for research and development work regarding procedures for the selection and classification of air-crew personnel. The propriety of this assumption was questioned by psychological personnel in the Adjutant General's Office of the War Department. On 13 April 1942 a directive (17) concerning classification and testing procedures for commissioned, warrant and enlisted personnel was approved and published by the War Department's Assistant Chief of Staff, G–1. This directive, while not specifically mentioning air crew, could be interpreted as giving the Adjutant General responsibility for developing classification procedures for air crew as well as all other military personnel.

A directive (18) was therefore prepared for the signature of the Assistant Chief of Staff, G–1, specifically concerning responsibility for the selection and classification of air-crew personnel. This was approved and published on 12 May 1942 and stated specifically that "the Commanding General, Army Air Forces will continue to have responsibility for and control of the selection and classification of military personnel for air-crew duty." The directive also charged the Commanding General, Army Air Forces with certain specific activities such as the development of practical examinations for use in the selection and classification of air-crew personnel.

The activation and staffing of the Headquarters AAF Flying Training Command early in 1942 made possible the delegation to that Command by Headquarters Army Air Forces of responsibility for the administration of the classification tests and the follow-up and analysis of the results. This was in accordance with the general policy of delegating all operating functions to field organizations, leaving only policy making and over-all supervision to the higher headquarters. AAF Regulation No. 35-24 (19) published 22 May 1942 listed those functions for which the Air Surgeon would be immediately responsible with respect to selection and classification of personnel for air-crew assignment and those functions delegated to the Commanding General, AAF Flying Training Command.

To provide the necessary psychological staff to carry out the functions delegated to Headquarters AAF Training Command by this directive, Col. Frank A. Geldard was transferred from the Psychological Branch in the Office of the Air Surgeon to the position of Chief, Psychological Section, Office of the Surgeon, Headquarters AAF Flying Training Command. Col. Geldard had been in charge of the Field Research Section in the Psychological Branch pending the activation of a fourth Psychological Research Unit.

Since it was decided to process all aviation cadets through three centers, this fourth unit was not needed. Since this directive delegated authority for "collecting classification-test results and data on success in air-crew training schools and performing the statistical analyses necessary to determine the accuracy of prediction of classification tests" to Headquarters AAF Flying Training Command, Lt. Col. Walter L. Deemer, Jr. was also transferred to that Command. Col. Deemer had been brought into the Psychological Branch to carry on this work and it was believed desirable that he continue to supervise it in its new location.

As general policy became clarified, and procedures for selection and classification became well established, further responsibility for psychological research was delegated to the Commanding General, AAF Training Command. In a directive (20) concerning the Aviation Cadet Qualifying Examination dated 1 September 1943, the responsibility for the preparation of additional forms of the Aviation Cadet Qualifying Examination under the general policies outlined by the Office of the Air Surgeon was delegated to the psychological research units under the Surgeon, Headquarters AAF Training Command. Similarly, responsibility for the immediate supervision and coordination of research and test development projects in the psychological research units of the AAF Training Command was delegated (21) to Headquarters AAF Training Command. This responsibility included the preparation of reports, the evaluation of results, and the preparation of recommendations for changes in the tests and weights for the Air-Crew Classification Test Battery.

## Modifying and Extending Procedures

Minor changes in the tests and weights used in the Air-Crew Classification Test Battery were made during the spring of 1942 as new tests became available. On 9 May 1942 a new directive and plan (22) was sent to Headquarters AAF Flying Training Command by the Chief of the Air Staff, Army Air Forces. At the same time the Psychological Division of the Office of the Air Surgeon prepared and distributed a new statement (23) concerning procedures for combining scores on Aviation Cadet Classification Examinations and a mimeographed booklet (24) containing directions for administering and scoring Aviation Cadet Classification Tests. On 3 June 1942 a directive (25) was sent to the Commanding General, AAF Flying Training Command, giving revised weights and directions for administering Aviation Cadet Classification Tests.

The printing of some new test booklets and the production of a number of copies of certain apparatus tests made it desirable to issue a new directive (26) from Headquarters Army Air Forces regarding classification testing procedures on 3 July 1942. Two major changes were made in classification procedures. It was stated, "Cadets with predicted scores of 4 or below for navigation training have almost invariably been found to be unsatisfactory material for navigation training. It is therefore directed that no one shall be assigned to navigation training who does not have a predicted score for navigation training of at least 5 (average).

"It has been established that much time is now being wasted by reclassifying cadets who have been eliminated from one type of air-crew school because of lack of aptitude, and attempting to train them for another type of air-crew duty for which they have even less aptitude. It is therefore directed that any cadet who has been eliminated from one type of air-crew training shall not be classified for another type of air-crew training unless his predicted score for this type of training is above average (at least 6)."

This directive, which established what became known as the July 1942 battery, was followed at periodic intervals by new directives establishing the December 1942 battery, the July 1943 battery, the November 1943 battery, and the September 1944 battery.

The changes in tests and weights for these subsequent batteries were based upon research findings and resulted from conferences of the directors of the psychological research groups and the chief of the psychological organization in Headquarters Army Air Forces.

The directive establishing the December 1942 battery (27) was prepared in Headquarters Army Air Forces following a conference at Headquarters AAF Flying Training Command and was dated 24 October 1942. By this time sufficient models of all apparatus tests had been produced so that the Air-Crew Classification Test Battery could be uniform for all three training centers. Also for the first time there was a substantial amount of data regarding the validities and intercorrelations of the various tests on which to base weights to be used in combining the scores. This directive also modified the procedures for making recommendations for training. This part of the directive is quoted as follows:

> After the preferences of cadets have been secured and their predictive scores for training in each of the specialties have been determined, the following procedure will be followed in making recommendations for training.
>
> a. The first step in determining the recommended classification of an aviation cadet will be to determine the air-crew duties for which he is psychologically qualified for air-crew training as follows:
>    (1) for bombardier or pilot training if his aptitude score for these types of training is 3 or above.
>    (2) for navigator training if his aptitude score for this type of training is 5 or above.
>
> b. The second step is to determine whether or not the cadet indicated on the Aviation Cadet Preference Blank (Form CE501E) that he desired assignment to the type of training for which he has the most aptitude. If the cadet did indicate such a desire, he will be recommended for the type of training for which he has the highest aptitude score, provided:
>    (1) that his aptitude score for such training is at least 3 points higher than for the type of training given the highest preference of those for which he is qualified, and
>    (2) that he confirms in an interview the desire to be so recommended.
>
> c. If a recommendation is not made under section "b" above, the next step will be to determine whether or not the cadet has an aptitude score of 3 or 4 (below aver-

age) for the type of training given the highest preference of those for which he is qualified. If the cadet did make such a score, he will be recommended for the specialty for which his aptitude score is highest, provided that

(1) if recommended for bombardier or pilot training, his aptitude score is 2 or more points higher for the recommended specialty than for his first preference,

(2) if recommended for navigation training, his aptitude score for this specialty is 3 or more points higher than for his first preference.

d. If a recommendation is not made under section "c" above, the cadet will be recommended for the type of training given the highest preference of those for which he is qualified.

All aviation cadets who are to be recommended for a type of training other than that for which they have the highest preference will be interviewed by an officer of the Psychological Unit. The purpose of the interview will be to verify the cadet's statement regarding assignment on the basis of aptitude scores, and to explain the reasons for recommending a classification other than the cadet's highest preference.

These procedures were modified somewhat by the Surgeon of Headquarters AAF Flying Training Command with the oral approval of the Air Surgeon. The directive (28) sent out by the Flying Training Command to the Classification Centers stated that the cadets not meeting these standards for any of the three types of air-crew training would be interviewed and brought before a board for final decision as to the Surgeon's recommendation. During the first few months after this directive was put into effect, these boards allowed a number of men with pilot stanines of 1 and 2 to enter pilot training. However, as the groups had more and more experience with these men, less and less of them were sent into pilot training until the number was practically negligible by the time that the July 1943 battery went into effect.

At a conference of the directors of the psychological research groups held in Washington at Headquarters Army Air Forces in March 1943, plans for the July 1943 test battery and procedures for preparing recommendations for assignment to training were discussed. It was agreed that results from the follow-up of several thousand men with low aptitude scores who had been sent into training definitely indicated that it was undesirable to qualify these men for flying training. The directive (29) issued by Headquarters Army Air Forces on 28 May 1943 specified that to qualify for pilot training a minimum pilot stanine of 4 was necessary. For navigator training a minimum navigator stanine of 6, and for bombardier training a minimum bombardier stanine of 4 together with a navigator stanine of 4 was necessary. The navigator stanine requirement for men to be trained as bombardiers was included because of the large number of men scheduled at that time for dual training as both navigators and bombardiers.

On 10 June the first indorsement (30) to this directive was signed by the Chief of Staff, Headquarters AAF Flying Training Command and sent to Headquarters Army Air Forces. This indorsement stated, "An insufficient number of students is being supplied to keep the college-training program up to full strength and there are apparently no indications that the procurement situation will be improved in the near future. .

* * * * * * *

"In view of the present aviation-cadet procurement problems which have already affected this Command (this Command is short 11,500 air-crew college trainees), it is recommended that the provisions in the basic communication raising the stanine requirement for pilot training from 3 to 4 be rescinded."

To clarify this situation, two studies were prepared by the Psychological Branch of the Office of the Air Surgeon. The first of these (31) was completed on 29 June 1943 and the contents were communicated to the Headquarters of the Flying Training Command by telephone. This study showed that as of 12 June 1943 the number of men recruited for the aviation-cadet training program and not yet enrolled in primary flying or other specialized training schools was 126,144. Of this number 56,395 were enrolled in the college program and the remainder were in basic training centers, classification centers, and preflight schools. It was pointed out that this figure was only 2,000 below the maximum permitted the Army Air Forces by the General Staff, that procurement since 12 June had been at a rate greater than that needed to meet current requirements, and that additional measures had been taken by Headquarters Army Air Forces to assure the continuance of the recent procurement rate.

It was further pointed out that "if the standards now prevailing in the Flying Training Centers are maintained, more aviation cadets will be graduated from Advanced Training Schools by entering 11,600 men with aptitude scores of 4 to 9 as provided in the basic communication than are now graduated by entering 12,100 with aptitude scores of 3 to 9. More graduates can be obtained by entering 500 men with aptitude scores of 4 to 9 than can be obtained from entering 1000 men with aptitude scores of 3. Even more important than this is the fact that in comparison with men having aptitude scores of 3, more than twice as many of the graduates with aptitude scores of 4 to 9 will be rated by their instructors as likely to become superior military pilots. This dilution of the quality of the graduates is a more important consideration in terms of the total war effort than the 500 additional eliminees whose time is being wasted."

On 7 July 1943 a second study (32) on this problem was prepared. This was sent to the Office of the Assistant Chief of Air Staff, Personnel and the Assistant Chief of Air Staff, Training and communicated by telephone to Headquarters AAF Flying Training Command. It recommended that the passing mark on the Aviation Cadet Qualifying Examination be established at the score of 180 which was approximately one-half standard deviation lower than the previous passing mark with the definite understanding that only men with pilot aptitude scores of 4 and above would be sent into pilot training by the AAF Flying Training Command. The study of the relationship between the Qualifying Examination score and pilot stanine indicated that, as compared with the existing procedure, the new plan would result in over-all increases in the numbers of aviation cadets sent into training and graduating of 14 percent and 17 percent respectively. Training reported on

15 July 1943 that it had not yet had an answer from Headquarters AAF Flying Training Command concurring in the proposal to raise the qualifying pilot stanine to 4 or above.

During the following week the Air Surgeon and the Assistant Chief of Air Staff, Training personally talked with officers at Headquarters AAF Flying Training Command by telephone. The Air Surgeon reported on 21 July 1943 that the Training Command had agreed to the standard of a pilot stanine of 4 with the reservation that men with lower aptitude would be sent into training if it was absolutely necessary to meet quotas.

A directive (33) was issued by Headquarters AAF Flying Training Command putting the procedures specified in the directive from Headquarters Army Air Forces to be used in preparing recommendations for air-crew training assignments in effect on 10 July 1943. However, this directive was modified to require a pilot stanine of 3 for qualification for pilot training rather than 4 as indicated in the original directive. On 12 August Headquarters AAF Flying Training Command issued a directive (34) instructing the Classification Centers to raise the qualifying pilot stanine to 4. This went into effect in the various units near the end of August.

Following a conference at Headquarters AAF Training Command in September 1943 a memorandum (35) outlining the data available at that time on the validation of classification tests for bombardiers, navigators, and pilots and proposing a battery of tests and weights was transmitted by the Surgeon, AAF Training Command to the Air Surgeon. This was approved (36) with certain minor changes and on 1 November 1943 the new battery was put into effect at the three Psychological Research Units and also at the seven new Medical and Psychological Examining Units which had been established at basic training centers to test applicants before they were sent to college training.

In order to activate these new psychological units, key personnel were selected from the existing three Psychological Research Units. The officers selected to head the new units were as follows:

Lt. Col. Lewis B. Ward, AAF Medical and Psychological Examining Unit No. 4, at AAF Basic Training Center No. 10, Greensboro, N. C.;

Lt. Col. A. C. Tucker, AAF Medical and Psychological Examining Unit No. 5, at AAF Basic Training Center No. 4, Miami Beach, Fla.;

Lt. Col. Frederic Wickert, AAF Medical and Psychological Examining Unit No. 6, at AAF Basic Training Center No. 2, Keesler Field, Miss.;

Lt. Col. Philip H. DuBois, AAF Medical and Psychological Examining Unit No. 7, at AAF Basic Training Center No. 1, Jefferson Barracks, Mo.;

Maj. Merrill F. Roff, AAF Medical and Psychological Examining Unit No. 8, at AAF Basic Training Center No. 3, Sheppard Field, Tex.;

Maj. Clarence W. Brown, AAF Medical and Psychological Examining Unit No. 9, at AAF Basic Training Center No. 8, Buckley Field, Colo.;

Maj. William E. Walton, AAF Medical and Psychological Examining Unit

No. 10, at AAF Basic Training Center No. 12, Amarillo Army Air Field, Amarillo, Tex.

The directive which was put into effect 1 November 1943 established minimum qualifying stanines of 6 for navigation training and 4 for pilot training. The bombardier qualifying standard was changed to a bombardier stanine of 5 together with a navigator stanine of 4.

The fact that a large surplus of trained pilots existed became evident in the fall of 1943 and the quotas for men entering into pilot training classes were reduced. The very large numbers of men available for pilot training made it possible at the same time to raise the qualifying standards for men sent into training. On 15 November the minimum qualifying stanines were set (37) at 5 for pilot, 7 for navigator, and 5 for bombardier together with a navigator stanine of 5. On 17 December the qualifying stanine for pilot was raised (38) to 6 and on 27 December the qualiyfing bombardier stanine was raised (39) to 6 together with a navigator stanine of 5. These standards remained in effect until the new classification battery of September 1944.

New tests and weights based on the minimum acceptable standards for air-crew personnel obtained from a survey of more than 300 of the squadron commanders and officers in charge of specialized personnel in Eighth, Ninth, Twelfth and Fifteenth Air Forces were contained in a directive (40) to the Commanding General, AAF Training Command signed by the Chief of Air Staff on 7 July 1944. This directive specified that stanines be computed for fighter pilot and bomber pilot separately. Besides the usual stanines for navigators and bombardiers, stanines were computed for three additional air-crew specialties. These were radio gunner, mechanic and armorer gunner, and career gunner.

The qualifying standards for fighter-pilot and bomber-pilot training were set at 6 and it was further specified that as many as possible of the fighter pilots should have stanines of 7 or higher. Navigator stanine remained at 7, and for bombardier training the single requirement of a revised bombardier stanine of 6 was set. The stanines for various types of gunners were to be used in classification of men for technical school and flexible-gunnery school training but the heavy demand for these specialists would not permit the disqualification of men because of low aptitude. This battery with these standards went into effect on 1 September 1944. By 24 October 1944, further reduction in quotas for training made it desirable to raise (41) the minimum qualifying scores for bombardier, navigator, and both types of pilot training to 7.

On 9 December 1944 this directive was modified (42) to require a career-gunner stanine of 6 or above for B-29 career-gunner training and an armorer-mechanic-gunner stanine of 6 or above for training as a B-29 remote-control turret gunner or a B-29 airplane-mechanic-trouble-shooter gunner. To qualify for training as a B-29 armorer-gunner, an armorer-mechanic stanine of 5 or above was required. On 23 December the directive was further modified (43) to introduce a new gunnery sighting test into the battery and modify the weights for armorer-and-air-mechanic gunner and career gunner to adapt them better to the selection of B-29 gunners. On 19 March 1945 a telegraphic directive (44) stated, "aptitude qualification for any B-29 gunner position will consist of aerial-(career) gunner stanine of 5 or better and no other psychological qualification will be required."

A directive (45) from Headquarters Army Air Forces to the Commanding General, AAF Training Command, dated 10 April 1945 stated, "It is desired that the radio-operator-gunner stanine, armorer-or-air-mechanic-gunner stanine and career-gunner stanine be replaced with an aerial-gunner stanine and a flight-engineer stanine." New weights were specified for the computation of these stanines.

The only other change made during the wartime program in the September 1944 test battery was the addition of a radar-observer stanine effective 1 June 1945 as directed (46) by Headquarters Army Air Forces on 23 May 1945. The minimum qualifying standards for both flight engineer and radar-observer were set at 7 at the time computation of these stanines was directed.

## THE RESEARCH PROGRAM ON INDIVIDUAL TRAINING AND COMBAT RETURNEES

Sending the first detachment of aviation psychologists to work at flying training schools was the result of a request (47) from the Director of Individual Training, Headquarters Army Air Forces, dated 25 August 1942. This request asked that the Office of the Air Surgeon set up the necessary classification and aptitude tests to select the most suitable men for training as low altitude (D-8) bombardiers from among the graduates of the Flexible Gunnery Schools. Research detachments were sent to three of the AAF Gunnery Schools from the three Psychological Research Units on 15 September1942. The officers selected to head these detachments at Las Vegas, Nev., Halingen, Tex., and Tyndall Field, Fla. were Maj. Clarence W. Brown, Maj Glen Finch, and Lt. Col. R. N. Hobbs, respectively.

In addition to their work on selecting low-altitude bombardiers, each of these detachments carried on research on the selection of flexible gunners and also on the training of flexible gunners. When the flow of men into low-altitude bombardier training was stopped after a couple of months, the officers in charge of flexible-gunnery training requested that a permanent Psychological Research Unit be established at one of the gunnery schools. On 17 February 1943, a psychological research detachment consisting of two of the officers who had worked in flexible-gunnery schools and several enlisted men with psychological training was established at Fort Myers, Fla. On 1 October 1943 this detachment was formally activated (48) as Psychological Research Unit No. 11. Lt. Col. R. N. Hobbs was made director of this unit.

In the meantime, aviation psychologists were becoming increasingly aware of the need for research on other aspects of personnel and training in the Army Air Forces. By this time, also, classification procedures were well established and the Psychological Research Units were fairly adequately staffed. Accordingly, in May 1943 a staff study (49) was prepared recommending that the psychological staff under the supervision of the Air Surgeon be utilized for a coordinated psychological research program on problems of air-crew personnel procurement and classification, training, and redistribution. The concurrence of the various divisions of the Air Staff was obtained for this recommendation and it was formally approved by the Deputy Chief of Air Staff on 21 July 1943. It should be noted that as originally written it was proposed to include research on problems related to the technical training of men for ground-crew duty as well as air crew. However, neither the Assistant Chief of the Air Staff, Personnel, nor the Assistant Chief of the Air Staff, Training, concurred in the inclusion of other than air crew.

In August 1943 the Psychological Branch prepared three directives related to this program. One (50) was to the Adjutant General for a study relating to procurement resources for air-crew personnel. Another (51) was to the newly established AAF Redistribution Center outlining a psychological program for this organization, and the third (52) was to the AAF Training Command proposing research on the selection of instructors, the evaluation of proficiency of men in flying training, the evaluation of the relative effectiveness of various training devices and procedures and the study of special psychological problems of instruction. The Air Surgeon decided that insufficient staff and facilities were available to carry out the study on procurement resources and this was abandoned.

The assignment to the AAF Redistribution Center of an aviation psychologist experienced in the Program was approved in August 1943. The Air Surgeon decided that rather than issue a specific directive on a psychological research program, an officer would be assigned to study the problem and to develop detailed plans for a psychological program in that organization.

Lt. Col. Laurance F. Shaffer was selected to develop and direct the psychological program in the AAF Redistribution Center and was made Chief, Psychological Division, Office of the Surgeon, Headquarters AAF Redistribution Center, in September 1943.

On 19 April 1944 Headquarters Army Air Forces directed (53) that the AAF Redistribution Center establish a psychological testing program with a unit at AAF Redistribution Station No. 2, Miami Beach, Fla. Lt. Col. Frederic Wickert was made Chief, Psychological Branch, Office of the Surgeon, at this station.

On 12 August 1944 a directive (54) was issued from Headquarters Army Air Forces establishing a psychological program in all Redistribution Stations primarily for the purpose of administering instructor-selection tests. Lt. Col. St. Clair A. Switzer was selected to be Chief, Psychological Branch, Office of the Surgeon, AAF Redistribution Station No. 1, Atlantic City; Maj. Merrill T. Hollinshead was placed in charge of the branch at AAF Redistribution Station No. 3, Santa Monica; Maj. Chester W. Harris was selected for the similar position at AAF Redistribution Station No. 4, Santa

Ana Army Air Base; and Maj. Frederick B. Davis was chosen for AAF Redistribution Station No. 5, Camp Davis. To coordinate the work of these branches under the general supervision of the Chief, Psychological Division, Maj. Merrill F. Roff was brought into Headquarters AAF Redistribution Center.

On 1 June 1944 the AAF Redistribution Center was reconstituted as the AAF Personnel Distribution Command and on 28 October 194: the Aviation Psychology Program in AAF Convalescent Hospitals was formally established (55). At that time Psychological Services Branches, under the technical supervision of the Psychological Division, Office of the Surgeon, Headquarters AAF Personnel Distribution Command, were organized in all AAF Convalescent Hospitals. To develop and coordinate this program under the supervision of the Chief, Psychological Division, Capt. Sidney W. Bijou was brought into Headquarters AAF Personnel Distribution Command. The officers placed in charge of the Psychological Services Branches are listed below together with the name of the station at which the AAF Convalescent Hospital was located:

| *Name* | *Station* |
|---|---|
| Capt. Paul R. Diller | Pawling, N. Y. |
| Capt. Donald E. Super | Miami Beach, Fla. |
| Maj. Harry V. McNeill | Fort Thomas, Ky. |
| Maj. Glen L. Heathers | Bowman Field, Ky. |
| Lt. Col. Hermann O. Schmidt | Fort Logan, Colo. |
| Lt. Col. Henry Obel | Fort George Wright, Wash. |
| Lt. Col. Lee E. Travis | Santa Ana Army Air Base, Calif. |
| Maj. Milton B. Jensen | Camp Davis, N. C. |
| Lt. Col. George Forlano | St. Petersburg, Fla. |
| Maj. Edward I. Strongin | Plattsburg Barracks, N. Y. |
| Lt. Col. R. N. Hobbs | Cochran Field, Ga. |

The Air Surgeon transmitted the directive on training research to the Surgeon of the AAF Training Command informally rather than through official correspondence channels. The Surgeon of the AAF Training Command indicated in discussions with the Air Surgeon that he concurred in its general provisions and would proceed along those lines without the necessity of a formal directive. Plans for increasing the proportion of time devoted to training research were discussed at a conference including representatives of all of the Psychological Research Units at Headquarters AAF Training Command in September 1943.

One of the first units to devote a large portion of its time to training research was the Psychological Test Film Unit which was activated (56) at Santa Ana Army Air Base on 9 October 1943. Lt. Col. James J. Gibson, who had been responsible for this type of work in the Psychological Section of Headquarters AAF Training Command, was made director of the new unit. In addition to developing tests for tryout as classification tests, this unit conducted extensive research on methods of aircraft recognition training and also carried out studies on effectiveness of motion picture training films.

On 15 January 1944 Psychological Research Projects were activated (57) for research on bombardier training at Midland Army Air Field, navigator training at Selman Field, and pilot training at Randolph Field. The officers selected as directors of these special research groups were: Lt. Col. Edward H. Kemp, Psychological Research Project (Bombardier); Maj. Launor F. Carter, Psychological Research Project (Navigator); and Maj. Neal E. Miller, Psychological Research Project (Pilot).

On 1 December 1944 a similar Psychological Research Project was activated (58) to study problems in connection with radar-observer training at Langley Field. This unit actually began functioning on 4 October 1944 when Maj. Stuart W. Cook who had been selected to direct the work of this project reported for duty.

Psychological Research Project (Combat Crew) was established (59) on 6 April 1945 at Lincoln Army Air Field for the purpose of administering proficiency tests and other evaluative devices to combat-crew personnel and the collating of training records leading to recommendations for the assignment of combat-crew and the designation of potential lead-crew material. Lt. Col. William M. Lepley was made director of this project.

The last Psychological Research Project to be activated was the one to study flight-engineer training at Hondo Army Air Field. This was formally activated (60) on 1 July 1945. Prior to that date the work was carried on by a detachment from Psychological Research Unit No. 2. Maj. Neil D. Warren was made director of this project.

## RESEARCH ON PROBLEMS OF OPERATIONAL TRAINING AND COMBAT

### Temporary Duty Detachments in Combat Theaters

The first official recognition of the importance to the Aviation Psychology Program of the AAF of the study of combat operations was the issuing of orders (61) 10 November 1943 directing Col. John C. Flanagan, Headquarters Army Air Forces, to proceed "from Washington, D. C. to the European theater and Mediterranean theater, reporting to the Commanding General of each theater on temporary duty approximately ninety (90) days." The mission of this temporary duty was to survey and study "air-crew personnel classification matters."

Following this officer's return to Washington on 8 March 1944 a memorandum (62) was submitted to the Chief of the Air Staff, recommending that research psychologists be sent to assist in the selection of personnel for lead crews in the United States Strategic Air Forces in Europe. On 24 March the Chief of Air Staff directed (63) that a cable be sent "to the Air Force Commander in the Theater explaining the object of the mission and asking whether or not he desires to have them come."

This telegram (64) was dispatched on 27 March and on 3 April a cable (65) was received from the Commanding General, United States Forces in the European Theater of Operations, London, England. This cable stated that the project was favorably considered except that the work in the theater should be limited to research for the purpose of developing aptitude tests under actual combat conditions and evaluation of air-crew personnel with respect to aptitude for combat as an aid to the selection of lead crews. The cable further stated that after such tests were developed they should be given in the training stations in the Zone of the Interior.

With the concurrence of the AAF Training Command, orders (66) were issued 18 April 1944 directing 6 officers and 15 enlisted men to proceed to Greensboro, N. C., reporting not later than 25 April 1944 to await overseas shipment. On 27 April 1944 the Air Surgeon issued a directive (67) defining the mission of this Air-Crew Evaluation and Research Detachment to the detachment commander, Lt. Col. A. Paul Horst. The Chief of the Psychological Branch delivered this directive to the Detachment Commander in person at Greensboro, and together with the Chief, Psychological Section, Headquarters AAF Training Command, spent 2 days discussing plans with the personnel of this detachment. The principal provisions of the directive are quoted below:

> The training and classification-test records of personnel now assigned to bombardment groups in the Eighth Air Force are to be collated from the basic AAF Training Command records which have been compiled for the detachment. These records are to be compared with the combat performance of men in such key assignments as lead-crew positions to determine the value of these records to squadron and group commanders in selecting men for these assignments.
>
> To supplement these records the detachment is to administer to appropriate combat personnel certain new tests developed as a result of the recent survey of combat operations conducted by a representative of this office. These results are also to be validated against combat performance.
>
> On the basis of the findings from the above studies the detachment is to revise and develop tests for the purpose of predicting combat performance as accurately as possible at the time air-crew personnel enter operational training units in the Zone of the Interior.
>
> This detachment is to report progress and results achieved directly to this office every 15 days so that the findings may be utilized in the training units in the Zone of the Interior.

The temporary duty orders of this group were extended (68) and on 7 September 1944 the detachment commander and 2 officers returned to Washington with the remainder of the detachment following 2 or 3 days later.

Two additional Air-Crew Evaluation and Research Detachments were sent to overseas theaters The directives (69) (70) on the missions of these two Air-Crew Evaluation and Research Detachments were very similar to that for the first Air-Crew Evaluation and Research Detachment quoted above. However, the latter two detachments did not develop or administer tests except for tests of the proficiency of various air-crew members.

Near the end of August a cable (71) was sent to the Fifteenth Air Force requesting concurrence in sending research personnel to that theater on temporary duty. On 11 September 1944 a cable (72) was received from the Commanding General, Allied Force Headquarters, Casserta, Italy, requesting these aviation psychologists and on 2 October, 4 officers and 8 enlisted men under the command of Maj. Neil D. Warren left this country for the Fifteenth Air Force.

Later the orders of this group were amended (73) to permit them to visit the Twelfth, Ninth, and Eighth Air Forces in addition to the Fifteenth Air Force and their temporary-duty period was extended to 6 months. The detachment returned from England, arriving in New York on 5 May 1945.

As a result of a cable (74) sent to the Commanding General, United States Army Air Forces, Pacific Ocean Areas, Hickam Field, Hawaii on 2 November 1944, concurrence was obtained for sending a detachment of 4 officers and 8 enlisted men under the command of Lt. Col. William M. Lepley on 90 days' temporary duty to the Pacific Theater. This detachment arrived in Hawaii on 16 December 1944, visited units of the Seventh and Twentieth Air Forces, and returned to Hamilton Field, California, on 16 March 1945.

### The Program in the Continental Air Forces

At the same time that arrangements were being made to send Air-Crew Evaluation and Research Detachments to the combat theaters, the approval of the Air Staff was being obtained for sending aviation psychologists to the four Air Forces in charge of operational training in the Zone of the Interior and to the Air Forces in the combat theaters on a permanent basis. On 26 June 1944 the Air Staff approved the recommendation (75) to assign suitable trained personnel to each of the Air Forces to develop the Psychological Research Program recommended as the result of the initial survey of air-crew personnel in the European and Mediterranean theaters.

On the basis of this authority, plans were made for the permanent assignment of groups of approximately 4 officers and 6 enlisted men each to the First, Second, Third, and Fourth Air Forces which were responsible for operational training. A conference of the officers selected to head these units together with officers from research units in the AAF Training Command, the AAF Personnel Distribution Command, and Headquarters Army Air Forces was held early in August in San Francisco.

The officers chosen to direct this work and made Chiefs of the Psychological Branches, Office of the Surgeon, in the headquarters of the various air forces were: Lt. Col. Richard T. Sollenberger, Headquarters First Air Force; Lt. Col. Lewis B. Ward, Headquarters Second Air Force; Maj. Clarence W. Brown, Headquarters Third Air Force; and Lt. Col. Edwin E. Ghiselli, Headquarters Fourth Air Force. After the delegation by Headquarters Army Air Forces of the direct supervision of these operational training air forces to a new headquarters created for this purpose and called Headquarters Continental Air Forces, Lt. Col. Meredith P. Crawford was made Chief, Psychological Branch, Office of the Surgeon of that headquarters to coordinate the work of these units.

On 23 August 1944 directives (76) were sent to Commanding Generals of the First, Second, Third, and Fourth Air Forces outlining the mission of the psychological research officers being assigned and giving some of the background of this project. These directives stated:

> The mission of these officers is as follows:
>
> a. To collect, assemble, and make available to commanding officers, classification-test scores, stanines, proficiency-test scores, training records, and combat-adjustment scores, and to advise concerning the use of such information in the selection of personnel for lead crew and other special types of operational training.
>
> b. To administer aptitude and/or proficiency tests when indicated for the purpose of securing additional information for the selection of personnel for special training.
>
> c. To collect and/or develop criterion data on the proficiency of individuals in carrying out various operational duties. Such data will be used in refining present testing and selection procedures or will be forwarded to Headquarters AAF Training Command for use in validation studies to be conducted by that Headquarters.
>
> d. To develop new tests or test specifications designed to measure aptitudes and proficiencies that are important for success in combat and that are not measured by existing tests.
>
> e. To undertake studies of attitudes, motivation, and leadership ability as indicated by local problems or as directed by Headquarters Army Air Forces.
>
> It is suggested that in order for these research officers to perform the above mission they be given access to training, proficiency, and operations records in your command and that they be authorized to interview personnel and to collect information concerning the proficiency of flying personnel insofar as those activities do not interfere with training activities. It is further suggested that the research officers be authorized to participate in aerial flights and to observe air-combat-crew training activities when necessary for the accomplishment of this mission.

On 3 October 1944, Manpower Division of Management Control, Headquarters Army Air Forces, in reply (77) to a request from the Office of the Air Surgeon for approximately one lieutenant colonel, one major, two captains, and six enlisted men for each Air Force stated, "Necessary action has been taken to increase bulk allotments of the four Continental Air Forces for aviation psychologists and personnel-consultant requirements as requested." Personnel were assigned to these Air Forces shortly after the first of October.

### Permanent Research Units for Combat Air Forces

The change in table of organization for the training air forces did not have to be approved by the War Department General Staff since the Air Forces had a bulk allotment for all establishments within the Zone of the Interior. The approval of the Air Staff resulted in the appropriate action being taken. It was necessary, however, to obtain War Department approval for changes in the Tables of Organization of the combat air forces.

After considerable discussion of methods of providing for the permanent assignment of research officers to Air Forces in combat theaters, the Troop Basis Division of Headquarters Army Air Forces sent a memorandum (78) on 19 September 1944 to the Assistant Chief of Staff (Attention G-3 Division). After reviewing the need for such personnel, this memorandum stated, "To provide for supervision of lead-crew selection by personnel specially trained for such a job, it is desired by this Headquarters to furnish an aviation psychologist (2251) to each Air Force Headquarters and an aviation psychologist and two enlisted men to selected wing headquarters. It is estimated that approximately 40 officers and 80 enlisted men would be required and provision can be made for this number within the present Army Air Forces Troop Basis.

"Informal discussion with representatives of your Division (Colonel Shugart and Major Moss) indicates that G-3 does not favorably consider a change to appropriate TO and E's to accomplish this objective. It is therefore recommended that this headquarters be permitted to initiate action to authorize personnel by special allotment and equipment by special lists as indicated on attached Tab A."

On 23 September 1944 a reply (79) signed by the Deputy Assistant Chief of Staff, G-3 was received stating, "Sufficient staff personnel, flight surgeons, statistical control, and operations analysis personnel are now available within the Army Air Forces to capably compile necessary information regarding aptitude scores, training records, and any other pertinent data necessary for proper selection of lead crews in combat groups.

"In view of the remarks of paragraph 2 above reference memorandum is returned not favorably considered."

Some months later the need for Central Medical Establishments in the overseas Air Forces made it possible to reopen this matter. Accordingly, the new Table of Organization and Equipment No. 8-460 (80) for Central Medical Establishments, Aviation, which was published by the War Department, Washington, D. C. on 12 July 1945, called for five aviation psychologists (2251), one lieutenant colonel, one major, two captains and one lieutenant. It established a Psychological Center in the Central Medical Establishment, Aviation, which was described as follows: "This Center acts as a clearing house for stanine scores and other classification records from the Zone of the Interior for all flying personnel and develops procedures for the identification of personnel with superior aptitude for special duty assignments and makes appropriate recommendations to commanding officers."

With the publication of this table for Central Medical Establishment, Aviation, plans were made for staffing Psychological Centers in five Central Medical Establishments, Aviation, in the Pacific theater. Following a conference with key personnel of the Aviation Psychology Program on 5 August 1945, the Chief of the Psychological Branch, Headquarters Army Air Forces, wrote (81) to the men selected to head these new units, Lt. Colonels W. M. Lepley and S. R. Wallace, and Majors N. E. Miller, C. W. Harris, and R. H. Henneman describing the functions of the psychological section and outlining plans for staffing, activating, and sending these units to the Pacific theater. Orders were issued sending some of these men to Warner-Robins Field for processing and training prior to overseas shipment but none of the psychologists was actually sent overseas for duty in Central Medical Establishments because of the cessation of hostilities in August 1945.

## Forming and Evaluating Potential Lead Crews

A conference was called regarding the processing of men for combat crews in Headquarters Army Air Forces beginning 26 February 1945. This was a problem which personnel of the Aviation Psychology Program had been working on for some time and representatives of the Training Command and Continental Air Force aviation psychology units attended. With the assistance of personnel at this conference, representing all phases of personnel processing, a directive (82) was sent to Headquarters AAF Training Command from Headquarters Army Air Forces establishing a single central processing station for all personnel leaving the AAF Training Command. The standard operating procedures (83) for in-bound combat crew processing at central processing stations was prepared at the conference and transmitted to Headquarters AAF Training Command along with the directive. One section of this procedure is quoted below:

> Individuals processed at Lincoln Army Air Base will be formed into crews, utilizing the best features of the procedures now used for crew matching by the Continental Air Forces. The purpose of these procedures will be to form crews which are most likely to develop maximum teamwork and effectiveness. Furthermore, by placing together individuals whose potentialities for crew training are greatest, the Continental Air Forces will be enabled to send overseas the largest possible number of outstandingly good crews from which group commanders can select leaders for their combat formations. It is suggested that the following factors should be given special consideration in crew formation: rank, experience, proficiency, and personality. Information concerning rank and experience is readily available. Proficiency data will include: (a) original aptitude test-scores, (b) ratings obtained from the individual's training records, and (c) scores on standardized objective proficiency tests and phase checks. In addition to the use of these data it is suggested that, insofar as practicable, personal factors which influence crew compatability such as age, education, and mutual interests, be considered in matching crews.

As previously mentioned, the Psychological Research Project (Combat Crew) was established at Lincoln Army Air Field on 6 April 1945 to perform this function.

A conference of aviation psychologists was held at Lincoln Army Air Field from 7 to 10 May 1945. Representatives of AAF Training Command, AAF Personnel Distribution Command, AAF School of Aviation Medicine, the Continental Air Forces, and Headquarters Army Air Forces were in attendance. The principal purpose of the conference was to develop procedures based on recent research findings to be used in assigning personnel to potential lead-crew positions at the time of processing at Lincoln Army Air Field, prior to shipment to the Continental Air Forces for operational training. Recent research findings were presented including reports by the Air-Crew Evaluation and Research Detachments recently returned from the Pacific Ocean Areas and from the European Theater of Operations and the Mediterranean Theater of Operations. Agreement was reached on the items and weights to be used in computing lead-crew aptitude scores.

An AAF Letter, No. 50-117 (84), Screening of Combat Crew Personnel, dated 7 June 1945 was published by Headquarters Army Air Forces. The procedures established by this letter are quoted in the following paragraphs:

> Potential lead crews will be composed of air-crew members who have the highest lead-crew aptitude scores; namely, 7, 8, and 9. Insofar as possible, potential lead crews will be made up of crew members all of whom have lead-crew aptitude scores of 9, all of whom have lead-crew aptitude scores of 8, or all of whom have lead-crew aptitude scores of 7. The remaining combat crews will be assembled without regard to lead-crew aptitude scores.
>
> A Combat Crew Record Form (copy attached) will be initiated at the AAF Combat Crew Processing and Distribution Center, which will indicate whether the crew is designated as a potential lead crew or as a combat crew. The name, rank, and serial number of each crew member assigned to the crew will be entered. This form will be forwarded in duplicate to the combat-crew training station to which the crew is sent. Authority is granted to the AAF Training Command for the reproduction of this form as shown in the attachment to this Letter, in sufficient quantities to furnish a three (3) months supply for use at the AAF Combat Crew Processing and Distribution Center.
>
> Upon completion of combat crew training, the designation of potential lead crew or of combat crew made in the AAF Training Command will be confirmed or revised at the training air force station and suitable notations will be made on the form. Appropriate remarks concerning the proficiency of the crew as a whole will be entered.
>
> The original of the Combat Crew Record Form will be forwarded to the theater for the information of the combat organization to which the crew is assigned. The duplicate will be forwarded to the commanding general of the air force in which combat-crew training is accomplished. The surgeon and the office responsible for training programs at the headquarters of the air force will use this information to evaluate the potential lead-crew selection program.

These procedures were put into operation in the summer of 1945. The cessation of hostilities prevented an adequate evaluation of their effectiveness.

## PSYCHOLOGICAL RESEARCH ON PROBLEMS OF AVIATION EQUIPMENT

From the earliest days of the Aviation Psychology Program in the Army Air Forces, it was clear that one area in which psychologists might make useful contributions was in connection with the psychological research to determine the effect of human capacities on the design of aviation equipment. A substantial number of projects had been completed prior to the spring of 1945 by various units in the Aviation Psychology Program, especially the Department of Psychology, School of Aviation Medicine. However, it was difficult to select the research problems of the greatest immediate benefit to those developing new planes and equipment without the intimate knowledge of plans and new designs which can be gained only by working closely with the scientists and engineers doing the developmental work. It was therefore believed that to be most effective a large part of this work would have to be done in close association with the engineering laboratories and other research and development groups.

On 2? May 1945, a directive (85) was approved by the Air Staff and sent to the Commanding General, Air Technical Service Command, Wright Field, Dayton, Ohio. This directive stated, "It is desired that facilities be established for conducting scientific psychological research on problems of aviation equipment. Because of the intimate relationship between psychological, medical, physiological, and biophysical research problems, it is believed that this psychological research project should be established as a branch of the Aero Medical Laboratory."

Lt. Col. Paul M. Fitts was transferred from Headquarters Army Air Forces to Wright Field to be Chief of this Psychological Branch.

Immediately after his assignment to Wright Field, the Chief of the Psychological Branch of the Aero Medical Laboratory was sent on 3 months temporary duty as described in the next section to study new developments and procedures of aviation psychologists in England and especially Germany. He returned to this country on 30 August 1945 and early in September additional personnel were added to the staff and detailed plans were developed for the Psychological Research Program on problems of equipment design.

## PROJECTS INVOLVING ALLIED AND ENEMY AIR FORCES

A fairly close liaison was maintained by the psychologists in the United States Army Air Forces with aviation psychologists in other Allied air forces. This led to certain formal exchanges of materials as well as much profitable informal exchange of experiences.

The first major release of materials developed in the United States Army Air Forces for use by an Allied nation was the making available of all of the tests of the Air-Crew Classification Test Battery to the Royal Air Force in the spring of 1944. The members of the Training Research Group in the Air Ministry, particularly Dr. Bott and Dr. Myers, had visited this country and studied the procedures which had been developed for the selection and classification of air crew. The plans for initiating an air-crew classification program were developed in the winter of 1943-1944 and Colonel Flanagan, who was in England at the time, consulted with the group regarding these plans. As a result, an official request (86) was received from the Royal Air Force on 3 January 1944 for the privilege of reproducing and using the United States Army Air Forces Air-Crew Classification Test Battery in classifying their candidates for air-crew personnel training. This request was granted (87) on 26 January 1944. Later, on 29 August 1944, a request (88) was received to have this privilege extended to all of the Dominion Air Forces of the British Empire. This was approved (89) on 16 September 1944. The classification procedures were placed in operation on 1 April 1944 and several reports on the English experience with these tests have been received.

In March 1944 at the request of one of the American members of the Joint Air Commission, the United States Army Air Forces Air-Crew Classification Tests and procedures were described to the members of that group by Colonel Flanagan who was in the Mediterranean Theater. As a result of this conference a cable (90) was received on 7 March 1944 stating that the Joint Air Commission recommended that the procedures be translated and adapted for use by the French Air Force in selecting the pilots and other air-crew members in North Africa before sending them to the United States for training. This request was approved (91) on 10 March 1944 and Lt. Col. Philip H. DuBois with two enlisted men reported to the Commanding General, North African Theater, on 26 April 1944 and supervised the final adaptation of the printed tests, the installation of the apparatus tests provided, the development of appropriate routines and procedures, and the training of French personnel to operate the selection and classification program.

With the reoccupation of the Philippine Islands, a request (92) was received on 27 July 1945 from the Theater Commander for assistance in adapting and installing the Air-Crew Classification Tests for use in screening the candidates for pilot and other air-crew training in the Philippines before sending them to this country for training. This request was approved (93) and on 17 September 1945 Col. Frank A. Geldard with 2 officers and 11 enlisted psychological assistants, proceeded to Manila by air to establish an air-crew selection and classification program for the Philippine Air Force.

While this group was in the Philippines, information was received from selection test research personnel in the Navy's Bureau of Personnel indicating that both the Japanese Army and Navy Air Forces had made extensive use of psychological tests for the selection of pilots. Accordingly, a cable (94) was sent to the Commander of Allied Forces in Japan suggesting that Col. F. A. Geldard and Maj. C. W. Harris proceed from the Philippines to Japan for the purpose of investigating the Japanese aviation psychology program. This suggestion was approved (95) and these officers visited Tokyo and several of the principal research and testing centers in Japan and collected valuable information regarding Japanese aviation psychology.

They returned to the Philippines and the entire detachment returned by air to Hamilton Field, Calif., arriving in mid-December 1945.

For the purpose of studying developments in German aviation psychology, Lt. Col. Paul M. Fitts was ordered (96) to the European Theater on 5 May 1945. This officer visited a number of research laboratories and interviewed psychologists who had been associated with the Luftwaffe Aviation Psychology Program before the program was abolished early in 1942. Valuable information was obtained regarding aviation psychology as it developed in the German Air Force.

## A COMPREHENSIVE AVIATION PSYCHOLOGY PROGRAM

By the summer of 1945 a need had developed for a formal official statement setting forth the functions and organization of the Aviation Psychology Program in the Army Air Forces. Various field organizations and air inspectors pointed out that the numerous isolated directives did not sent to the Commanding General, Air Technical Service Command, Wright Field, Dayton, Ohio. This directive stated, "It is desired that facilities be established for conducting scientific psychological research on problems of aviation equipment. Because of the intimate relationship between psychological, medical, physiological, and biophysical research problems, it is believed that this psychological research project should be established as a branch of the Aero Medical Laboratory."

Lt. Col. Paul M. Fitts was transferred from Headquarters Army Air Forces to Wright Field to be Chief of this Psychological Branch.

Immediately after his assignment to Wright Field, the Chief of the Psychological Branch of the Aero Medical Laboratory was sent on 3 months temporary duty as described in the next section to study new developments and procedures of aviation psychologists in England and especially Germany. He returned to this country on 30 August 1945 and early in September additional personnel were added to the staff and detailed plans were developed for the Psychological Research Program on problems of equipment design.

## PROJECTS INVOLVING ALLIED AND ENEMY AIR FORCES

A fairly close liaison was maintained by the psychologists in the United States Army Air Forces with aviation psychologists in other Allied air forces. This led to certain formal exchanges of materials as well as much profitable informal exchange of experiences.

The first major release of materials developed in the United States Army Air Forces for use by an Allied nation was the making available of all of the tests of the Air-Crew Classification Test Battery to the Royal Air Force in the spring of 1944. The members of the Training Research Group in the Air Ministry, particularly Dr. Bott and Dr. Myers, had visited this country and studied the procedures which had been developed for the selection and classification of air crew. The plans for initiating an air-crew classification program were developed in the winter of 1943-1944 and Colonel Flanagan, who was in England at the time, consulted with the group regarding these plans. As a result, an official request (86) was received from the Royal Air Force on 3 January 1944 for the privilege of reproducing and using the United States Army Air Forces Air-Crew Classification Test Battery in classifying their candidates for air-crew personnel training. This request was granted (87) on 26 January 1944. Later, on 29 August 1944, a request (88) was received to have this privilege extended to all of the Dominion Air Forces of the British Empire. This was approved (89) on 16 September 1944. The classification procedures were placed in operation on 1 April 1944 and several reports on the English experience with these tests have been received.

In March 1944 at the request of one of the American members of the Joint Air Commission, the United States Army Air Forces Air-Crew Classification Tests and procedures were described to the members of that group for the month of June will include a summary report of research activities for the fiscal year ending that month."

In the early spring of 1946 there were conflicting ideas in various divisions of the Air Staff regarding the organization of psychological research in the peacetime Army Air Forces. At the request of Maj. Gen. Curtis E. LeMay, a conference on psychological research in the AAF was called on 14 and 15 March 1946 in Headquarters Army Air Forces in Washington. The Office of the Secretary of War invited representatives of the various groups who had been active in psychological research in the AAF during the war to meet together at this conference for the purpose of preparing recommendations on the general scope of the work to be done and the organization for carrying it out during the peacetime period.

This group made recommendations (101) to General LeMay concerning the scope of future work, the operation of a research program, staff, and supervision, conditions of work for talented personnel, and organization. With respect to organization, the following recommendations are quoted:

a. It is highly recommended that scientific research in the AAF including psychological research, be carried on by means of an Air Science Corps which is directed by a group of military and civilian scientists under the Deputy Chief of Air Staff for Research and Development. This calls for an organization somewhat similar to the present Air Medical Corps in which the responsibility and authority for the control of research matters lies with members of the Scientific Corps and not with the local commanders. This also means that it is recommended that the Deputy Chief of Air Staff for Research and Development act as the director of a separate operating research agency with the assistance of an Air Scientist and other science specialists. As the first step in activating a coordinated Psychological Research Program under such an Air Science Corps it is recommended that an outstanding military or civilian psychologist, with the rank of colonel or the grade of P-8, be assigned to the Office of the Deputy Chief of Air Staff for Research and Development, that he be provided with the necessary staff, and be charged with the preparation of a detailed study on organization and work, based on the conclusions stated in this memo.

b. If there is to be no organization in the AAF, such as an Air Science Corps, for carrying on general scientific research, then it is recommended that Psychological Research be carried on by an organization under the general supervision of the Air Surgeon. It is important to note that such an organization will be responsible for Psychological Research which is of interest not only to the Air Surgeon but to all other parts of the AAF as well, and particularly in those commands concerned with classification and training. This will require that the scope of the psychological research now contemplated by the Air Surgeon must be enlarged to include the general scope recommended in this memo. It will also probably require that the contemplated staff be enlarged. Free interchange of information between this organization and other parts of the AAF and close coordination with them is essential. It is therefore recommended that the Director of Psychological Research report directly to the Air Surgeon and have a direct channel of communication to the Deputy Chief of Air Staff for Research and Development on all technical matters pertaining to psychological research and on matters pertaining to the preparation and defense of the budget for Psychological Research. As the first step in activating a coordinated Psychological Research Program in the AAF under the Air Surgeon, it is recommended that an outstanding military or civilian psychologist, with the rank of colonel or grade of P-8, be assigned to the Office of the Air Surgeon, that he be provided with the necessary staff, and be charged with the preparation of a detailed study on organization and work, based on the conclusions stated in this memo.

c. The least desirable situation, and one which existed during the war, is to have several groups in the AAF carrying on psychological research of special interest to the section of the AAF concerned. If this is the only organization possible, then it is recommended that everything possible be done to coordinate the work of the separate groups and to ensure that the recommended general scope of work is covered.

In commenting on these recommendations the Deputy Chief of Air Staff for Research and Development stated that such an organization as recommended under paragraph "a" was not feasible at that time and he therefore directed the Air Surgeon to prepare suitable directives to put the recommended program in effect. This resulted in the preparation of a revision (102) of AAF Regulation No. 20-59, the text of which is quoted in full below:

1. *General.* A comprehensive psychological research program has been established in the AAF to conduct research in order to provide basic scientific information which can be used to improve:

a. The selection and classification of all AAF personnel.

b. The effectiveness of training in the various specialties.

c. The utilization of personnel in various types of assignments.

d. The design and operation of equipment with regard to the human capacities of operating personnel.

2. *Functions.* Psychological research is conducted to provide scientific information to staff officers responsible for the formulation of policy. This basic scientific information is obtained by professional research personnel working in cooperation with operating personnel and is for the use of all staff agencies. The research personnel of this group should avoid participation in service activities not essential to research. Specifically, this program will be responsible for initiating, coordinating, and conducting research in the following major fields:

a. Development and refinement of procedures for the initial selection and classification for individual specialties on the basis of aptitude, interests and personality, and experience.

b. Research on advanced selection and classification for individual, crew, and unit assignments on the basis of aptitude, interests and personality, measured proficiency, and experience.

c. Research on the design of equipment, including cockpits, controls, instruments, and display systems with reference to the human capacities of the personnel who will operate this equipment.

d. Research on training problems, including job analyses; evaluation of training standards; objective study of teaching methods; improvement of curricula and instructional materials; effectiveness of training aids and devices; selection, training, and evaluation of instructors; and optimal methods of retaining proficiency after training.

e. Development and refinement of methods of measuring proficiency at various levels of training in the different types of officer and enlisted personnel duties.

f. Investigation of the psychological problems involved in the operation of new types of aircraft and weapons.

g. Investigation of the opinions, attitudes, and motivation of individuals and groups and of the effects of propaganda on these.

h. Research on problems of personnel management, including such aspects as leadership, morale, personal adjustment, and methods of influencing individuals and groups toward more effective activity.

i. Follow-up of personnel through training and subsequent AAF careers to determine what psychological measures or other records are effective in predicting later performance.

3. *Administration.* The Aviation Psychological Program is administered as a coordinated scientific research program as an integral part of the general research and development program of the AAF. The Director of the Aviation Psychological Program, Office of the Air Surgeon, Headquarters, AAF is responsible for the technical supervision of all psychological research in the AAF, and has staff responsibility for the establishment of research policies, coordination of psychological procedures, assignment of priorities for research, supervision of contractual relations with civilian research laboratories, reporting of results to proper agencies for use, and liaison with other organizations, (military and civilian) in the United States and foreign countries carrying out research in the same fields. Specific research activities will be supervised through periodic research reports and the assignment of priorities and deadlines. In order to make psychological research results available to interested staff agencies, reports of this research will be published.

4. *Organization.* Psychological personnel will be assigned to the various commands as the Commanding General, AAF may direct to carry out the functions indicated above. Psychological personnel assigned to specific headquarters, laboratories, or stations will work closely with the personnel of interested operating agencies, but will be assigned normally to the surgeon for administrative purposes. Contracts for cooperative research may be made with university laboratories and other civilian research organizations in order to obtain the services of the most expert professional talent available.

5. *Activities.* The activities of psychological personnel in specific stations and the projects of civilian and other Government laboratories will be planned so as to make the maximum contribution to the solution of the general problems of research and development assigned to the Aviation Psychological Program. Time and facilities will be scheduled for the administration of tests, flight checks and other proficiency measures, and examinations for the purpose of selecting, classifying, and evaluating personnel as a research activity. Time and facilities will be made available for research testing and other research studies which are essential to the research and development needs of the AAF as outlined above.

The scope and functions of the Aviation Psychology Program as outlined in this revision of the regulation were agreed to by the Air Staff and on 24 May 1946 the Deputy Chief of Air Staff for Research and Development sent a directive (103) to AC/AS-1 directing that the broadened program be activated and personnel procured to carry it out.

# Part II

## Specific Solutions to Problems of the Aviation Psychology Program

# CHAPTER FOUR

# The Selection and Classification of Air-Crew Personnel

## INTRODUCTION

This part of the report discusses specific solutions obtained for problems assigned or selected for study by personnel of the Aviation Psychology Program. It stresses accomplishments which were of immediate value, describing the policies and procedures developed and giving evidence as to their usefulness to the Army Air Forces. In a later part of the report there is a discussion of techniques and psychological findings with emphasis on their broader implications for psychological research.

As has been previously indicated, the initial problem assigned to the personnel of the Aviation Psychology Program was the development of procedures for the selection and classification of men for pilot training. This was later expanded to include all members of the air crew. Because of the belief in the fundamental importance of the selection and classification problem, this problem not only comprised a large portion of the research effort of the personnel of the program during the first stages of the war, but also supplied the framework within which other problems were developed and attacked. For example, the problem of the evaluation of proficiency of air-crew members grew directly out of an attempt to evaluate the effectiveness of the selection and classification procedures developed.

By the spring of 1943 the progress in selection and classification research made it possible to assign a substantial number of aviation psychologists to the study of training problems such as methods of instruction, curricula, training devices and equipment, and the selection and training of instructors. Even the problems regarding the effects of combat and attitude toward a second tour of duty were closely interrelated to the original problems of selection and classification, since a principal issue involved was the question of predisposition of individuals to anxiety reaction as a result of combat stress.

## INITIAL SELECTION OF AIR CREW WITH THE AAF QUALIFYING EXAMINATION

### The Problem and Proposed Solution

Flying through the air seems to be so different from the tasks ordinarily undertaken by human beings that the need for special selection procedures to decide which individuals are most capable of adapting to the new conditions and demands of aerial flight has long been recognized. The publication, *Notes on Psychology and Personality Studies in Aviation Medicine*[1] prepared at the School of Aviation Medicine, Randolph Field, and published in January 1941, reports some of the history of selecting men for military aviation during the First World War and in the interim preceding this past one. Paragraphs 142 and 143 state in part:

> During the early part of the World War the Allies selected their pilots in a haphazard manner. Frequently they were assigned to the Air Corps because of their inability to continue performance of ground duties. After many terrible accidents with their tremendous toll of manpower and materiel, the Allies began to consider the problems connected with flying and to wonder if all individuals were adequately endowed to meet these problems. In the beginning, courage was considered the only trait essential to piloting an airplane. If an individual possessed that to a high degree, there was nothing to prevent him from flying. Bitter experience taught the folly of any such assumption. No more do all individuals possess the aptitude for military flying than do all individuals possess the aptitude for painting, sculpturing, music, par golf, or any of the other accomplishments to which a favored few may attain.
>
> * * *
>
> The first Research Board of the Air Service was established October 18, 1917, at Mineola, New York. * * * In the fall of 1917, 67 trained examining units were established in the larger cities. These units examined thousands and thousands of young men for the Air Corps. Many medical problems connected with aviation were solved through the efforts of these pioneer investigators. This early work has been continued in turn by the Medical Research Laboratory, the School for Flight Surgeons, and at the present time by the School of Aviation Medicine, Randolph Field, Texas, and the Physiological Laboratory at Wright Field, Dayton, Ohio. An immense amount of research has been accomplished by the personnel of these institutions and as a result of their work, plus experience gained by medical officers' flying, there is ability to state which man should fly and which man should not, and also to predict with some degree of accuracy which student has the ability to learn to fly, and which one has not.
>
> Throughout the United States there are some several thousand young men between the ages of 20 and 26 examined annually for flying-cadet training. These young men are from the various colleges and universities. Of those examined, about 80 percent fail to pass the physical examination. This is an extremely high percentage of rejection when considering that the educational requirement (2 years of college) has already served to prevent a large percentage of young men from even appearing for the examination. The 20 percent who pass the physical examination represent an extremely highly selected group. They are the cream of the country when measured by these standards.

There are several factors which made the problem of selection of pilots as discussed above much more difficult in wartime. In peacetime, selection could be based on an intensive analysis of each applicant by a specially trained Flight Surgeon. The tremendous demand for pilots and other members of the air crew following the declaration of war in December 1941 made such procedures impracticable. In peacetime there was also a large amount of self-selection. Only a relatively small group of the young men applied for training in military aviation and these were in general those

[1] *Notes on Psychology and Personality Studies in Aviation Medicine*, Technical Manual 8-320, Washington: War Department, 27 January 1941.

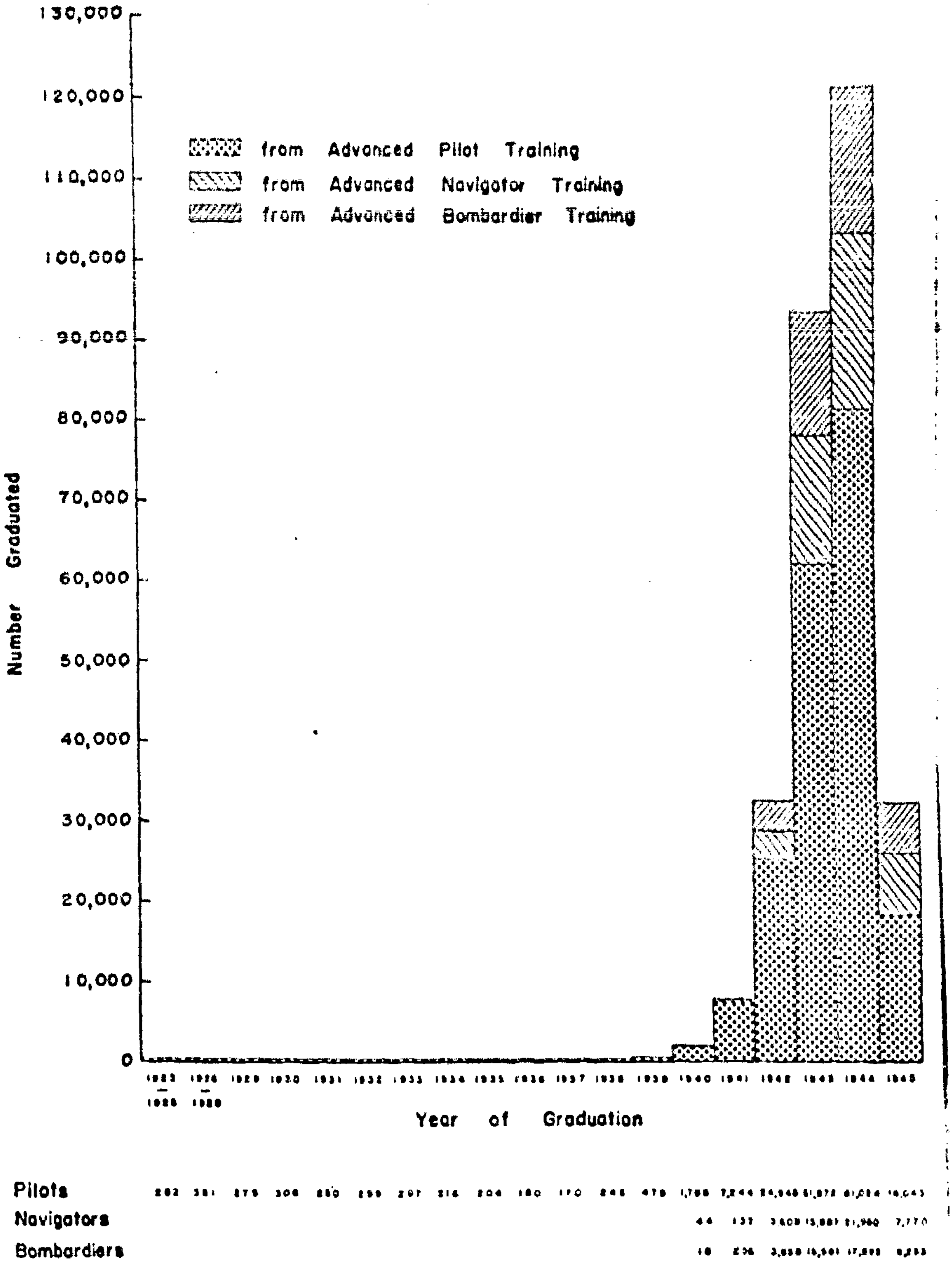



FIGURE 4.1.—Number of pilots, navigators, and bombardiers graduated from advanced training schools each year from 1923 through 1945.

who believed that their special aptitudes fitted them especially for this work.

With the advent of compulsory military training, other factors played a larger part in inducing men to apply for aviation training. Some individuals with little aptitude for this type of training were attracted by the social prestige, salary, or similar expected gain or merely as a way to avoid something considered less desirable, such as marching, mud, and mortars. When the number of men being trained was small it was possible to restrict applicants to those who had had a broad educational background by imposing such formal requirements as 2 years of college. It was also possible to restrict applicants in various other ways, such as requiring that they be single and of certain ages. However, such requirements were only very indirectly related to the actual needs for success in military aviation.

The problem confronting the Army Air Forces in 1941 was how to develop a simple efficient procedure for initially screening men in all parts of the country in quantities very much larger than had ever been previously handled for this type of work.

The magnitude and importance of the task is illustrated by figure 4.1 which shows the number of men trained in successive years from 1923 through 1945.

The proposal in November by the psychological group that a general examination be prepared suitable for the selection of qualified cadets for all types of training, was tentatively approved by the Air Staff in December 1941 and was put into general use on 15 January 1942 in several hundred Aviation Cadet Examining Boards established throughout the country.

The use of the AAF Qualifying Examination eliminated the need for previous formal educational requirements and the academic type educational tests designed to demonstrate an "equivalent" amount of educational training. The purpose of the examination was to make it possible for every young man who possessed the required aptitude, knowledge, and skills to have an opportunity to become a bombardier, pilot, or navigator. The examination was designed to qualify all men who had good prospects of succeeding and to eliminate only those who were definitely unqualified for these types of training.

This was not a test of the speed with which a series of abstract and artificial questions could be answered; instead, it was a test designed to provide a carefully standardized sample of the man's ability to do some of the kinds of things which an aviation cadet had to be able to do in the training schools. Although most men completed the examination in less than 2 hours, the applicant could spend as much as 3 hours on the examination if he wished. The large number of items and the inclusion of several sections measuring different abilities and characteristics provided a sound and reasonable basis for the selection of prospective officers in the Air Forces. The kinds of materials included are described briefly below.

703323—47—5

An analysis of the duties of aviation cadets and officers in the Army Air Forces showed that one of the important qualifications for these individuals is the ability to comprehend instructions and other materials and to follow directions. This ability is chiefly dependent on understanding the precise meaning of words and a proficiency in organizing, relating, interpreting, and drawing conclusions concerning materials read.

In addition to ability to comprehend the material he reads, the aviation cadet needed as general background for his training as an officer, skill and accuracy in fundamental mathematics and an aptitude for the interpretation of mechanical devices and diagrams. Air Force flying officers should be able to interpret maps, charts, and weather reports. They must know how to plot a course and check the accuracy with which they are following it. These operations have to be performed rapidly and accurately. The aviation cadet should be able to solve problems involving proportions, fractions, ratios, decimals, formulas, and elementary algebra and also to read and interpret graphs, tables, and charts. In order to obtain a working knowledge of the basic principles of operation of the airplane and the incidental mechanical equipment, aptitude for the comprehension of mechanical diagrams and devices was essential. The questions concerning mathematics and mechanical comprehension were intended to reveal the applicants' aptitudes and potentialities rather than the results of extensive and detailed formal training in these fields.

Finally, as a prospective officer in the Army Air Forces, the aviation cadet should possess qualities of leadership. The potential leader should be alert to recent changes and developments in aviation and related fields. For effective leadership, demonstrated initiative and inquisitiveness, as revealed by an awareness of recent happenings and an understanding of the significance of new developments, were believed to be of more value than an accumulation of academic credits. Another attribute of the potential leader which was regarded as important for aviation cadets is sound judgment in practical situations. Poor judgment makes the cadet a source of danger, not only to himself but to others. To test these characteristics, questions were prepared requiring a knowledge of recent important events and developments, and also questions presenting practical problems which might be met, not only in flying, but in everyday activities of the type with which applicants could be expected to have had experience.

### Development and Use[8]

Over a period of 4 years, 17 forms of the AAF Qualifying Examination were constructed and published. Numerous studies were made of the results of preceding forms in an effort to improve the quality of the men selected for air-crew training with respect to those characteristics found

[8] The development of the AAF Qualifying Examination was carried out under the immediate supervision of Frederick B. Davis after the construction of the first form of the test. He was assisted by Mary B. Willis, William G. Mollenkopf, Dorothy C. Bechtoldt, John T. Dailey, Chester W. Harris, and William J. McCabe during various periods. Valuable advice was also received from Irving D. Lorge, who served as a Special Consultant.

to be of most importance for their later work as cadets and as officers. The form of the AAF Qualifying Examination in use at the end of the war in 1945 included 15 reading-comprehension items, 50 general-information items sampling interests in aviation and related fields, 60 mechanical-comprehension items, and 25 hidden-figures items. For administrative convenience a time limit of 3 hours was used, but practically all applicants completed the test before the end of the time limit.

A sampling of answer sheets received from the various Aviation Cadet Examining Boards was rescored from time to time in the Psychological Branch of the Office of the Air Surgeon and various statistical tabulations of scores and proportions passing and failing in various Boards were made to check on the conditions of test administration. In general, the results were found to be very satisfactory.

Although accurate records regarding the number of men who were tested with the AAF Qualifying Examination are not available, figures indicate that approximately 550,000 men took the AAF Qualifying Examination in 1942 and approximately 50 percent of them were qualified and 50 percent rejected on the basis of this examination. During 1943 approximately 350,000 men were examined and during 1944 the number was about 250,000. During these 2 years about two-thirds of the men were qualified on the basis of scores made on this examination and about one-third were rejected. The increase in the proportion passing in the last 2 years appears to have been due in part to self-selection on the part of the applicants and in part to a reduction in the passing score on the AAF Qualifying Examination which accompanied the raising of the standards for pilot, bombardier and navigator in terms of the Air-Crew Classification Test stanines in July of 1943. These estimates of numbers appear to be a little low, since a substantial number of men qualified on the AAF Qualifying Examination were disqualified on the physical examination, and more than 600,000 men were qualified at AAF Examining Boards and sent to take the Air-Crew Classification Tests during this period.

## Evaluation Studies

Several studies were made to determine the proportion of various types of population who could be expected to qualify on the basis of this examination. It was found that in the graduating class at the United States Military Academy at West Point, the lowest score was 105 on the first form of the AAF Qualifying Examination, Test AC10A. This form contained 150 items and the minimum qualifying score on it was 90. Thus the West Point cadets all easily qualified on this examination. In 1942 the freshmen at five colleges were examined with various forms of the AAF Qualifying Examination. The proportion passing varied from 90 percent in one of the better colleges where entrance requirements are very high to approximately 40 percent in a university having rather low requirements for admission. In the other three colleges which were fairly typical state universities, between 50 and 75 percent of the freshmen passed the examination. It seems probable that about two-thirds of the freshmen in a typical college could be expected to pass the AAF Qualifying Examination.

During the war, high school seniors in cities and towns in four different sections of the country took various forms of the Qualifying Examination. There was much more variation in the average scores for the seniors in different high schools in the same cities than there was between the average scores for those in different cities. In some high schools as few as 10 percent of the seniors obtained a qualifying score. In other schools as many as 60 percent of the high school seniors achieved a passing mark. Using the passing score, 90, on the first test, AC10A, and its equivalent value, 75, on test AC14L which was in use at the end of the war, it was found that approximately one-fourth of high school seniors were able to qualify under these standards on this examination at the time of graduation from high school. As previously indicated, somewhat lower standards were in effect for initial screening for a time during the war after classification standards were raised.

In a study made of applicants at the Philadelphia Aviation Cadet Examining Board during the first 3 months of 1942, it was found that of the 1,640 men applying, approximately two-thirds were high school graduates. The others were distributed fairly evenly among those with 2 or more years of college, 1 year of college, and less than 4 years of high school. Of the applicants with either 2 or more years of college, the former educational requirement, and also for those with 1 year of college, approximately 80 percent passed the Aviation Cadet Qualifying Examination. For the high school graduates the proportion was a little less than 60 percent and for those with less than 4 years of high school the number qualifying was approximately 20 percent. Comparing these figures with those in the previous paragraph, it is seen that in all categories there is evidence of self-selection on the part of the applicants. On the basis of an early study comparing the AAF Qualifying Examination with the Army General Classification Test, it was estimated that approximately one-fourth of the men who were inducted into the Army would be able to pass the AAF Qualifying Examination.

A number of studies were done to define more adequately the functions measured by the AAF Qualifying Examination. The correlation between the Army General Classification Test and Test AC10A was found to be 0.64 in a sample of 282 aviation cadets.[3] For a later form of the examination, Test AC12I, which did not include vocabulary and mathematics, the correlation in a sample of 660 candidates for air-crew training was found to be 0.47.[4] The correlation between score on the AAF Qualifying Examination, AC10A, and the average of academic grades for the 4-year

[3] The mean and standard deviation for the AAF Qualifying Examination were respectively 103.6 and 16.6 and for the Army General Classification Test they were 57.5 and 14.3 in terms of raw scores. (In terms of Standard Scores these values were approximately 125 and 10.5.)

[4] The mean and standard deviation for the AAF Qualifying Examination were respectively 218.6 and 31.Q and for the Army General Classification Test they were 116.5 and 12.1 in terms of Standard Scores.

course at West Point was found to be 0.50 for the class of 1942 at the United States Military Academy.

In January 1942 over 2,000 aviation students and aviation cadets who had been selected under the previous requirements were given the first form of the AAF Qualifying Examination at Maxwell and Kelly Fields. Of the group, containing approximately three-quarters of these men, who obtained scores which would have qualified them for aviation-cadet training, 40 percent were eliminated in pilot training. Of the remaining quarter who did not obtain scores which would have qualified them for pilot training 59 percent were eliminated.

A group of 780 men were tested with Test AC10A at Maxwell and Kelly Fields in January, February, March, and April 1942 and sent into navigation training. Approximately three-quarters of this group passed navigation training and the remaining one-quarter failed. The biserial correlation coefficient obtained between test score and pass-fail was 0.40. For the 30 items in the mathematics section of the test taken separately, the biserial correlation coefficient was 0.59. It should be noted that the size of these coefficients is somewhat reduced by restriction of range because this group was partly made up of persons who had been selected under the old 2-year college requirement and partly made up of people who had qualified on a previous administration of Test AC10A. Furthermore, the mean score of the men sent into navigation training was about 8 points higher than for the men being sent into pilot training at that time, and the standard deviation was approximately 13½ instead of 16, the value that had been found in testing cadets qualified under the old procedures.

For a group of 1,650 men tested with Test AC12I of the AAF Qualifying Examination while in pilot preflight school at San Antonio Aviation Cadet Center, the biserial correlation between the score on the total test and pass-fail in pilot training was 0.26. It should be remembered that this group had been previously selected on the basis of an earlier form of the AAF Qualifying Examination, had been found qualified on the Air-Crew Classification Test Battery, and had been classified for pilot training. Therefore, the coefficient obtained was definitely lower than the value which could be expected in an unrestricted sample.

Because of the difficulty of interpreting the coefficients obtained from selected groups and the need for a clear-cut evaluation of the effectiveness of the AAF Qualifying Examination and the Air-Crew Classification Tests, an experimental group of slightly more than a thousand men were accepted for pilot training regardless of their scores on the AAF Qualifying Examination and regardless of their pilot stanine as determined by the Air-Crew Classification Tests. Of the 598 men who obtained at least a passing mark on the AAF Qualifying Examination, AC12I, 35 percent graduated from advanced pilot training while only 11 percent of the 405 men who failed to make a qualifying score and would ordinarily have been rejected were graduated from advanced pilot training. It should be noted that the standards on this form of the Qualifying Examination had been reduced somewhat to make possible an increase in standards for pilot training on the Air-Crew Classification Tests in terms of the pilot stanine. The biserial correlation coefficient between graduation and elimination including all phases of pilot training was found to be 0.50.

The biserial correlation between graduation and elimination and amount of education which the applicant had received as expressed in terms of the highest grade reached was 0.21. The similar coefficient calculated between graduation and elimination and the Army General Classification Test score was 0.31. The superiority of the AAF Qualifying Examination over these other bases for this particular purpose is obvious.

This initial screening and selection for air-crew duty was of tremendous practical importance to the Air Forces. Approximately 600,000 men were selected by the AAF Qualifying Examination from nearly twice that number of applicants. Probably the most important advance in recruiting procedures during the war was the discovery of the great improvement in quality which could be gained by removing requirements in terms of formal educational credit and experience and substituting in their place demonstrated proficiency, interest and capacity as established by suitable examination procedures.

## THE AIR-CREW CLASSIFICATION PROGRAM

### The Classification Problem

The second problem assigned to the psychological research group in the Office of the Air Surgeon, Headquarters Army Air Forces, was the problem of classification. The problem of classification or placement as it is sometimes called in the textbooks is a much more complicated one than selection. To get maximum efficiency in selection one merely places the individuals in rank order according to desirability for the particular position and then proceeds to take people from the top until a sufficient number have been obtained to fill the quota. If the position is a very complex one the problem consists of first obtaining measures of all of the various traits of importance. The next step is to use these scores in such a way that the individuals are placed in order of desirability with each individual in the rank order clearly more desirable than anyone below him on the list for the particular job on the basis of the obtainable evidence.

For the classification problem it is necessary to consider simultaneously the relative desirability of all the available people for the available positions. The problem is to maximize the effectiveness of the personnel assigned to all types of duty. Consider first the situation in which all jobs are equal in importance and relative difficulty. Let us further assume that the jobs require entirely different skills so that the rank order of desirability for one position is not correlated with the rank order of desirability for another position. Under these conditions the problem is fairly simple.

If there are just two such positions, a quarter of the people will be above average in their ability for both of the positions, an additional quarter will be above average for the first position but not for the second. The reverse will be true for a third quarter and only one-fourth of the original group will be below average for both positions.

If a third position requiring abilities not related to either of those necessary for the first two positions is now added, we find only one-eighth who are in the top half for all three positions and similarly only one-eighth who are in the bottom half for all three positions. Thus we now have seven-eighths of the personnel who can be assigned to one of the three positions with the expectation that they will be found to be above average in suitability for this particular type of work. The reasoning can be easily extended to additional positions and each time a position is added the number of people who cannot be assigned to a group for which they have above average aptitude will be cut in half. This can be seen in figure 4.2 for the cases of two and three positions.

To take another extreme possibility, if the various types of duty can be placed in hierarchical order so that all assignments for the first type of duty are more important than any assignment in the second type of duty and similarly on down for all positions, the problem simplifies to one of successive selection. In this case persons are placed in rank order for the most important position and enough selected to fill the vacancies, then they are placed in rank order for the next most important position, and so on until all positions have been filled.

For the other simple case of a number of equally important positions requiring unrelated skills it is merely necessary to assign each individual to the position for which his aptitude is highest. Thus, if one assigned personnel to two positions at random, the average aptitude score in terms of stanines for each position would be 5.00. On the other hand, if one assigned a group of individuals to these two positions, each individual being assigned to the one for which his aptitude score is highest, the personnel in each position will have average stanines for that position of 6.13. This is illustrated by the front face of the cube in the upper drawing of figure 4.2 by the points $MA$ and $MB$ which indicate the mean scores of the two groups which would be classified for position A and position B respectively.[8]

Similarly, for three positions having unrelated requirements and requiring equal numbers of men assigned, the assignment of each man to the position for which his aptitude score was highest would raise the mean stanine for each position from 5.00 to 6.69. The drawing at the lower right in figure

[8] The mean value for the two position situations is obtained in the following way: (1) Since all surfaces are symmetrical, they may be rotated. (2) Rotating the axes in the front face of the upper figure 45°, it is clear that the mean value of each half section, assuming a normal distribution, is 0.79788 as obtained from the usual formula for mean deviation of a tail of a normal distribution (formula 53, Kelley, T. L. *Statistical Method*, New York, Macmillan Co., 1924). (3) In terms of the original A and B axes this mean value is equal to 0.79788 (Cos 45°) = 0.564. (4) In terms of stanine units this value is equal to 1.13 and therefore the mean values are 6.13 and 3.87 for the two aptitude scores for each section when the division is made on a 45° split.

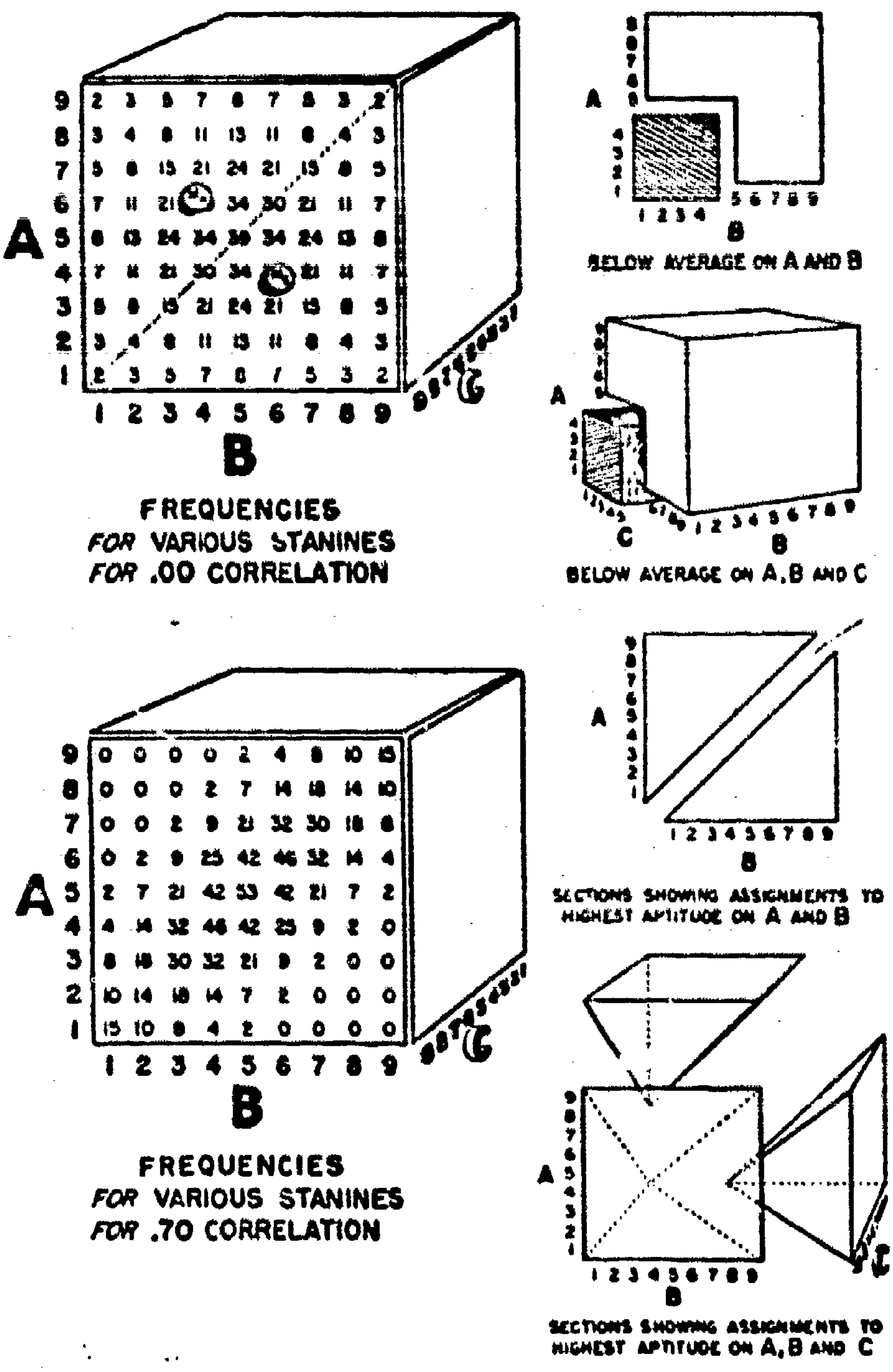



*FIGURE 4.2*

**ILLUSTRATING THE CLASSIFICATION PROBLEM**

4.2 shows the three sections into which the cube representing the independent aptitude scores for the three positions would be divided to maximize the average scores for all three positions.[8]

The practical situation is much more complicated than this. The positions are not of equal importance and in many situations their relative importance is not even stable. There is also a positive correlation between the requirements for most positions. Usually somewhat different numbers of personnel are required for the different assignments also. Preferences, interests, and attitudes are relevant complicating factors for making assignments in most situations. Fortunately, large differences in the relative importance of various positions and also large differences in the numbers of personnel required for the various positions usually tend to simplify the practical problems connected with classification.

The correlation between the aptitudes required for various positions tends to reduce the extent of the improvement that can be effected. This is illustrated by the diagram at the bottom of figure 4.2. It is seen that the mean aptitude scores for positions A and B are closer to the means for the total group as indicated by points $MA$ and $MB$ when a substantial correlation between the aptitudes necessary for the two positions exists. In many situations the problem of classification does not arise because independent recruitment procedures are developed for each of the various specialties and the source of personnel is sufficiently unlimited so that the inefficiency of such procedures is not seriously noticed.

## The Origin of the Air-Crew Classification Problem

When bombardier- and navigator-training programs were started in the summer of 1940 the first classes of men sent into training were composed of personnel who had been eliminated from pilot training because of failure to make satisfactory progress in the flying-training schools. In the first classes no selection procedures were used and many of the students in the resulting classes were found to be unsatisfactory. Col. W. M. Garland and Maj. Gen. Fred L. Anderson made informal studies and observations of the students in the bombardier school and largely as a result of their recom-

[8] For the three position situations the reasoning is as follows: (1) Rotating the axes of the cube, or more precisely the contour sphere, 45° and dividing it into two hemispheres in a manner similar to that used in the two position situation above, we find that one hemisphere contains all of the section classified for position B and half of those classified for position C. (2) It is further observed that the half classified for position C is symmetrical with respect to their aptitude scores for position B. The sum of the scores for position C of this group with respect to position B is therefore equal to zero. (3) The mean score of the individuals in this hemisphere with respect to position B is found, as in the two position situations previously discussed, to be 0.564. This may be expressed in terms of the sections classified for position B and position C as follows:

$$M_{(\text{hemisphere})} = \frac{\text{Summation } B \text{ scores for position } B \text{ section} + 0}{0.5\,N}$$

(4) The mean of the B scores for the position B section can be expressed as:

$$M_{B\,(\text{position } B)} = \frac{\text{Summation } B \text{ scores for position } B \text{ section}}{0.333\,N}$$

(5) Therefore, $M_{(\text{hemisphere})} = \frac{2\,M_{B\,(\text{position } B)}}{3}$ and the mean B aptitude score for those classified for

$$\text{position } B = \frac{3\,M_{(\text{hemisphere})}}{2} = 0.846.$$

(6) Transforming to stanine units, this becomes 1.69 and therefore the mean values are 6.69. Since the situation is similar, the mean values for the aptitude scores for those classified for each of the three positions are equal to 6.69 for the position for which they are classified.

mendations a research project was undertaken by the Training and Operations Division of the Office, Chief of Air Corps, in the spring of 1941 for the purpose of developing procedures for the selection of bombardiers and navigators. Such procedures were developed and put into use in the late fall of 1941.

The procedure was essentially to replace the 2-year college requirement by the requirement of high-school graduation and also add the requirement of obtaining a qualifying score on a battery of three tests consisting of the Army General Classification Test, the Army Mechanical Comprehension Test, and a physics test. The lack of efficiency of this independent recruiting program was recognized by the Army Air Forces and in January 1942 the new selection and classification program was initiated.

An essential element of the plan developed for air-crew recruiting and classification was the emphasis on the other members of the air crew. Although about one bombardier and one navigator were needed for every five pilots, almost all of the applicants for air-crew training wanted to be pilots. Early in 1942 a campaign was started to build up the prestige of the bombardier and navigator by emphasizing the importance of the combat team. Later, as the classification was expanded to include gunners the publicity emphasized the importance of all members of the air crew.

These publicity campaigns and other influences caused between 10 and 20 percent of the applicants to list either bombardier or navigator training as first choice in relation to pilot training. Of even more importance was the increase in the prestige of the bombardier and the navigator so that a larger and larger percentage of the applicants indicated that they had a very strong interest in these types of training and regarded them as only slightly less desirable than pilot training.

### The Development of the Air-Crew Classification Tests

In addition to using preferences, a comprehensive battery of tests, measuring various aptitudes, interests, and abilities which it was believed would have predictive value for one or more of the training specialties, was developed.

The first step in the development of this battery of tests consisted of a systematic survey and analysis of the requirements for success in various activities. In making this study, emphasis was placed upon the critical requirements (critical requirements are defined as those which are critical or crucial in the sense that they are known to have been the factors determining success or failure in the particular task for a substantial number of instances). The first extensive study of the requirements for success as a member of the air crew was the analysis of the reasons for failure reported in the board proceedings of 1,000 pilots eliminated in the last six months of 1941. This study was followed by a number of other studies of requirements which were published in a series of "Analysis of Duties Bulletins." In identifying the traits to be measured, a number of other sources was explored. These included the review of the findings of previous studies, actual participation in the activities, and interviewing and observing persons actually engaging in the activity either successfully or unsuccessfully.

In addition to selecting and defining traits believed to have predictive value in determining success or failure in a particular air-crew assignment, an attempt was made to select traits which were as (1) independent of each other as possible; (2) simple in composition so that they were not likely to include two elements, one of which was important for bombardiers and another for pilots; and (3) stable or not greatly affected by small amounts of practice or differences in experience and training.

To serve as a guide to some of the personnel in the program with limited experience in test development, the first of a series of technical bulletins was prepared by the research personnel in the Psychological Division, Headquarters Army Air Forces, and circulated to all of the field units. This bulletin suggested that, in developing a test to measure a trait believed to be important for success in a particular air-crew assignment, the points which should be specifically considered should include (1) validity, (2) independence, (3) simplicity, (4) stability, (5) objectivity, (6) acceptability, (7) practicality, (8) atypical performance, and (9) discrimination.

From the outset it was decided that all types of testing, observational, questionnaire, and interview procedures would be tried out. Because of the large numbers of persons to be tested it was agreed that objective, printed, multiple-choice tests which could be scored by machines were to be preferred if the trait were susceptible to measurement by this method. It was believed essential to include certain apparatus tests of coordination and speed of decision, at least until such time as it could be demonstrated that the traits measured by these tests could be adequately measured by more efficient devices. It was planned to include an interview as part of the classification processing. Observational procedures, projective techniques, and motion-picture tests were also explored. Unfortunately, none of these procedures was found to produce any improvement in the predictions which could be made with simpler devices.

To develop effective procedures for testing a trait judged to be of importance, primary responsibility for work on tests in each of four different categories was delegated to specific units in the Aviation Psychology Program. To coordinate the test development, a system of circulating test ideas for criticism was established.

In January 1942 a conference was held to plan the initiation of classification testing and the development of necessary tests for use in this program. Classification testing was initiated approximately a week later, using available tests, and work was begun immediately on the procurement of additional copies of apparatus tests measuring those functions which previous research workers had found to be of importance. Because of the heavy load involved in initiating the classification testing work in the field units, the Psychological Division in Washington undertook the development of a number of printed tests during the spring of 1942.[7]

By July 1942 it was possible to initiate use of a battery of 12 printed tests and 5 apparatus tests which had been especially developed to measure those traits believed to be essential to success in the various air-crew specialties. This battery included four different types of mathematics tests believed to be especially important for the navigator; tests of dial and table reading also believed to be of primary importance in selecting navigators; three tests involving speed of perception and recognition of forms which were considered to be especially important to pilots and the bombardier; a test of mechanical comprehension considered to be essential for the pilot; a test of reading and judgment considered to be important for all three positions, but especially important for the navigator; and a technical-vocabulary test containing separate parts for pilot, bombardier, and navigator scores. The apparatus tests of complex coordination and two-hand coordination were regarded as primarily requirements of the pilot and the three other apparatus tests measuring finger dexterity, steadiness, and speed of reaction were believed to be important for the specialized duties of the bombardier.

By December 1942, when the first major change was made in this battery, confirmation had been obtained regarding the validity of a large number of the tests. Revised forms were substituted for several of the printed tests. The Rotary Pursuit Test, believed to measure a type of coordination important for pilot training, was added to the apparatus tests of the battery. A new form of the steadiness test, also including a "stress" element, called the Aiming Stress Test was substituted.

In July 1943 the principal changes introduced into the battery were the expansion of the scope of the Technical Vocabulary Test to include other types of general information and the addition to the battery of a biographical data blank similar to that which had been developed by the National Research Council for the CAA and Navy programs. The Rotary Pursuit Test was revised to include a "divided attention" feature.

In the November 1943 battery the Rudder Control Test, which had been found to measure aspects of pilot aptitude not adequately covered by the test battery in use at that time, was substituted for the Aiming Stress Test.

---

[7] Responsibility for the development of the Air-Crew Classification Tests was delegated to various field units in the Aviation Psychology Program.

The development of tests in the field of personality, temperament, and emotion was primarily the responsibility of Laurance F. Shaffer. He was assisted chiefly by Donald E. Super, Neal E. Miller, and Glen L. Heathers in this work.

The development of tests of intellectual and ability factors was the responsibility of J. P. Guilford. His principal assistants included Merrill F. Roff, Neil D. Warren, Stuart W. Cook, Lloyd G. Humphreys, Wilbur S. Gregory, Milton Burdman, and John I. Lacey.

The development of tests in the perceptual and motion-picture fields was delegated to James J. Gibson working under the supervision of Frank A. Geldard. Assisting in this work were S. Rains Wallace, Richard H. Henneman, Robert M. Gagne, and George F. J. Lehner.

The responsibility for developing the apparatus tests was delegated to Arthur W. Melton in collaboration with Robert T. Rock. The latter's place was later taken when he left the Aviation Psychology Program in April 1943 by Anthony C. Tucker and subsequently by Meredith P. Crawford. Active in this developmental work were Roger B. Loucks, Jack Buel, Judson S. Brown, Walter F. Grether, Glen Finch, and Joseph Weitz.

The work on classification tests in the Psychological Division, Office of the Air Surgeon, Headquarter Army Air Forces, was supervised directly by the editor of this report and was done principally by Paul M. Fitts, Launor F. Carter, and Frank J. Dudek.

Previous forms of steadiness tests given under "stress" conditions had been found to have very nearly zero validity. A new test was added requiring the visualization of the attitude and orientation of a plane in flight on the basis of data shown on diagrammatic sketches of a compass and flight indicator and the identification of the one of five photographs of planes giving the correct representation. This test was found to have important predictive value for pilot training not adequately represented in the battery then in use.

No substantial changes were made in the scope of the test battery as revised for use beginning September 1944. New forms of certain tests were substituted. However, in December 1944 it was directed that a Gunnery Sighting Test be added to the battery, specifically to assist in the selection of B-29 gunners. The only other change made in the scope of the Air-Crew Classification Battery during the wartime period was the substitution on 1 June 1945 of the Two-Hand Pursuit Test for the Two-Hand Coordination Test to improve the validity of the battery for selecting gunners and the addition to the battery on the same date of the Coordinate Reading Test which had been found to be important for the radar observer.

During most of the period there were approximately 14 objective type printed tests of various special aptitudes and abilities and 6 apparatus tests in the Air-Crew Classification Test Battery. All of the pencil-and-paper tests were of the multiple-choice type and the men recorded their answers on separate answer sheets which were scored with the electrical test-scoring machines. The apparatus tests were automatically timed and paced by special control units and the man's score depended solely upon his own performance as recorded on electric time clocks and counters. The printed tests were given to groups of about 200 men. The men were usually seated at specially constructed tables having separate partitions for each man so that the men could be seated fairly close together without any problem of distraction from being able to see anybody else's work. These printed tests were given in two sessions in one day. The sessions were each about 3 hours in length. All of the tests were very accurately timed, although in many of them most of the individuals were able to complete their work well within the time limit.

The program of administering a battery of six apparatus tests to such a large number of individuals involved the solution of a number of new problems. The directions to these tests as well as the printed tests were carefully tried out and standardized and were either read or recited by the examiner. The examiners were carefully trained to pace their instructions in a specified manner. Since each of the apparatus tests was scheduled for 15 minutes and since the men rotated from one to the other of the testing rooms containing these tests, any delay in one test would throw the whole procedure into confusion.

To assist the examiner further and to improve the standardization of the tests the timing of trials on the apparatus tests was done automatically by control units. To permit the handling of such large numbers of individuals, each examination room contained four units of a particular apparatus. All of the units were controlled by the same control unit and the four clocks or counters were arranged in order on the examiner's control desk. Thus one examiner was easily able to test four individuals at one time.

## The Classification Procedures

One of the major problems relating to the use of test scores and other information in evaluating individuals with regard to their aptitude for specific positions is the question of the manner in which the various test scores will be used. The procedures which have been most frequently used in the military services are multiple cut-offs and clinical evaluations. The multiple cut-off procedure considers each test singly and establishes for it a qualifying or cutting score. In determining whether a man is qualified or not under this system one merely considers his scores on each of the tests in turn and continues until one is found on which he is disqualified, or if he does not fall below the cutting score on any of the tests he is considered as qualified.

Where the method of clinical evaluations is used there are generally no fixed qualifying scores or cut-offs for any of the tests or other types of information. The classification officer examines all of the test scores and presumably carries out in his head the application of various procedures such as multiple cut-offs, the use of information regarding the pattern of abilities, and compensations of scores on one trait for those found on another. He then makes an over-all judgment regarding the man's qualifications and presumably records this judgment in some way.

The procedure selected for use in the Aviation Psychology Program was the assignment of weights to the various test scores and the computation of a weighted average score for each individual. Since the importance of the

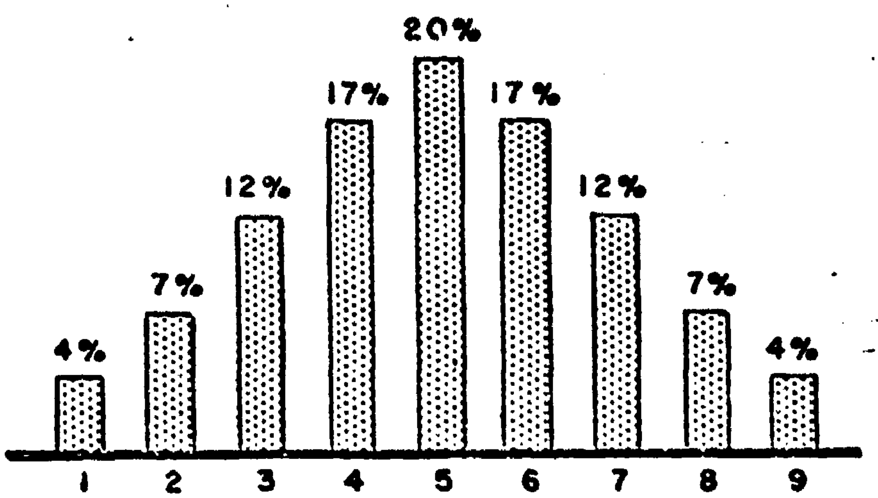



FIGURE 13.—Showing the definition of stanines.

various tests was generally different for each of the various air-crew positions there were as many weighted averages obtained as there were different air-crew positions for which the man might be classified.

To simplify the interpretation and use of these weighted average aptitude scores they were converted to a simple and convenient nine-point scale. This was done by assigning various percentages of the total distribution to the various score intervals beginning with the lowest 4 percent receiving scores of 1 and going on until the highest 4 percent received scores of 9. This is illustrated in figure 4.3.

These values represented normalized standard scores with a mean of five and a standard deviation of approximately two, on a nine-point scale. The term stanine was coined to facilitate reference to these scores.

The weighting procedure allowed one test score which was unusually high on a trait which had been found to be important to compensate for a weak score on another important test or on each of several less important ones. On the basis of experience it was possible to modify the weights to approximate more closely the optimal ranking of the individual with respect to aptitude for the particular position.

The recommendations made by the Psychological Research Unit for the classification of men as pilots, bombardiers, and navigators were based on two principal sets of information, first the man's aptitude scores as expressed in terms of stanines for the particular aircrew position, and, second, his preferences which were stated in terms of a first preference and a strength of interest rating on a nine-point scale on which nine represented extremely strong interest in that particular type of training and one represented little or no interest.

The preference form also contained four statements regarding the manner in which he wished to have his aptitude scores used in relation to his preferences in assigning him to one of the types of training. If he checked either of the first two statements he indicated that he wished to be sent to the type of training for which he was found to have most aptitude or preferred to be sent there if the difference in aptitude between that and his first preference were great. The third statement indicated that he wished to be assigned to the type of training for which he had indicated first preference unless it appeared that he was likely to fail in that type of training. The fourth and last statement indicated that he wished to be assigned to the type of training indicated by his first preference regardless of his aptitudes as found on the basis of the tests.

For a sample of approximately 200,000 aviation cadets and aviation students who checked these statements from 1942 to 1945, approximately 40 percent checked one of the first two statements, 52 percent the third statement, and only 8 percent the last statement.

The general principles followed in making recommendations were: (1) to give those having high aptitude for everything their first choice; (2) to interview and encourage those individuals with much higher aptitude scores for some type of training other than their first preference to modify their first preference; and finally (3) to recommend those whose aptitudes for all types of training were marginal for that type of training in which their aptitude was highest. It was intended that no one would be assigned to a type of training other than his first preference without an interview in which he was told in general terms of the nature of his aptitudes as shown by the test results.

When the classification procedures were first initiated, all persons who had passed the AAF Qualifying Examination and were sent to classification centers were assigned to either pilot, bombardier, or navigator training if they were found physically qualified for flying. Shortly after the procedures were initiated, however, it was found that men with low navigator stanines were unable to do satisfactory work in most instances in navigation training schools. Since the quotas for this type of training were relatively small, an effort was made almost from the outset to classify men for this type of training who had at least average aptitude for navigator.

In July 1942 the average stanine of five was established as the minimum qualifying stanine for navigation training. As was seen in the previous chapter, as further experience accumulated regarding pilot training, qualifying standards were set for this and also bombardier training. After October 1944 a stanine of 7 or better was required in any type of training leading to a commissioned rank. With the establishment of such high standards the Air-Crew Classification Tests became essentially a second screening. The classification procedure was simplified. Those who were qualified for all types of training were given their first preference. Those remaining were assigned to the most preferred type of training among those for which they were qualified. These procedures had to be modified occasionally to meet quota requirements.

The general reaction of the aviation cadets to the classification procedures used was excellent. As would be expected there were occasional individuals who appeared to be upset at not having been assigned to the type of training they had listed as first preference. However, the general experience was that they adjusted very quickly after being assigned to a specific type of training. In the main, the attitude toward assignment in accordance with the aptitude-test results was excellent as indicated by the percentages checking the various statements as reported above. A fairly typical reaction is indicated by an excerpt from a personal letter written by an aviation cadet to a friend as quoted below:

> . . . I enlisted in the Air Corps in February at Camp Forest, Tennessee, and was ordered to Maxwell Field on May 6, 1942. Have therefore been here about a week as aviation cadet and like it very much. Just finished today the psychological tests for classification purposes. . . . I'd like to say that they're the most thorough and accurate I could have imagined. Naturally it's my desire to be classified as a pilot. I enlisted with that in view and I haven't changed my mind, nor do I intend to unless and until I learn I am not qualified to fly. However, that will take care of itself.
>
> I like the place very much—the fine uniform, excellent food, and a finely planned training program. Impatience is my only worry so far.

During the war more than 600,000 men took the complete Air-Crew Classification Test Battery. Of this group about 250,000 were assigned to pilot training, approximately 50,000 were assigned to navigation training, and about the same number were assigned to bombardier training. A few thousand were sent directly into training as radar observers and as flight engineers. When given the more complete physical examination for flying training at the classification centers or basic training centers, more than 50,000 were found physically disqualified with respect to the types of training previously mentioned. Of those found fully qualified on the physical examination for flying duty, nearly 100,000 were disqualified on the basis of low aptitude scores (stanines). Many of the men who passed the AAF Qualifying Examination but lacked certain of the special aptitudes necessary for these types of training for officer assignments were trained as mechanics, radio operators, and gunners and became enlisted members of the air crew.

## ROUTINE STUDIES OF THE CLASSIFICATION TESTS

### Early Validation Studies

As indicated previously, it took several months to get a satisfactory battery of classification tests into use in the classification centers. However, by October 1942 data were available for a number of tests based on the training records of several hundred individuals in primary flying schools. These data gave partial confirmation and guidance in selecting the battery of tests to be used beginning the first of December 1942.

By February 1943 adequate data were available to confirm the validity of the pilot stanine. A booklet entitled "Psychological Traits of Successful Pilots (Research Findings of 1942)" was prepared by the Office of the Air Surgeon, Headquarters Army Air Forces and published in February 1943. This booklet included a chart based on a follow-up of the results on classes 43-D and 43-E for the Gulf Coast and West Coast training centers. The aviation cadets in these classes were tested during the latter part of the summer of 1942 at the Psychological Research Units in the respective training centers. This chart, which is reproduced as figure 4.4, was based on nearly 10,000 cases.

In these classes approximately a thousand men received pilot stanines of 9. Out of this group only 37 men or 3.7 percent failed to complete primary training because of flying deficiency. Only a small number (314 in all) of the men with a score of 1 were sent into training, but out of these men who were sent into pilot training in spite of low stanines, 69 percent were eliminated for flying deficiency before they completed primary training. It is probable that many of the 30 percent who did finish this stage were later eliminated during basic and advanced training.

This booklet also included charts indicating the importance of tests of judgment, mechanical comprehension, and visual-motor coordination, each of which was based on a sample of more than a thousand cadets. Data were presented for smaller samples, including several hundred cases indicating the

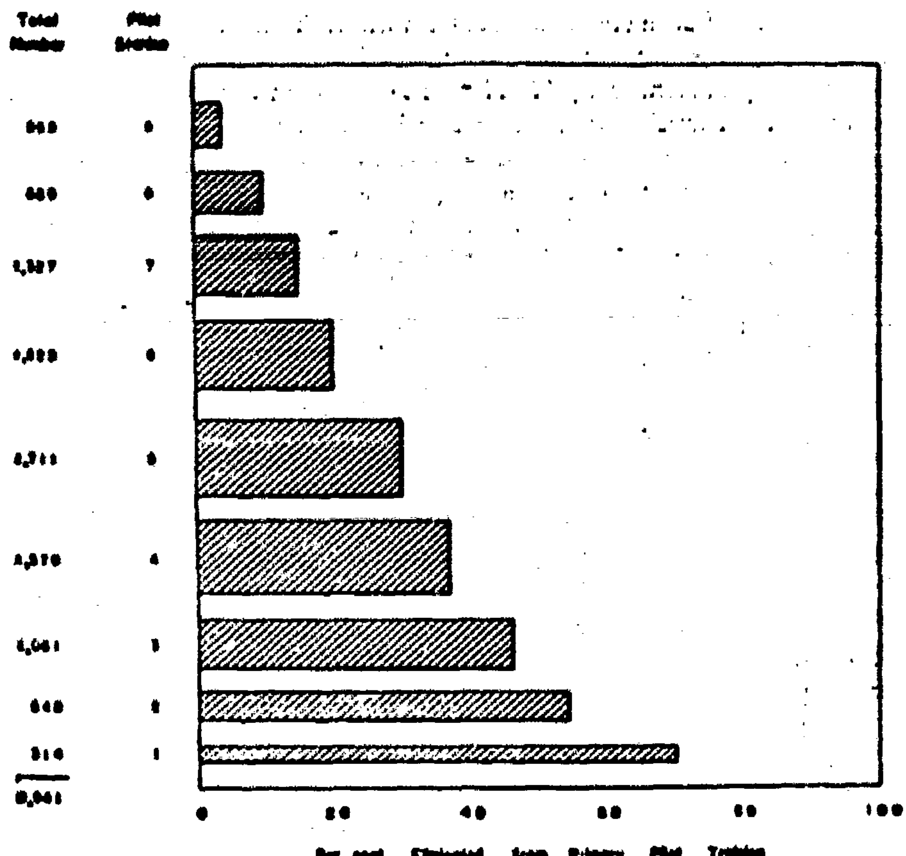



FIGURE 4.4.—The percent of cadets eliminated from primary pilot training at each stanine level. These results are based on Classes 43-D and 43-E for the Gulf Coast and West Coast training centers. The stanines include credit for previous flying experience. Elimination was for flying deficiency.

probable validity of tests of foresight, the ability to visualize, flexibility of attention, and interest in aviation.

On the basis of samples of more than a thousand cases each, it was also reported that tests of vocabulary and mathematics had been found to be of no value in selecting men for pilot training. The importance of this finding for classification was emphasized in the discussion in the booklet since both of these latter types of tests had been found to have substantial predictive value for navigation training.

By the spring of 1943 sufficient data had accumulated regarding the validity of various classification tests for pilots so that relative weights for computing pilot stanines could be prepared with knowledge of the intercorrelation of the tests based on several thousand cases and validity coefficients between the various tests and graduation-elimination from primary flying training obtained from 4,000 to 8,000 cases for most tests and on more than 10,000 cases for four of the tests.

Because of the smaller number of men sent into navigator training, data regarding the predictive value of tests for this purpose were somewhat slower

703123—47—6

in accumulating. However, by February 1943 data on 500 to 1,000 cases were available for quite a large number of classification tests. Practically all types of mathematics tests were found to have substantial predictive value for the selection of navigators. Other tests of importance were dial-and-table-reading tests, tests of reading comprehension and judgment, and certain perceptual tests involving speed of identification of forms and the use of maps and photographs.

The very small number of individuals eliminated from bombardier training schools, together with the lack of any agreement between the final grades received in these schools and the records of the circular errors of bombs dropped by the students, made the evaluation of tests for bombardier training extremely difficult. The accuracy of the bombing done by the students appeared to be a naturally relevant and desirable criterion of success, but it was found that so many extraneous factors influenced the bombardier's actual results that there was little consistency in the accuracy of his bombing from one training mission to the next. Under these circumstances exceptionally large numbers of cases were required to determine the relative predictive value of the various tests with a satisfactory degree of confidence.

### Illustrative Classification Data

The functioning of classification procedures can be better understood if some actual examples are given. Table 4.1 shows the indicated strengths of interest, stanines, preference waivers, original assignments, and a brief statement concerning the ultimate disposition of a group of 45 aviation cadets from a prominent New England college. These men were tested together at the AAF Classification Center at Nashville in the spring of 1943 and followed up to illustrate the disposition of a fairly homogeneous group with a better than average background. The records of the entire group of 45 men are reproduced here.

TABLE 4.1.—*Illustration of classification procedures as they operated on 45 individuals from a prominent New England college who entered training as a group in the spring of 1943.*

| Name | Stanine B | Stanine N | Stanine P | Preference B | Preference N | Preference P | Preference Waiver[a] | Type Training | Training Results | Post Training Career |
|---|---|---|---|---|---|---|---|---|---|---|
| Frank, N. | 5 | 6 | 3 | 9 | 4 | 8 | 3 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | 7 months overseas service. Received Air Medal with 4 Oak Leaf Clusters and EAME Theater Ribbon with 3 Bronze Stars. Was 1st Lt. 31 Dec. 45 (last date reported). Efficiency rating 4.3, excellent. |
| Charles, M. | 9 | 8 | 9 | 5 | 4 | 9 | 3 | P | Graduated | 13 months overseas service. 150 combat hours. Received Air Medal with 2 Oak Leaf Clusters. Was 2nd Lt. as of May 1945 (last date reported). |
| William, A. | 7 | 7 | 6 | 3 | 1 | 6 | 3 | P | do | No overseas service—flying instructor. Was 1st Lt. 30 June 1945 (last date reported). Efficiency rating 4.5, excellent. |
| Hamilton, C. | 8 | 8 | 8 | 5 | 4 | 9 | 3 | P | do | 11 months overseas service; noncombat. 1st Lt. 30 Aug 45 (last date reported). Efficiency rating 4.3, excellent. |

TABLE 4.1.—*Illustration of classification procedures as they operated on 45 individuals from a prominent New England college who entered training as a group in the spring of 1943.—(Continued).*

| Name | Stanine B | Stanine N | Stanine P | Preference B | Preference N | Preference P | Preference Waiver[4] | Type Training | Training Results | Post Training Career |
|---|---|---|---|---|---|---|---|---|---|---|
| Dorsett, C. | 6 | 7 | 4 | 3 | 3 | 9 | 3 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | On last record available, April 1944 was at an Air Force Training Station in Idaho. |
| Frederick, M. | 6 | 4 | 6 | 8 | 4 | 9 | 3 | P | Graduated | Length of overseas service unknown. Was 1st Lt. 30 June 1945 (last date reported). Efficiency rating 4.8, excellent. |
| David, E.[2] | | | | | | | | | | |
| Tracy, S. | 3 | 6 | 7 | 5 | 2 | 7 | 3 | P | do | 35 combat missions; 13 sqdn. leads and 2 element leads. Received Air Medal with 4 Oak Leaf Clusters and the DFC. Was 1st Lt. 31 Dec. 1945 (last date reported). |
| Philip, P. | 5 | 8 | 3 | 6 | 9 | 8 | 3 | N | do | Served overseas. Was 1st Lt. in Aug. 1945 (last date reported). Efficiency rating 5.7, superior. |
| James, G. | 5 | 5 | 6 | (9) | (9) | (9) | | N | do | 11½ months overseas service; 28 combat missions. Received DFC, Air Medal with 3 Oak Leaf Clusters and EAME Theater Ribbon with 2 stars. 1st Lt. 1 July 45 (last date reported). |
| John, W. | 6 | 8 | 7+3 | (9) | (9) | (9) | | P | do | 26 combat missions. Received Asiatic-Pacific Theater Ribbon with 4 Bronze Stars and the Philippine Liberation Ribbon with Bronze Star; also Purple Heart. |
| Richard, C. | 8 | 8 | 7 | 2 | 3 | 9 | 4 | P | do | 16 months overseas service. 111 combat missions. Received the DFC, Air Medal with 5 Oak Leaf Clusters and EAME Theater Ribbon with 6 stars. Was 1st Lt. 14 May 1945 (last date reported). |
| George, A. | 4 | 6 | 2 | 6 | 3 | 9 | 4 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | 37 combat missions. Received Air Medal with 3 Oak Leaf Clusters and EAME Theater Ribbon with 2 stars. 1st Lt. 15 May 1945 (last date reported). Efficiency rating 4.1, excellent. |
| Alfred, J. | 3 | 6 | 4 | (9) | (9) | (9) | | P | Graduated | (9) |
| James, H., Jr. | | | | (9) | (9) | (9) | | P | Eliminated for physical reasons. | (9) |
| James, H. | 4 | 6 | 4 | (9) | (9) | (9) | | P | Eliminated for flying deficiency. | (9) |
| Earl, F. | 3 | 6 | 3 | 7 | 7 | 7 | 1 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | 13 months overseas service. 30 combat missions. Received Air Medal with 1 Oak Leaf Cluster, Asiatic-Pacific Theater Ribbon with 4 stars and Philippine Liberation Ribbon with 1 star. 1st Lt. 30 June 1945 (last date reported). |
| Cornelius, H. | 2 | 4 | 2 | (9) | (9) | (9) | | P | Eliminated for flying deficiency. | (9) |
| John P. | 7 | 8 | 8+3 | 2 | 6 | 9 | 3 | P | Graduated | Overseas service 7 months; 30 combat missions. Received DFC, Air Medal with 3 Oak Leaf Clusters and EAME Theater Ribbon with 2 stars. Was Captain 6 Mar. 1946 (last date reported). |
| Don, W. | 9 | 9 | 9 | 3 | 6 | 9 | 3 | P | do | No overseas service—flight instructor. Rating 6.2, superior. Was 1st Lt. 14 May 1945 (last date reported). |
| Robert, B. | 8 | 9 | 8 | 3 | 7 | 9 | 3 | P | Eliminated for flying deficiency. Went into bombardier training. Graduated. | (9) |
| George, E. | 8 | 8 | 6 | 7 | 8 | 9 | 3 | N | Graduated | Length of overseas service unknown. Was German P.O.W. for some time. |

TABLE 4.1.—*Illustration of classification procedures as they operated on 45 individuals from a prominent New England college who entered training as a group in the spring of 1943.*—(*Continued*).

| Name | Stanine B | Stanine N | Stanine P | Preference B | Preference N | Preference P | Preference Waiver | Type Training | Training Results | Post Training Career |
|---|---|---|---|---|---|---|---|---|---|---|
| Robert, W. | 6 | 5 | 6 | (9) | (9) | (9) | | P | do | (9) |
| Thomas, W. | 7 | 6 | 8 | 6 | 6 | 9 | 3 | P | do | 61 combat missions. Received Air Medal and 9 Oak Leaf Clusters, EAME Theater Ribbon with 3 Bronze Stars and the DFC. Captain 4 Aug. 1945 (last date reported). Efficiency rating 5.2, excellent. |
| Richard, M. | 5 | 6 | 8 | 7 | 9 | 8 | 3 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | Overseas service 7 months; 49 combat missions. Received Air Medal with Oak Leaf Cluster, Purple Heart, DFC with 1 Oak Leaf Cluster. Was 1st Lt. 28 March 1945 (last date reported). Efficiency rating 4.2, excellent. |
| Frank, L., Jr. | 4 | 6 | 3 | 7 | 4 | 9 | 3 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | Overseas service 8 months. 48 combat missions. Received DFC, Air Medal with 1 Oak Leaf Cluster and Asiatic-Pacific Theater Ribbon with 4 stars. Was 1st Lt. 8 June 1945 (last date reported). Efficiency rating 4.0, excellent. |
| Henry, D. | 4 | 4 | 4 | (9) | (9) | (9) | | P | Eliminated for flying deficiency. | (9) |
| Arthur, L. | 7 | 3 | 6 | 6 | 1 | 9 | 3 | P | Graduated | Overseas service 13 months; 3 combat missions. Received the EAME Ribbon with 1 star. Was 1st Lt. 13 Nov. 1945 (last date reported). Was made 1st Lt. in Air Corps Reserve 8 Feb. 1946. |
| William, R. | 3 | 5 | 6 | (8) | (9) | (9) | | P | do | Length of overseas service not known. Received DFC, the Air Medal with 5 Oak Leaf Clusters and the Presidential Unit Citation. Had 39 combat missions as pathfinder lead pilot. Was 1st Lt. 30 June 1945 (last date reported). |
| Malcolm, M. | 9 | 6 | 7 | (8) | (9) | (9) | | P | do | Length of overseas service not known. Was 1st Lt. 1 May 1945 at time of liberation as German P.O.W. |
| Charles, W. | 6 | 6 | 6 | 7 | 5 | 9 | 3 | P | do | Overseas service 13 months; 50 combat missions. Received the Air Medal with 3 Oak Leaf Clusters and EAME Theater Ribbon with 3 Bronze Stars. Final grade 1st Lt. Appointed 1st Lt. in Air Corps Reserve 30 March 1946. |
| Campbell, M. | 5 | 4 | 5 | 5 | 3 | 9 | 3 | P | do | Overseas service 14 months; 61 combat missions. Received Air Medal with 3 Oak Leaf Clusters, EAME Theater Ribbon with 4 Bronze Stars. Was captain 16 Jan. 1945 (last date reported). Efficiency rating 6.0, superior. |
| Joseph, B., Jr. | 5 | 6 | 4 | 3 | 4 | 9 | 3 | P | Eliminated. Went into navigation. Graduated. | No overseas service. Was continuation trainer, Ellington Field as of 30 June 1945. Was 1st Lt. 30 June 1945 (last date reported). |
| Thomas, J. | 6 | 8 | 4 | 1 | 6 | 9 | 3 | P | Eliminated for flying deficiency. Went into bombardier training. Graduated. | 6 months overseas service; 31 combat missions. Received Air Medal with 3 Oak Leaf Clusters, DFC and EAME Theater Ribbon with 2 stars. Was 1st Lt. Dec. 1945 (last date reported). Efficiency rating 4.3, excellent. |
| Stanley, L. | 3 | 8 | 4 | 3 | 3 | 7 | 1 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | 30 combat missions. Received DFC, Air Medal with 3 Oak Leaf Clusters, ETO Ribbon with 1 Bronze Battle Star, EAME Ribbon with 2 stars and Air Crew Member Badge. Was 1st Lt. 30 June 1945 (last date reported). |

TABLE 4.1.—*Illustration of classification procedures as they operated on 43 individuals from a prominent New England college who entered training as a group in the spring of 1943—(Continued).*

| Name | Stanine B | Stanine N | Stanine P | Preference B | Preference N | Preference P | Preference Waiver[1] | Type Training | Training Results | Post Training Career |
|---|---|---|---|---|---|---|---|---|---|---|
| Arthur, A. | 3 | 4 | 2 | (3) | (3) | (3) | | P | Eliminated for flying deficiency. | (3) |
| Herdmann, S. | 7 | 6 | 7 | 6 | 5 | 9 | 5 | P | Graduated | 30 combat missions. Received the DFC, Air Medal with 3 Oak Leaf Clusters and EAME Theater Ribbon. Was 1st Lt. 21 July 1945 (last date reported). Efficiency rating 4.0, excellent. |
| John, S. | 6 | 7 | 5 | 8 | 8 | 9 | 3 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | Length of overseas service not known. Was 1st Lt. 1 Jan. 1946 (last date reported). |
| William, C. | 6 | 7 | 6 | 3 | 3 | 9 | 4 | P | Eliminated for flying deficiency. Transferred to bombardier training. Graduated. | Overseas service 6 months; 35 combat missions. Received the DFC, Air Medal with 3 Oak Leaf Clusters and EAME Ribbon with 4 stars. Was 1st Lt. 30 June 45 (last date reported). |
| Donald, O. | 5 | 7 | 6 | 3 | 3 | 9 | 3 | P | Graduated | Length of overseas service not known; 3 combat missions. EAME Theater Ribbon with 1 star. Was 1st Lt. 30 June 1945 (last date reported). Efficiency rating 3.87, excellent. |
| Arthur, H.[4] | 7 | 6 | 7 | | | | | | | |
| William, H. | 5 | 4 | 6+2 | 8 | 3 | 9 | 1 | P | Graduated | Length of overseas service not known. Escapee—bailed out on 11th mission over Germany. P.O.W. from 20 July 1944 until escape 3 Sept. 1944. He rejoined Allied Forces 3 Sept. 1944. Received Purple Heart and EAME Ribbon with 1 Bronze Star. Was 1st Lt. 24 Aug. 1945 (last date reported). |
| Robert, L. | 4 | 7 | 4 | 6 | 3 | 8 | 3 | P | Eliminated for flying deficiency. Went into navigation. Graduated. | 7 months overseas service; 35 combat missions. Received Air Medal with 3 Oak Leaf Clusters and EAME Theater Ribbon with 3 Bronze Stars. Was 1st Lt. July 1945 (last date reported). |
| Lowell, T. | 8 | 7 | 7 | (3) | (3) | (3) | | P | Graduated | No overseas service. Flying instructor. Was 2nd Lt. 26 Sept. 1943 (last date reported). |
| Larned, A. | 7 | 7 | 6+3 | 4 | 3 | 9 | 3 | P | do | No overseas service. Instructor. Was 2nd Lt. 30 June 1945. (last date reported). |

[1]Refer to page 71 for description of preference waivers
[2]Discharged on CDD 4 March 1943.
[3]No record located.
[4]Deceased. Result of aircraft accident 21 July 1943.

The first individual, Frank, N., was sent into pilot training by the personnel officers in spite of having indicated a greater interest in bombardier training.' This is surprising since his pilot stanine of 3 indicated a fairly low aptitude for pilot training. He showed higher aptitudes for both bombardier and navigator training as indicated by stanines of 5 and 6 respectively for these positions. As might have been expected from his low pilot stanine, he was eliminated for flying deficiency. His navigator stanine of 6 enabled him to enter navigator training after this elimination. He had originally indicated only a moderate degree of interest, 4, in navigator training and highest interest, 9, in bombardier training. The rules in effect at that time required that he have a stanine of at least 6 in the specialty to be eligible for a second type of aviation cadet training after having been eliminated once. He was successful in navigation training and graduated from the course and served satisfactorily overseas as a member of a combat crew. This man should have been sent into bombardier or navigator training originally. The final decision could probably best have been made on the basis of an interview with him regarding his aptitudes and preferences and with regard to quota requirements.

The next three men on the list all had fairly high stanines for all three types of training. They all had a high interest in pilot training and not much interest in other types of training. As indicated by the 3's in the Preference Waiver column, they all wished to be sent into the type of training of their first preference unless their aptitude scores were so low as to indicate their probable failure in that type of training. They all graduated from pilot training and served in useful assignments during the war period.

The fifth man, Dorsett, C., has a stanine pattern much like the first man and a very similar history of elimination from pilot training and later graduation from navigation training. Frederick, H., had a fairly low navigator stanine, 4, but an adequate pilot stanine and completed pilot training. David, K. was found not qualified for military service of any type on the physical examination and given a medical discharge. Tracy, S. had a high pilot stanine and made an excellent record in combat after completing pilot training.

The case of Philip, P. is interesting in that he was found to have very high aptitude for navigation training, a navigator stanine of 8, and a very low pilot stanine of 3. As is quite frequently the case, his interests paralleled his aptitudes to some extent, his highest interest also being for navigation training. He was sent directly into navigator training and graduated from that course. For some reason James, G. was also sent into navigator training in spite of a very high pilot stanine. Unfortunately, no record was found of his preferences in the files at hand. He appears to have done an adequate job in his navigation work.

The next two cases follow a familiar pattern. The first of these had a pilot stanine of 7 with three points extra credit for having held a private pilot's license prior to entering the service. They were both successful in pilot training. George, A. with a very low pilot stanine was eliminated as a pilot but later graduated from navigation training. Alfred, J. graduated from pilot training in spite of a fairly low pilot stanine.

The other cases tend to follow the same pattern. Occasionally a cadet with a moderately high stanine is eliminated and once in a while a low-stanine cadet graduates. However, the general results illustrate the agreement of training success with the predictions from the aptitude scores as shown by the stanines. Statistical studies based on more extensive data are reported in the next section.

## Cumulative Follow-Up Studies

The results shown in table 4.1 are intended merely to illustrate the actual procedures for a small group of specific individuals. With the aid of an electrical collating and tabulating machine installation at Headquarters AAF

Training Command, the records of all men tested were recorded on punched cards and their subsequent records in the training schools were also punched from rosters received from the schools and collated with the original test scores and stanines. Figure 4.5 shows the relative success in primary flying schools of more than 185,000 men who were assigned pilot stanines prior to being sent into training on the basis of the Air-crew Classification Test scores. This chart shows eliminations for all reasons and indicates the substantial predictive value of the pilot stanine as it was maintained throughout the war.

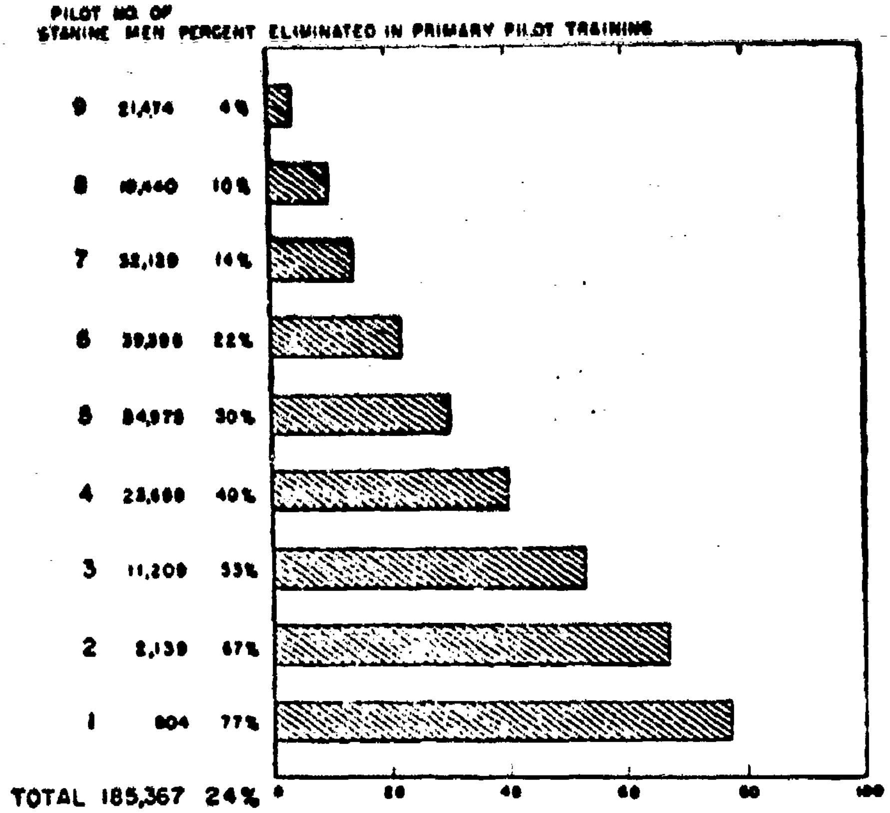



FIGURE 4.5.—The percent of cadets eliminated from primary pilot training at each stanine level. These results are based on Classes 43-F through 45-H at all training commands. Elimination was for flying deficiency, fear and own request. Flying experience credit is included in the stanine score.

It will be noted that the number of men with low aptitude scores sent into training is relatively small. As previously indicated, the qualifying standards in terms of stanines were established after several months of experience had proved their value and these qualifying standards were raised successively during the war period. This improvement in the quality of the personnel entering into pilot and other types of air-crew training enabled the training-school authorities to raise the standards with respect to instruction and also the requirements for graduation.

A similar chart is shown in figure 4.6 giving the results of follow-up studies on more than 15,000 men sent into navigation training schools. It will be noted that the numbers of men with very low stanines sent into navigation were negligibly small after the summer of 1942. The chart shows that very few of the men who met the high standards in effect in the latter part of the war were eliminated from navigation training.

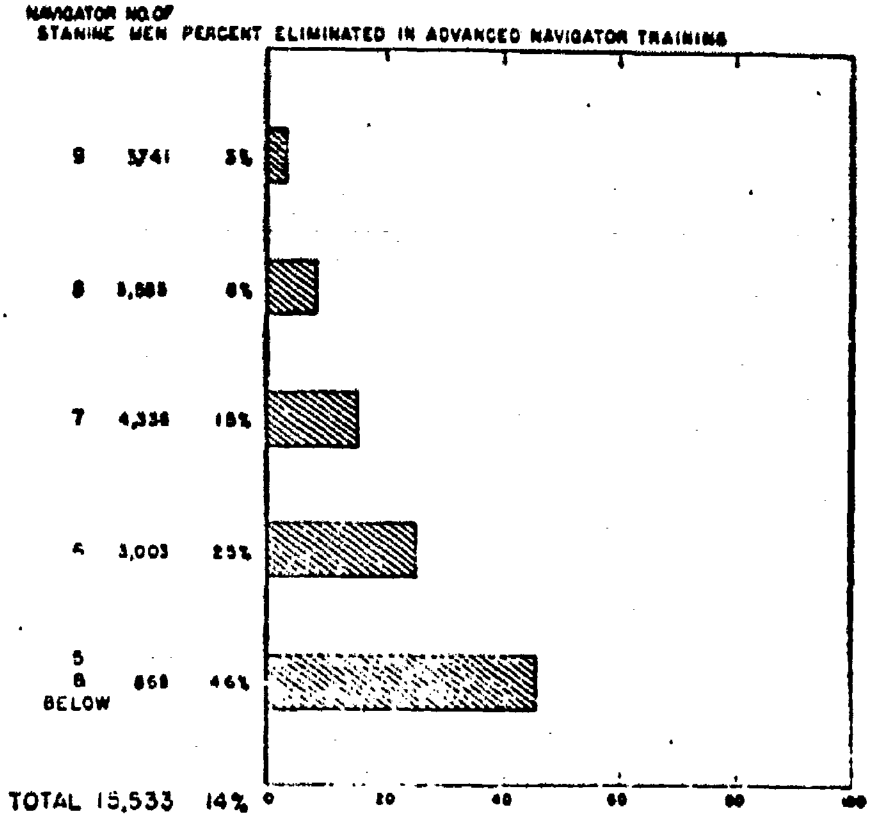



FIGURE 4.6.—The percent of cadets eliminated from advanced navigator training at each stanine level. These results are based on classes 43-12, 43-13, 43-14, 43-15, and 44-1 through 45-13. Elimination was for flying deficiency, fear, and own request. Only new aviation cadets are included.

It has already been pointed out that bombardier evaluation problems presented special difficulties. In figure 4.7 are shown the results from following up several thousand men sent into bombardier training schools. It will be noted that the proportion of men both entering and graduating from bombardier training who had low stanines is significantly greater than for the other two types of training.

The rejection of men with low stanines prior to the time they were sent into training, changes in training conditions and standards, and the unknown effect of the initial screening accomplished by the various forms of the AAF Qualifying Examination, emphasized the desirability of a simple and unambiguous evaluation of selection and classification procedures. This led to the initiation of certain special projects which will be described in the next section.

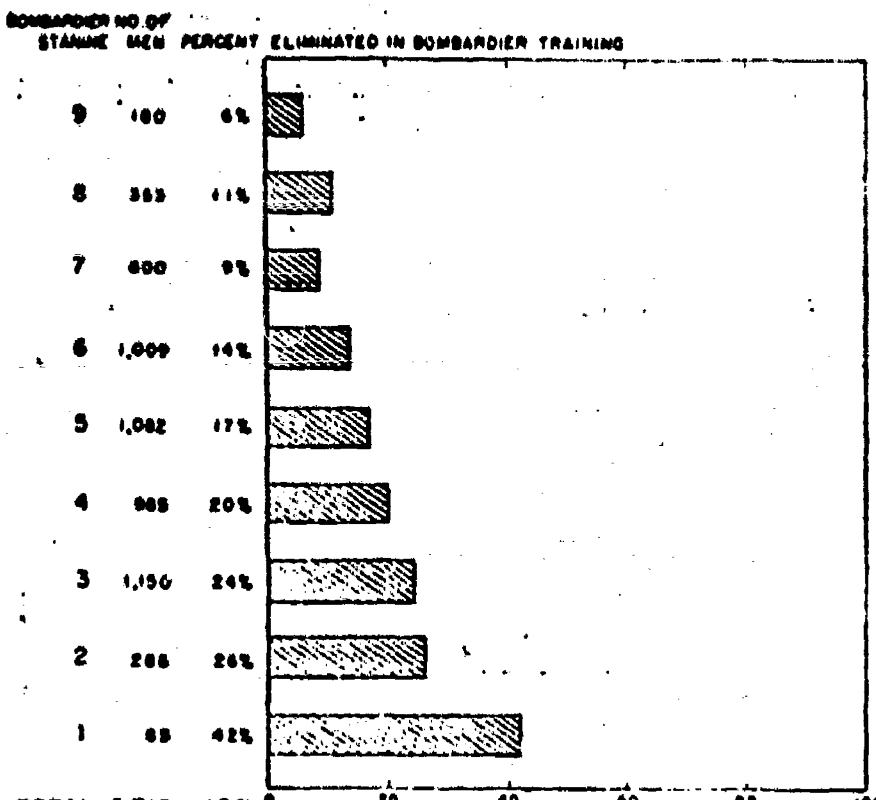



FIGURE 4.7.—The percent of cadets eliminated from advanced bombardier training at each stanine level. These results are based on Classes 43-8 through 16. Elimination was for unsatisfactory progress, fear, and own request. Only new aviation cadets are included.

## THE EXPERIMENTAL STUDY OF A THOUSAND APPLICANTS SENT INTO PILOT TRAINING

### The Planning of the Experiment

A common problem for research workers concerned with the development and improvement of procedures for the selection and training of personnel is the adequate evaluation of procedures after they have been established. Educational institutions, business and industrial concerns, and government organizations, having once accepted certain procedures are generally opposed to suspending the use of these procedures for a large enough group to obtain an adequate evaluation of the new procedures. This makes it very difficult to refine and make further improvements on the procedures.

Because of the very large numbers of men involved and the great importance of the procedures for the selection of air crew, such an evaluation appeared especially desirable in this situation. It was believed that a check on the value and interrelation of both the initial screening procedures and the procedures for qualifying men for pilot training on the more comprehensive Air-crew Classification Tests should be made. This could be accomplished by examining a large sample of applicants with these tests and sending all of the men tested into training regardless of the test results. Accordingly, a memorandum was prepared entitled "Experimental Study of Eligibility Requirements for Aviation Cadets" by the Chief of the Psychological Branch in May 1943.

Varied responses were obtained to this proposal from representatives of other divisions of the Air Staff. Certain of the Regular Army officers felt that since procedures had been accepted and appeared to be working well, it was unwise to conduct a study which might reveal serious defects and weaknesses. Others stated that research of this type should be carried on in peacetime and should not be allowed to interfere with established routines for the selection and training of men during the war period. One officer suggested that the proposal to bring in a thousand applicants regardless of test results was inconsistent with proposals by aviation psychologists that qualifying standards in terms of pilot stanines be raised. Other officers questioned the study on the grounds that the value of these new procedures had already been established and that further studies were therefore unnecessary. However, the argument that the procedures were not perfect and that further improvement depended upon such an evaluation won out and on 21 June 1943 the study was approved and a letter was sent from the Commanding General, Army Air Forces, to the Commanding General, Army Service Forces, requesting the cooperation of the Aviation Cadet Examining Boards in the nine Service Commands in recruiting this group.

During the preliminary discussions in the Office of the Air Surgeon it was decided to require full qualification of this group on the regular physical examination. However, the surgeons of the Aviation Cadet Examining Boards were told that if the applicant was otherwise physically qualified he should not be disqualified by reason of a low Adaptability Rating for Military Aeronautics. At the classification centers, instructions were also given that no one was to be rejected from the group except for purely physical reasons.

Approximately 40 boards, representing all of the 9 Service Commands and including all sections of the country, were authorized to recruit members of the experimental group. Each board was given a definite quota. The quotas varied with the size of the population of the area serviced by the board. The smallest quotas were for 20 aviation cadets and the largest for 75. In establishing the quotas for the various Service Commands the numbers recruited from that Service Command in previous months were also considered. This was especially important since some of the Service Commands contained a number of boards based at Army posts or stations at which men already in the service could apply. The quotas for all Service Commands totaled 1,450 men. It was believed that this would allow for a certain number of later physical disqualifications and other losses and still provide a group of more than a thousand entering pilot training.

## Recruiting the Group*

To insure that the personnel of the Boards understood the general plan and the specific procedures to be followed, an officer from the Psychological Branch, Research Division, Office of the Air Surgeon, was sent at some time in July to each of the Boards which had been given a quota. At the time these men were being recruited the normal procedure was, first, to be sent to basic training centers for 6-weeks' basic training, then to college for approximately 5-months' preaviation-cadet college training, and after that to preflight school for about 2 months. Following this the individual was sent to primary flying or one of the other air-crew specialty schools.

Since it was desired that the results of this experiment be available as quickly as possible, it was decided that the preaviation-cadet college course would be omitted for these men. Accordingly, beginning about 1 August 1943 all applicants at the authorized boards were given a statement to sign. This statement said, "I wish to enter pilot training. If I am found qualified by the Examining Board I agree (1) to enter pilot training after a shortened period of basic military training without first taking the preaviation-cadet college-training course, and (2) to volunteer for induction within ten (10) days following the day on which I am found qualified by the Examining Board." For enlisted men a similar blank form was provided except that it had no reference to basic military training or to volunteering. The examiner also read a statement to the men, pointing out the advantages of becoming an aviation cadet 5 months earlier and also having the opportunity to earn a pilot rating and become an officer that much sooner.

All applicants who signed the waiver were given the AAF Qualifying Examination and regardless of their score on this test given a physical examination and an interview by the board. If they were found physically qualified and had no criminal record they were qualified by the board for air-crew training. Records on these specially recruited men were sent directly to the War Department. In Washington special orders were written sending a large group of them at one time to a basic training center with special instructions for their disposition.

From the basic training center they were sent to a classification center where the Air-crew Classification Tests were given them. If found physically qualified they were sent into pilot preflight school regardless of the scores made on the Air-crew Classification Tests. The orders assigning these men to classification centers indicated that they were members of the experimental group. Upon completing their classification processing they were sent along with other aviation cadets to preflight schools with no designation as to which ones were members of the experimental group.

Thus, in preflight schools and in the training schools the members of the experimental group were not identified by the orders assigning them and received no special treatment. Since the service records of these men did not

*Chester W. Harris was responsible for planning the details of the recruiting procedures and selecting and visiting the AAF Examining boards. In the work of visiting these boards and explaining the recruiting procedures to them he shared this responsibility with William G. Mollenkopf.

contain their stanines for either pilot or other air-crew specialties the officers in charge of these schools were instructed to wire the AAF Training Command Headquarters for the disposition of any men whose stanines did not appear in their service records.

Orders were issued from Washington on 1,311 men recruited by the various AAF Examining Boards in accordance with the plan of this study. Of these 1,275 reached the AAF Classification Centers and were given the Air-crew Classification Tests. The test results of these men were processed in the usual fashion and sent to Headquarters AAF Training Command after stanines had been computed. When the more thorough physical examination was given at the classification center, a number of men were found disqualified for aviation-cadet training.

Of this group 671 men were tested at Psychological Research Unit No. 1, 365 at Psychological Research Unit No. 2, and the remaining 239 were scattered among the seven Medical and Psychological Examining Units. A small number were disqualified on the Adaptability Rating for Military Aeronautics during the physical examination in spite of directions to the contrary. A number of others were eliminated at the classification centers and no record sent to Headquarters as to the reason for elimination. The remaining 1,143 men were assigned to pilot preflight schools and this constitutes the primary sample on which this study is based.

### Description of the Sample

It is believed that the sample comprising the basic group for this experiment was thoroughly typical of *applicants* for aviation-cadet training. The average age was a little more than 21 with approximately 30 percent of the group 18 and 19. By far the largest age group was 19, and 10 percent were more than 26. From the standpoint of education, 2 percent were college graduates, an additional 16 percent had had some college training, 58 percent were high-school graduates, and the remaining 25 percent had not finished high school, including 1 percent who had never attended high school.

Approximately half of them were recruited from the Army and half from civilian status. With regard to previous flying experience, nearly 5 percent had flown solo and an additional 4 percent had had previous instruction. About 58 percent had been a passenger in a plane but had received no instruction, and 33 percent had never been a passenger in a plane. In this group 25 percent were married, 74 percent single, and 1 percent widowed, divorced, or separated.

Their average score on the Army General Classification Test was 113.0 with a standard deviation for the group of 13.8. Approximately 10 percent of the group achieved Army General Classification Test scores above 130 which placed them in category 1 and approximately 10 percent obtained scores below 95.

In this original group 58 percent obtained scores which would have normally passed them on the AAF Qualifying Examination and 42 percent would have been rejected on the basis of their scores. The average score was a few points higher than the passing mark and the standard deviation was approximately that which had previously been found for unselected applicants.

It is clear from their educational background, their Army General Classification Test scores, and their scores on the AAF Qualifying Examination, that this group does not represent a random sample of men of Army age. Rather, it represents approximately the usual amount of self-selection which can be expected in a group of applicants who have chosen to compete for a highly desirable job for which the requirements are relatively high both in terms of the examinations at the time of entrance and standards for retention in and graduation from the training schools.

To check whether the physical disqualifications and other losses at the classification centers had any important influence on the nature of the group, the average test scores of this group of 1,143 were compared with those of the total group tested. For practically all of these tests the differences between the means of the two samples were less than one- or two-hundredths of a standard deviation and in only one instance did it exceed five-hundredths of a standard deviation. It was therefore concluded that the losses in the classification centers had not introduced any significant bias in the samples.

### The Results*

Of the 1,143 men who were assigned to pilot preflight schools, 582 were eliminated in primary flying training schools, 83 were eliminated in basic training schools, and 24 eliminated in advanced flying schools. The remaining 265 graduated from advanced flying training and received ratings as qualified pilots. Of the 878 men eliminated, 99 were eliminated for academic deficiencies in preflight school, 591 were eliminated for flying deficiency at one of the three phases of flying training, and 65 were eliminated at their own request or because of fear of flying. The remaining 122 men were eliminated for administrative reasons, including physical disqualification. Approximately half of these were eliminated during preflight school.

Thus in this group of applicants who were allowed to enter pilot training without any screening for aptitudes, interests, or ability, only 23 percent were successful in completing the course of pilot training and becoming rated pilots. The question which the experiment was designed to answer was, "How well did the initial screening-test results, the various classification-test scores, and the pilot stanine predict which ones of this group would succeed?"

It was reported in a previous section that the AAF Qualifying Examination was found to be providing a rather satisfactory initial screening. Figure 4.8 shows the success of the pilot stanine in predicting which of these applicants would be successful. Very few of the 8's and 9's were eliminated in the training schools and of those that were, many were eliminated for physical or administrative reasons which the tests were not designed to predict. Nearly

*The principal analyses of results were carried on under the immediate supervision of Robert L. Thorndike and Walter L. Deemer in the Psychological Sections in Army Air Forces Headquarters and AAF Training Command Headquarters.

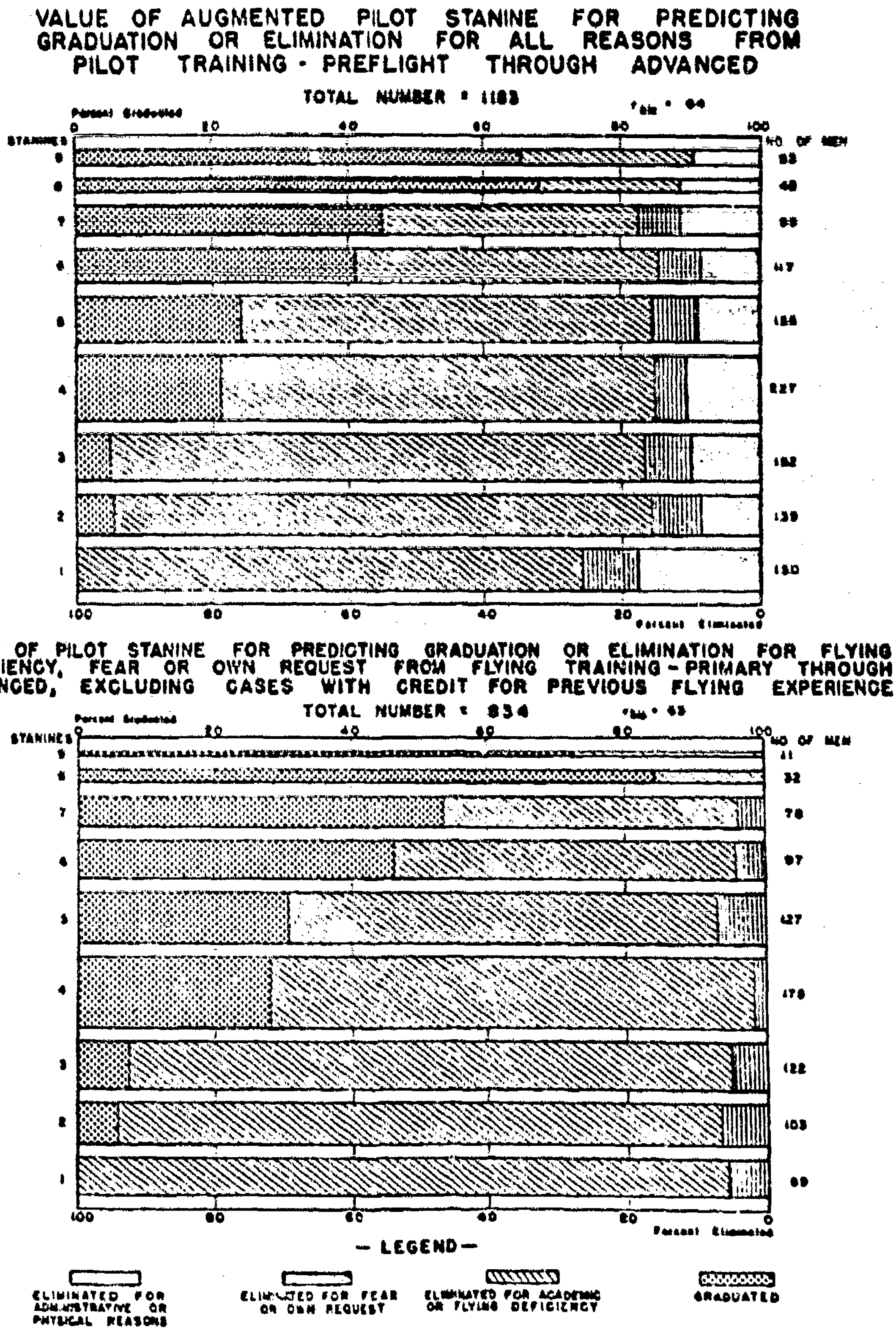
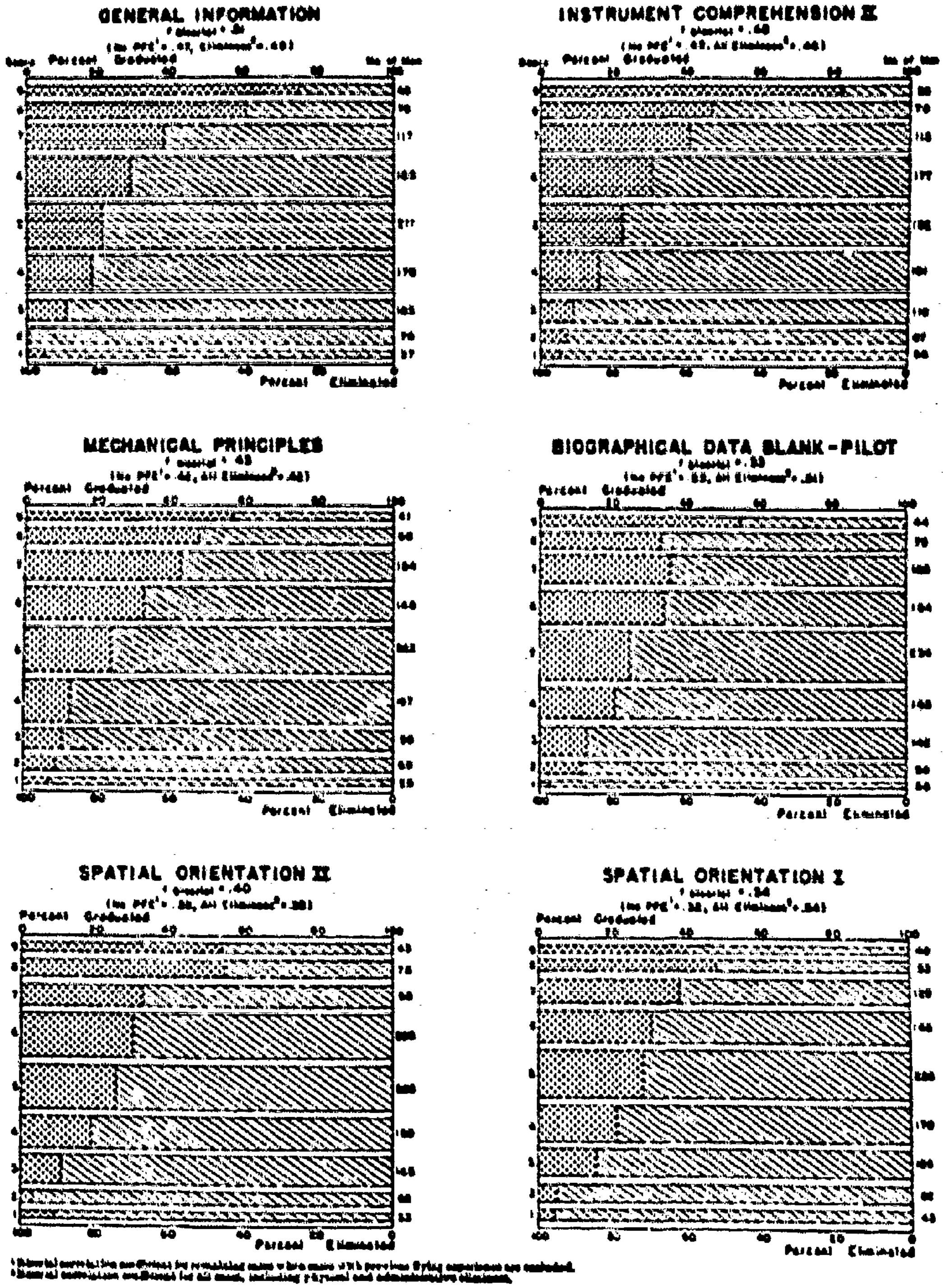
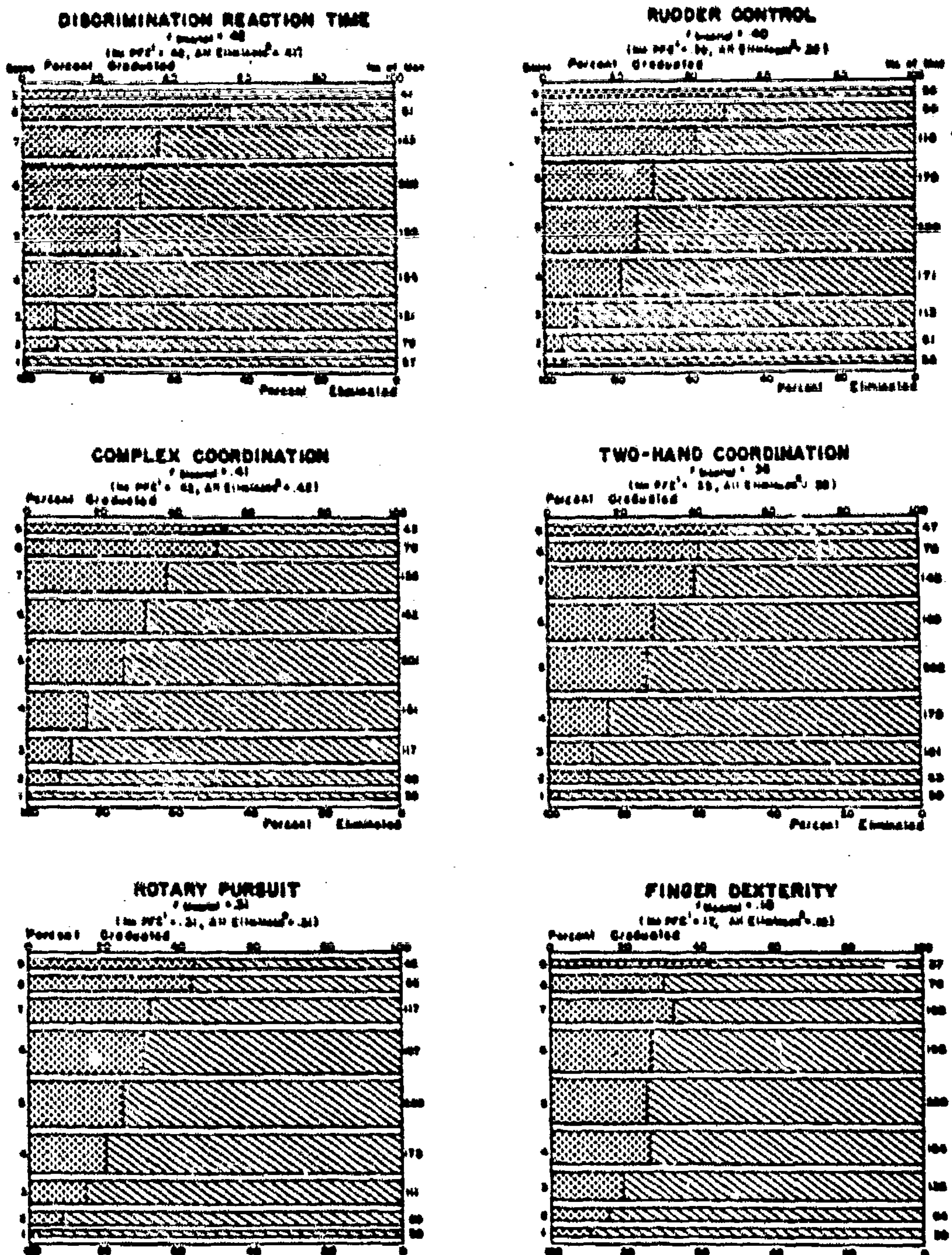



FIGURE 4.8.—Value of pilot stanine in experimental group.




[1] Biserial correlation coefficient for remaining cases when cases with previous flying experience are excluded.
[2] Biserial correlation coefficient for all cases, including physical and administrative eliminees.

FIGURE 4.9.—Predictive value for success in pilot training of printed tests with substantial weight in determining the pilot stanine. Data based on records of experimental group. Elimination was for flying deficiency, fear, and own request from preflight through advanced pilot training.




[1] Biserial correlation coefficient for remaining cases when those with previous flying experience are excluded.
[2] Biserial correlation coefficient for all cases, including physical and administrative eliminations.

FIGURE 4.10.—Predictive value for success in pilot training of printed tests developed primarily for the prediction of success in navigator and bombardier training. Data based on records of experimental group. Elimination was for flying deficiency, fear, and own request from preflight through advanced pilot training.

half of the 7's were successful in completing training, but only a quarter of the 4's and 5's and only a very small percentage of the 2's and 3's. *None* of the 1's was successful in completing pilot training.

The chart in the lower half of figure 4.8 represents a similar study including only those cases with no previous flying experience (no pilot credit) who graduated from preflight and entered elementary flying schools and also excluding from consideration men who were eliminated for any reason other than flying deficiency or fear of flying. This chart also indicates the marked success of the pilot stanine in predicting which men would graduate from flying training.

In figure 4.9 are presented some charts showing the predictive value of the printed tests which have substantial weights in determining the pilot stanine. The two best tests by quite a large margin were found to be the General Information Test and Instrument Comprehension Test II. These two printed tests were also found to be superior to any of the apparatus tests in predictive value. Both tests represent novel ideas developed within the Aviation Psychology Program.

The Mechanical Principles Test and Spatial Orientation Test II were also found to have substantial predictive value. The Biographical Data Blank (Pilot), and Spatial Orientation Test I were found to be of more limited value. The findings regarding the Mechanical Principles Test and the Biographical Data Blank (Pilot) are of special interest because these tests are quite similar to tests which the United States Navy and the pilot committee of the National Research Council had found to be of value early in the war. These tests constituted the principal tests of the United States Navy in its pilot-selection program throughout the war.

The Spatial Orientation Tests were developed by the Psychological Division in the Office of the Air Surgeon very early in the war and have continued in use ever since with very little modification. These tests involve the use of aerial photographs and sectional maps and were developed to measure perceptual aptitudes in the general area of alertness and observation which preliminary analysis indicated were important for success in pilot training. The second part of the test which involves the finding of areas shown by aerial photographs on a larger area portrayed by a sectional map was found to have more validity than the similar problem in which areas shown by aerial photographs were to be located in larger areas also shown as aerial photographs.

In figure 4.10 are shown the pilot validities of a number of tests which were developed primarily for the prediction of navigator and bombardier-training success. All of these tests have been found to have substantial validity for predicting success in navigator training. The Dial and Table Reading Test gives a better prediction of success in preflight school than any of the other tests in the Air-crew Classification Test Battery.

Instrument Comprehension Test I, which is similar in some ways to the Dial and Table Reading Test, was found on the basis of approximately 1,500 ca o have a substantially lower predictive value for success in primary flying ining schools than Instrument Comprehension Test II. Because of its high correlation with this latter test, statistical analyses indicated that it could be profitably used to suppress certain extraneous factors present in Instrument Comprehension Test II and thus improve the predictive value of the pilot stanine. Unfortunately, in subsequent samples the correlation between the two tests was found to be smaller than had been originally obtained. Also, the validity of test I was found to be somewhat larger for *primary training* than previously found, and even close to the validity of test II for *basic* and *advanced training*. For *preflight training* test I was superior to test II in its predictive value. Thus its early promise as a suppression test was not fulfilled and it was later dropped from the battery, since other tests, primarily Instrument Comprehension Test II and the Dial and Table Reading Test, appeared to provide adequate coverage of the functions measured by this test.

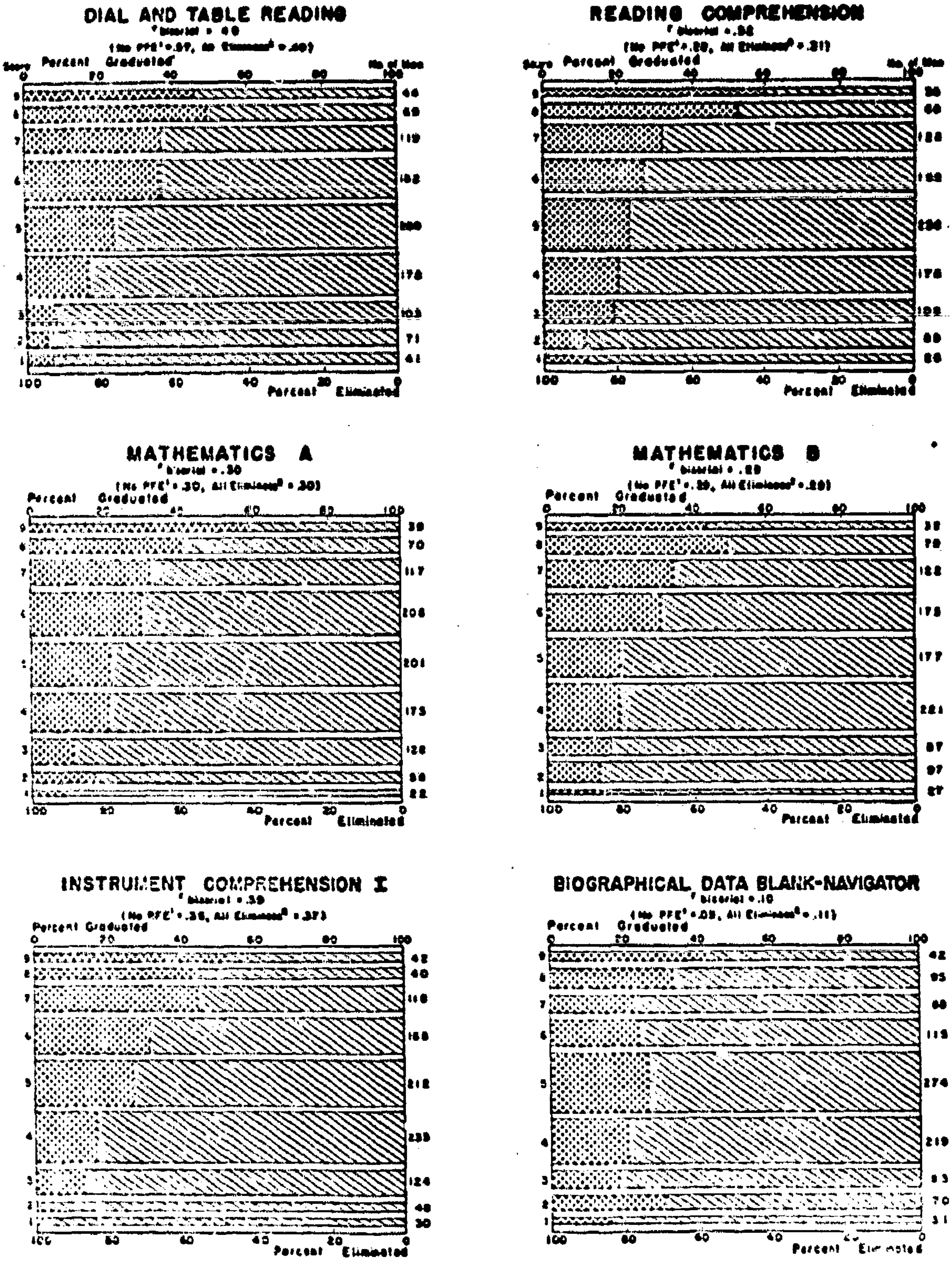
The mathematics tests, the test of reading comprehension, and the navigator key for the Biographical Data Blank were especially useful as classification tests because of their only moderate validity for predicting success in pilot training and their substantial predictive value for navigation-training success.

The predictive value of the apparatus tests used in the Air-crew Classification Test Battery at the time the experimental group was tested are shown in figure 4.11. It is seen that the Discrimination Reaction Time Test, the Rudder Control Test, and the Complex Coordination Test all have substantial predictive value for pilot training. The Two-Hand Coordination Test had somewhat less predictive value and the Rotary Pursuit Test was of limited value for this sample. The Finger Dexterity Test was not weighted for predicting success in pilot training.

The Rudder Control Test had the greatest predictive value for success in primary training schools and for predicting flying elimination when cases with previous flying experience were included. The Discrimination Reaction Time Test and the Complex Coordination Test were superior in predictive value to the Rudder Control Test for predicting basic training and preflight training. The Discrimination Reaction Time Test was especially good for predicting success in preflight school.

For comparison, the predictive value of certain other variables is shown in figure 4.12. It is seen that there is a very marked relationship between previous flying experience and success in pilot training. Education shows a very much smaller relationship. The General Classification Test has some predictive value in this unselected group but would not add to the over-all accuracy of predictions of the air-crew classification tests. In this sample, age and marital status have practically no relationship to success in flying training.

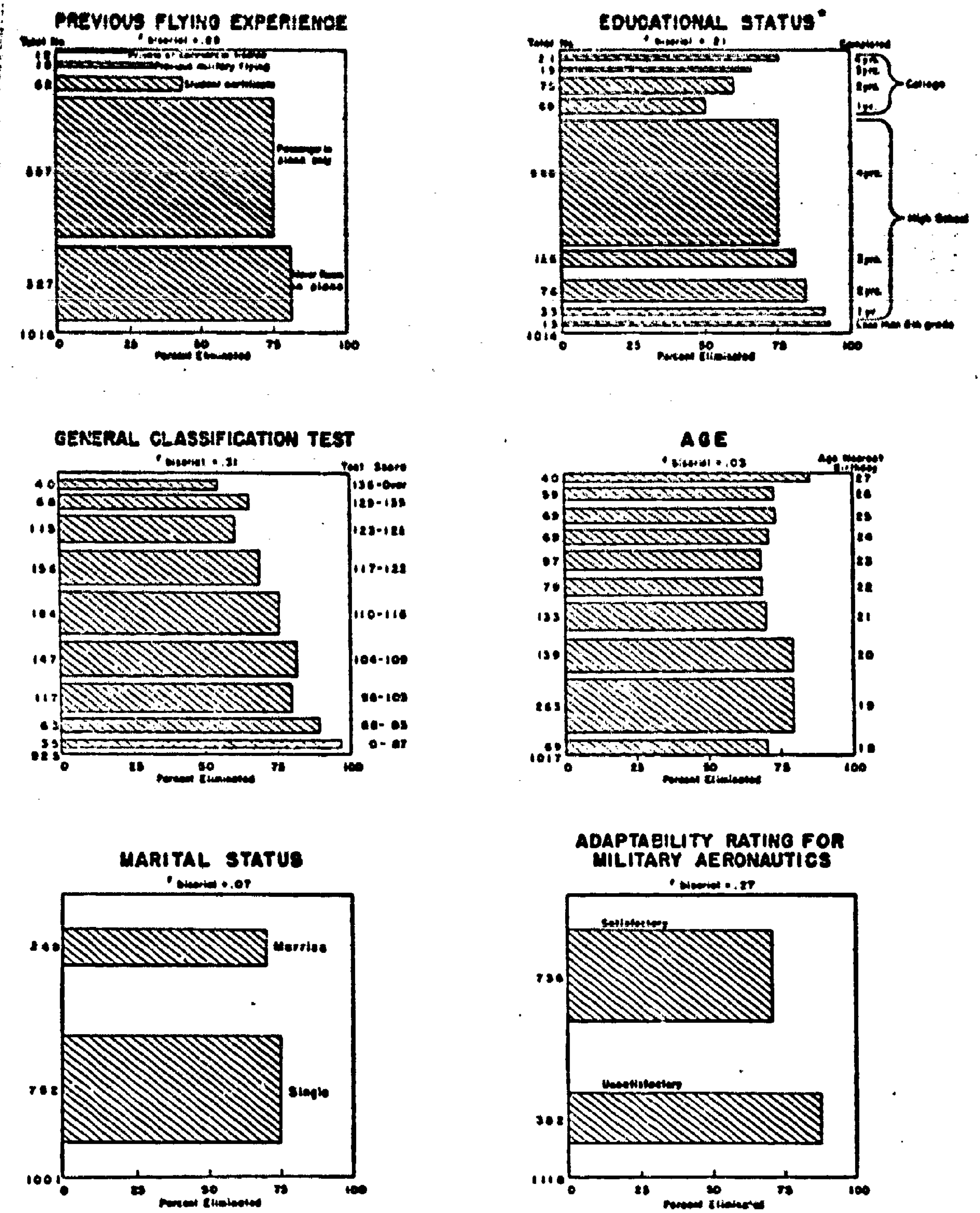
The Adaptability Rating for Military Aeronautics appears to have some predictive value for pilot training. An intensive analysis of the interview




[1] [illegible] coefficient for remaining cases when cases with previous flying experience are excluded.
[2] [illegible] all cases, including physical and administrative eliminees.

Figure 31.—[illegible] value of apparatus tests for success in pilot training. Data based on records of [illegible] Elimination was for flying deficiency, fear, and own request from preflight through advanced pilot training.

[illegible]—47 (Face p. [illegible])




* Two individuals had completed more than 4 years of college. One of these was eliminated.

FIGURE [illegible].—Predictive value of six variables for success in pilot training. Data based on records of experimental group. Elimination was for flying deficiency, fear, and own request from preflight through advanced pilot training.

sheets used by 10 examiners at the San Antonio Aviation Cadet Center suggested that the principal contributors were education, vocational achievement, interest in flying, national origin, and family income. There is a slight indication that the men who were rated as relaxed and listless during the interview were more successful in flying training than those who were rated as eager or tense. Neither the extent of hand tremor nor flushing were found to have predictive value for success in pilot training.

A number of statistical studies were carried out to evaluate the effectiveness of the Air-crew Test Battery and the pilot stanine in predicting success in pilot training. A table containing the product-moment intercorrelations of all of the variables in the Air-crew Classification Test Battery was prepared in order that certain analytical studies of combinations of tests and weights for specific tests could be studied in a precise fashion. This table of intercorrelations is reproduced in appendix B as table B.1. A number of analyses were made using the intercorrelations in this table, and the biserial correlation coefficients obtained between the test scores and the stanine and success in pilot training.

In calculating these coefficients, men eliminated for physical and administrative reasons were excluded from consideration. The two categories consisted of 262 men who graduated from advanced training and 755 who were eliminated in preflight, primary, basic, or advanced schools because of academic failure, flying deficiency, or fear of flying. The results of these analyses are reproduced below in table 4.2.

TABLE 4.2.—*The predictive value of various combinations of tests for success in pilot training as determined from an experimental group of 1,017 men.*

| *Combinations of predictions used* | *Correlation coefficient with pilot-training graduates—eliminees (academic and flying deficiency and fear of flying).* |
|---|---|
| Pilot Stanine | 0.460 |
| Best-weighted combination of air crew classification tests for this sample | .490 |
| Best-weighted combination of printed tests in air crew classification battery for this sample | .441 |
| Best-weighted combination of apparatus tests in air crew classification battery for this sample | .378 |

Using this set of validity coefficients and intercorrelations, the "best weights" give a prediction of success in pilot training which is characterized by a correlation coefficient 0.03 higher than that obtained from the particular set of weights used in computing the pilot stanine at the time these men entered training. It is known that correlation coefficients obtained in this way tend to show some shrinkage in a new sample, even though the sample is relatively large as in this case. Hence, we can conclude that the weights in use at that time were fairly close to the optimal ones.

As indicated in the table, it is also possible to predict success in pilot training with printed tests alone with an accuracy only moderately diminished, a correlation coefficient 0.05 smaller, than with the complete battery.

703323—47—7

Using the apparatus tests alone, the corresponding reduction in the coefficient is 0.11.

A type of problem frequently encountered in selection research is the question of the effect of selection on the basis of one variable on the predictive value found for a second set of scores. To make an empirical check on this, biserial correlation coefficients were computed for various tests excluding all of those individuals who would have normally been rejected on the basis of the AAF Qualifying Examination score. The correlation coefficients obtained for this group of 540 men were compared with those obtained for the uncurtailed group of 1,036 in predicting success in preflight and primary training schools. It was found that the average of the coefficients for these tests was approximately 0.05 lower in the restricted group. The validity of the pilot stanine was also 0.05 lower in this curtailed group.

A special study was made of the aircraft accident records of this group. Of the total group of about a thousand men, 20 had aircraft accident in training planes in the AAF Training Command. There were five accidents that involved pilots with pilot stanines of 7, 8, or 9. These higher stanine groups produced approximately a hundred of the graduates from pilot training. The lower stanine groups, which produced 150 graduates, had a total of 15 accidents.

Four of the accidents were fatal and these all involved individuals in the lower stanine groups. For the four men involved in fatal accidents, the stanines for bombardier, navigator, and pilot training were, in that order, 324, 636, 445, and 996. The first three were all violating flying regulations at the time of the accidents. The fourth individual overshot his turn from base-leg to final approach in lining up with the runway. In trying to bring the plane back, he stalled out and went into a half-snap. The instructor then took over but the plane hit on the left wing and cartwheeled.

## Detailed Individual Follow-Ups[10]

In a selection and classification program involving the testing and follow-up of hundreds of thousands of men, it is easy to lose sight of the individual man. Because of the special nature of the experimental group and the extensive amount of individual data already collected concerning these men, it was believed desirable to make an intensive study of certain individuals. It was believed that most could be learned by studying the cases for which the predictions were not fulfilled. Accordingly, a group of 31 men including 15 men with pilot stanines of 8 or 9 who were eliminated from training and 16 men with pilot stanines of 2 or 3 who graduated from training were made the subjects of a special individual follow-up conducted by an aviation psychologist from the AAF Training Command. Complete case studies were prepared for each of the 31 individuals. The sources of information included (a) psychological records of test scores, interests, and ARMA ratings, (b) preflight-school records including grades, demerits received, and highest rank held,

[10] The detailed individual follow-up study reported in this section was conducted by William E. Walton.

(c) sick call and hospital records, (d) training records including continuation or transition courses, (e) trainees' 201 file, (f) a personal interview with the trainee, and (g) a personal interview with certain of the student's supervisors. This study was begun on 15 June and concluded on 22 September 1945. The following statement is quoted from the summary submitted by the investigator along with the detailed case studies:

> The high-stanine men failed because they were weakly motivated, lacked emotional control, received poor instruction, had personality clashes with their instructors, had formed previous flying habits which interfered with their learning, or had personal problems which preoccupied their minds * * *.
>
> The low-stanine men learned to fly because of strong motivation, emotional maturity, good instruction, self discipline, and favorable personalities * * *.
>
> It is admitted that the conclusions are subjective in nature. It is believed that they are logical, however, and based upon adequate data. The evaluation of many official records, including final statements and board proceedings and the comparison of reports on the same man from a number of fields or from a number of instructors, leads the writer to conclude that less reliance can be placed upon those than upon the statements of the trainees themselves. It is believed that reasonable caution was exercised in their interpretation.
>
> Factors which seemed to have little or no bearing upon the success or failure of the trainees were health, leadership, temporary ratings (cadet ranks) ground-school grades, and data obtained from the personal histories.

It was hoped that this type of case-study material might bring out some rather specific ideas concerning tests or procedures which could be added to the Air-crew Classification Test Battery to improve the accuracy of prediction. Thus far, the analysis of these case materials has not been productive of such results. No clear-cut hypotheses for predicting these particular cases seemed to emerge. In many cases it appears that the instructors and check pilots were at fault and in other instances personal matters interfered with the normal progress of the instructional process.

Since individuals failing examinations frequently claimed that the cause of their poor test scores was illness at the time of taking the test, a special investigation was made of this matter for those individuals whose later success indicated that their test scores may have been too low. However, only one individual indicated that he had been ill at the time of examining. Since he is now rated as a rather poor-quality pilot it is probable that even in this instance his abscessed tooth may have had only a negligible effect on his test scores.

All of the objective data concerning these two groups of individuals were carefully examined. No specific patterns for test scores were discernible which might aid in prediction. The amount of education and height and weight were quite similar for the two groups. However, in these groups rather striking differences were observed in age and marital status. Among the 15 men with pilot stanines of 8 and 9 who were eliminated, 12 were 23 years of age or over, while in the group of sixteen graduates who had low pilot stanines, only 2 were more than 22 years of age. Similarly, eight of the eliminated group were or had been married whereas only two of the graduates had ever been married. Although these findings were suggestive, a check of the stanine 4's who graduated and the stanine 7's who failed did not confirm the general importance of these factors. Of the group graduating in spite of low stanines all indicated extremely strong interest in pilot training. On the other hand, in the high-stanine eliminees, two indicated greater interest in navigation training.

Although the investigator in his analyses of the reasons why the high group failed and the low group succeeded stressed emotional maturity and motivation and the lack of them as the principal reasons for the success or failure of these groups, the special 30-minute interview on which the Adaptability Rating for Military Aeronautics was based and which was designed to reveal these matters failed entirely to differentiate between the two groups. Only one of the high-stanine group was given an unsatisfactory rating on the basis of the interview whereas four of the low-stanine group who graduated were given unsatisfactory ratings.

### Implications

This study of 1,000 applicants and their success in pilot training in relation to their scores on the selection and classification tests has clearly demonstrated the effectiveness of these procedures when applied to groups of men recruited from civilian life or from the Army. Of 405 men who failed on the AAF Qualifying Examination and were subsequently sent into pilot training only 12 achieved pilot stanines of 7, 8, or 9 and only 4 of these and 42 others of this group of men who failed the Qualifying Examination were graduated from pilot training.

The value of the second screening by the Air-crew Classification Tests was dramatically demonstrated by the graduation of only 16 men out of 442 with pilot stanines of 1, 2, and 3 sent into preflight training. At the same time 113 men graduated of the 199 with pilot stanines of 7, 8, and 9 sent into preflight training.

The correlation coefficient of 0.66 obtained between pilot stanine and success in pilot training compares favorably with the best predictions which have been obtained in educational and industrial work. It now appears that improvement of instructional techniques and procedures for passing and failing students need to be made before a substantial amount of further refinement in the selection and classification procedures can be expected.

## THE STUDY OF THE UNITED STATES MILITARY ACADEMY CADETS IN PILOT TRAINING

### Preliminary Studies and Plans

One of the large peacetime sources of men for pilot training is the United States Military Academy at West Point. During the war period a special aviation training program was introduced for cadets at the Academy. As part of the accelerated program for training at the Military Academy, those cadets who had elected to take flying training and who had been found physically qualified were sent to civilian contract schools in the AAF Training Command where they received primary pilot training from April to June of the year preceding the year of their graduation from the Academy. Basic training was received during July and August at Stewart Field at West Point. This was followed by a period from September to March during which they flew just enough to maintain proficiency. Advanced training was received from March to May of their final year so that they graduated as rated pilots. The increase in the number of cadets eliminated for flying deficiency in the spring of 1944 as compared with the preceding years was of some concern to authorities of the Military Academy. Not only did such eliminations result in a wastage of cadet time, which was especially important during the accelerated program, but they also produced the problem of fitting eliminated cadets back into the work of their class.

A study was made of the comparative success in primary flying training schools of the United States Military Academy cadets and aviation cadets for the classes trained in the spring of 1942, 1943, and 1944. The findings of

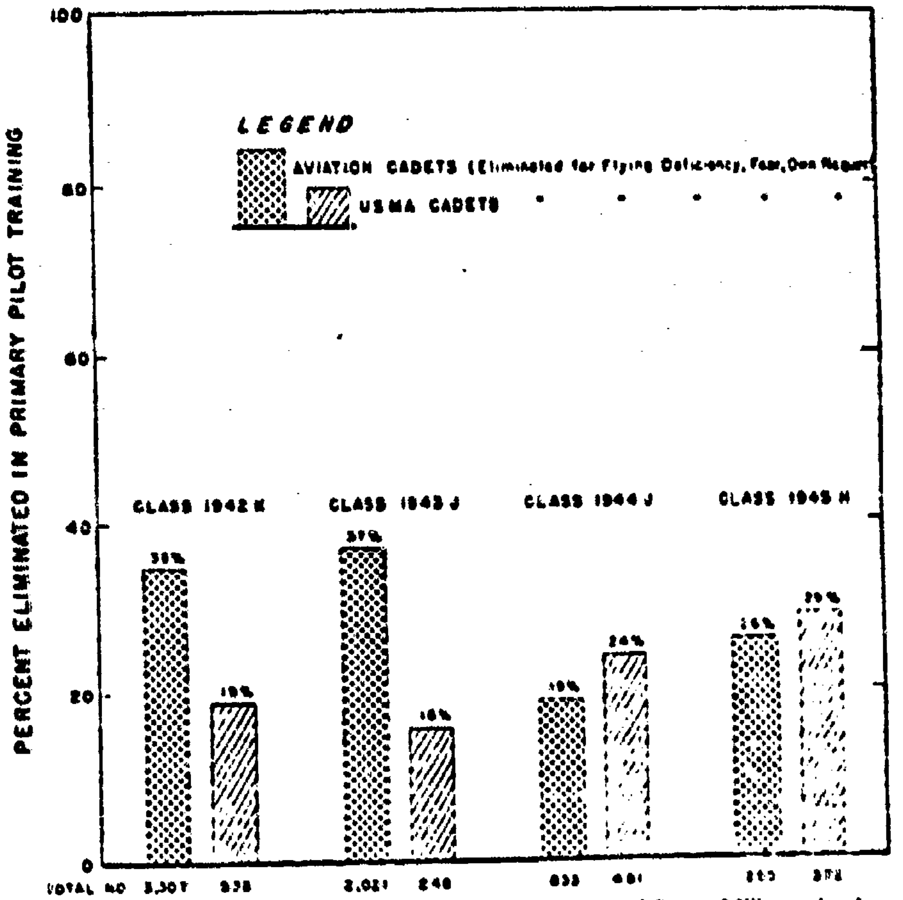



FIGURE 4.13.—Elimination rates for aviation cadets and United States Military Academy in the same classes and in the same primary pilot training schools.

this study are presented in figure 4.13 on which similar results for the class trained in the spring of 1945 have been added. Since West Point cadets, although a select group themselves, are assigned to pilot training without having taken the Air-crew Classification Tests they provided a fairly stable base line with which to compare changes in the procedures used in selecting and classifying cadets.

The 1942 aviation cadets shown on the chart were selected on the old 2-year college requirements in effect prior to January 1942 and very few of them had taken any selection or classification tests. It is seen that nearly twice as many of them were eliminated for flying deficiency as were West Point Cadets.

The 1943 aviation cadets had been selected on the basis of the AAF Qualifying Examination and there was a small amount of additional selection made on the basis of the classification tests. In this group, more than twice as many aviation cadets were eliminated as for the West Point Cadets.

The 1944 aviation cadets were selected not only on the basis of the AAF Qualifying Examination but also all of the low-stanine individuals were disqualified. It is seen that in this group a somewhat smaller proportion of the aviation cadets were eliminated than of the West Point Cadets.

The 1945 aviation cadets were also selected on the basis of the AAF Qualifying Examination and the requirements for qualification in terms of pilot stanines had been raised so that an even greater proportion of the aviation cadets with low pilot stanines had been rejected. In this class the difference is also in favor of the aviation cadets.

The AAF officers in charge of pilot training at Stewart Field were familiar with the demonstrated effectiveness of the Air-crew Classification Tests used in qualifying aviation cadets for training. These officers suggested to the United States Military Academy authorities and to the aviation psychologists in the Office of the Surgeon, Headquarters AAF Training Command that these tests be tried out on a class at the United States Military Academy to determine their effectiveness for prediction within a group of this type.

Accordingly, arrangements were made during the summer of 1944 to administer the AAF Air-crew Classification Tests to the class of 1946 at the United States Military Academy. The testing was carried out during the fall of 1944[11].

The printed tests were administered during the 4 days at the end of August between the time that cadets returned from summer camps and the time that academic work started. These were administered to groups of approximately 225 cadets at a time. The apparatus tests were administered to the cadets in groups of four individuals according to the standard procedures used in testing aviation cadets. These tests were taken by the cadets in November 1944 at the same time that they were taking the physical ex-

[11] The testing and general supervision of the collection and analysis of the data and the preparation of a report on this project were under the immediate supervision of Robert L. Thorndike of Headquarters AAF Training Command. Most of the tabulations and computations were carried on under the supervision of Walter L. Deemer, George Simon, Leonard Berwick, and Armen Alchian.

amination for flying. All tests were administered by trained examiners from the AAF Training Command. Practically the only changes made in the tests and testing procedures used in the AAF Air-crew Classification Program were a few necessary adaptations or eliminations of items on the Biographical Data Blank which were not appropriate for the Military Academy group.

## Study of Air-Crew Classification Test Scores

In figure 4.14 are shown graphically the comparisons between the scores of the United States Military Academy Cadets on the tests of the Air-Crew Classification Test Battery and those of aviation cadets. The mean score of aviation cadets tested at AAF Classification Centers is represented as the middle line labelled "5." The other lines indicate the mid-points of the respective stanine values. The distribution of West Point cadets is shown by the solid bars which include the middle two-thirds of the scores extending from one standard deviation below to one standard deviation above the mean. The horizontal bars indicate the mean.

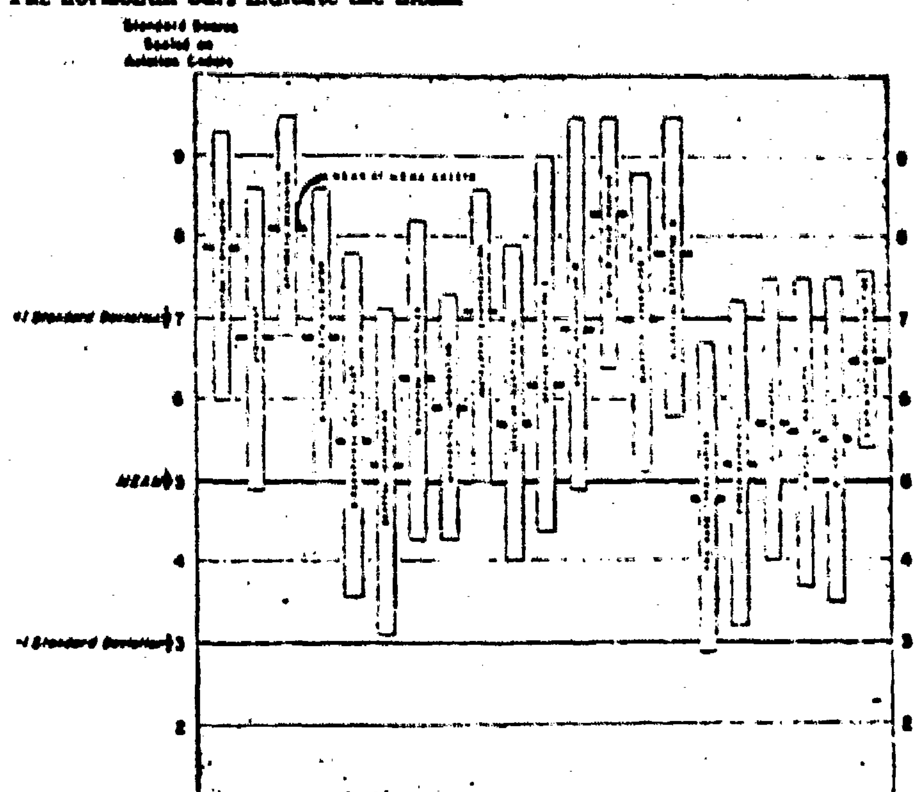



FIGURE 4.14.—The scores of United States Military Academy cadets on tests of the AAF Air-Crew Classification Battery in relation to those of aviation cadets.
NOTE—The bars extend one standard deviation above and below the mean.

It is apparent that the Military Academy cadets are markedly superior to the men tested at the AAF Classification Centers in such things as Arithmetic Reasoning, Numerical Operations, Reading Comprehension, and Dial and Table Reading. They are also quite a bit better than the air-crew candidates on such tests as Spatial Orientation I and II, Judgment, Instrument Comprehension, Mechanical Principles, and Discrimination Reaction Time. On the remaining apparatus tests including Rotary Pursuit, Complex Coordination, Rudder Control, Finger Dexterity, and Two-Hand Coordination, the test scores are not greatly different from those of the men tested in the AAF Centers. This is also true of tests of Speed of Identification, Mechanical Information, Biographical Data Blank (Pilot), and General Information.

Figure 4.15 shows a similar chart for the four stanines for fighter pilot, bomber pilot, bombardier, and navigator training. As could be expected from the preceding chart showing the results of the various tests, the superiority of the United States Military Academy cadets is most outstanding on the navigator stanine. They are least different from the men tested at the AAF Classification Centers in terms of fighter-pilot stanines. At the time that this testing was carried out, a stanine of 7 or better was required for an aviation cadet to qualify for any of these types of air-crew training. It is apparent that approximately 90 percent of the Military Academy cadets would qualify for navigator training, 76 percent for bombardier, 66 percent for bomber pilots, and only 46 percent for fighter-pilot training. The Military Academy cadets were clearly superior on most tests to the group selected on the AAF Quali-

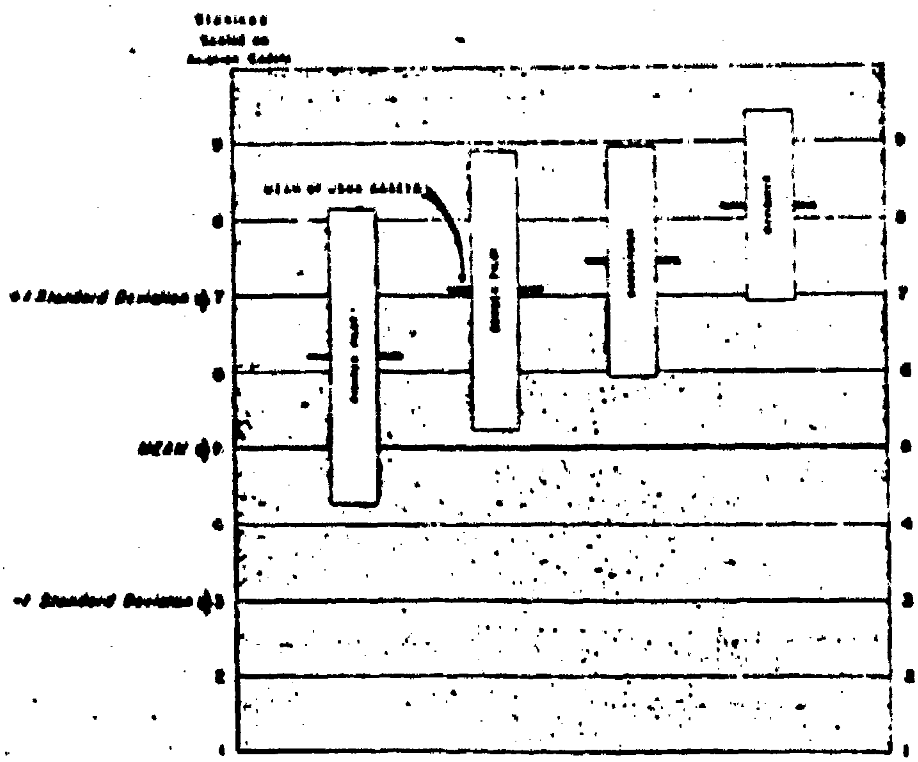



FIGURE 4.15.—The stanines of United States Military Academy cadets in relation to those of aviation cadets.

NOTE—The bars extend one standard deviation above and below the mean.

fying Examination and sent to AAF Classification Centers as candidates for air-crew training. However, their mean stanine was lower than that of the group being qualified and assigned to specialized training schools of these types except in the case of navigation training.

In evaluating the nature of the Air-Crew Classification Tests it is useful to analyze their relation to such established criteria as Military Academy grades. A complete table of all of the intercorrelations of the variables included in the West Point study is included as table B.2 in appendix B.

The Arithmetic Reasoning Test, Reading Comprehension Test, and Dial and Table Reading Test all have rather substantial predictive value for determining average West Point grades in the preceding year. The correlation of the "best-weighted" combination of these three scores with the average grades during the preceding year in terms of total proportional parts was 0.64. This is a fairly satisfactory degree of predictive value for a battery of three short tests requiring a total administration time of only 90 minutes.

The subject most accurately predicted by the tests was Military Topography and Graphics. A number of the tests were related to grades in this course and a combination of Arithmetic Reasoning, Dial and Table Reading, Spatial Orientation II, Reading Comprehension, Instrument Comprehension, and Mechanical Principles gave multiple correlations of 0.72 and 0.79 for the grades earned in the fourth and third year classes respectively. Grades in physics, chemistry, mathematics, and tactics were also predicted quite well by the Air-Crew Classification Tests. The only tests in the battery showing much relation to the English and history grades were the tests of reading comprehension and arithmetic reasoning. None of the air-crew classification tests showed a substantial correlation with language grades. This is not surprising in view of the nature of the technical specialties which the air-crew classification tests were designed to predict.

To try to shed some light on the nature of the motor abilities tested by the apparatus tests, scores on these and the other air-crew classification tests were correlated with the measures of physical proficiency obtained at the time of entrance to the Military Academy. These included a 300-yard run, dodge run, standing broad jump, vertical jump, bar vault, rope climb, sit ups, chins, dips, and soft-ball throw. The correlation coefficients indicated that there was practically no relation between these physical measures and either the apparatus tests or the printed tests in the Air-Crew Classification Test Battery.

## Analysis of Leadership Ratings

Another analysis was carried out to see whether the air-crew classification tests might predict United States Military Academy Aptitude Ratings. Ratings were available for analysis which had been made by other cadets, tactical officers, and academic officers. Each average rating represented the composite of a number of separate ratings made by individual officers or cadets at different times during the year. During a full year about 75 ratings were submitted on each cadet. Ratings were made on the three qualities of leadership, attitude, and military bearing. These were weighted in the proportions of 50, 35, and 15.

It is interesting to observe that the tests predicted the academic officers' rating with a moderate degree of accuracy. The academic officers' ratings are, however, substantially correlated with the academic grades given by these officers to the men. The tactical officers' ratings bear only a very small relation to any of the Air-Crew Classification Test scores. It should be noted, however, that all of these correlation coefficients are positive. The correlation between these test scores and the cadet ratings are even lower and a number of very small negative coefficients were obtained.

The cadet ratings show a very small positive correlation with academic grades and a moderately substantial correlation with military physical-efficiency grades which are essentially physical-education grades. These physical-education grades are substantially related to the physical measures such as the 300-yard run previously mentioned and these latter also show a small correlation with the cadets' ratings of military aptitude.

Thus, it appears that the only known component of cadet ratings of military leadership and aptitude is athletic prowess. The academic officers' ratings contained a substantial component of academic achievement and the tactical officers' ratings appeared to involve a small amount of each of these two components. It is clear that the problem of military leadership is in need of further investigation.

### Prediction of Success in Pilot Training

In table 4.3 on the following page are presented the biserial correlation coefficients for the stanines and the tests in the Air-Crew Classification Battery with success in primary flying training. It is apparent that the fighter-pilot stanine and bomber-pilot stanine both predict the success of these United States Military Academy cadets in primary flying training with approximately the same accuracy that these stanines show with other candidates for pilot ratings. This is shown graphically in figure 4.16.

The validities for the various Air-Crew Classification Tests obtained from the West Point group are very similar to the values ordinarily found among groups of aviation cadets. The tests with highest validity are in order: Rudder Control, Instrument Comprehension, Rotary Pursuit, Mechanical Principles, Complex Coordination, and Two-Hand Coordination. Although this group of a little more than 350 cases is not sufficiently large to establish genuine differences, it is of interest to note that the Rotary Pursuit Test has a somewhat higher validity coefficient than would have been expected on the basis of previous data. Similarly, the Spatial Orientation II and General Information Tests have somewhat lower validity coefficients than might have been expected from previous data. The Military Academy Class of 1947 has also been tested now and it is hoped that analysis of this class will provide further evidence as to whether these differences in predictive value are due to the nature of the West Point group or are merely sampling fluctuations.

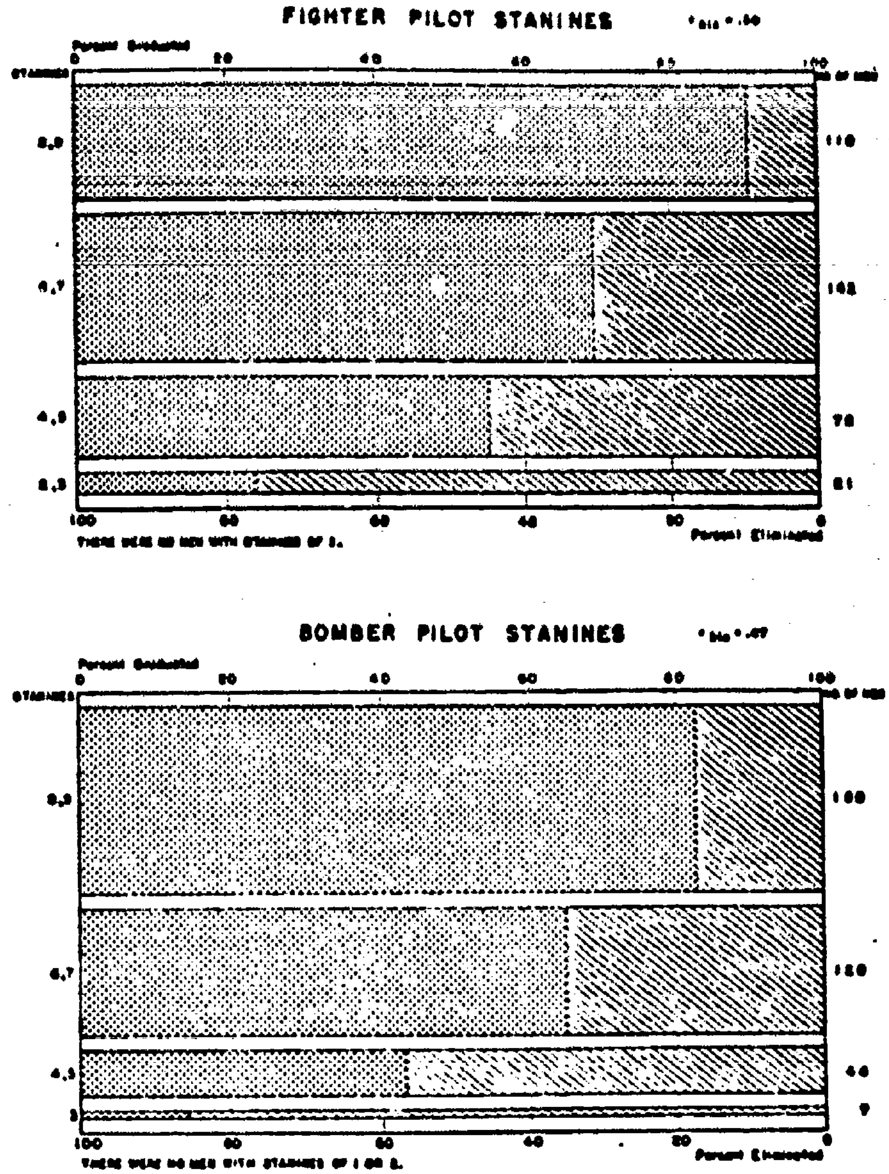



FIGURE 4.16.—The predictive value of fighter-pilot and bomber-pilot stanines for United States Military Academy cadets of the Class of 1946, tested in the fall of 1944 and trained in primary flying schools in the Spring of 1945. [Total number = 351]

It is also of interest to note the relation between the mean test scores of the United States Military Academy cadets who elected to take flying training as compared with their entire West Point class. The average fighter-pilot and bomber-pilot stanines of the group which entered flying training are approximately 0.1 standard deviation larger than the corresponding values for the class as a whole. These small differences are due almost entirely to the average scores for those entering flying training being approximately 0.2 standard deviation higher on the General Information Test and 0.3 standard deviation higher on the Biographical Data Blank (Pilot) than for the class as a whole. There are small differences favoring those going into flying training for all of the apparatus tests, but none of these is as large as 0.1 standard deviation. One further point of interest is the fact that the group entering pilot training is approximately 0.1 standard deviation lower than the total class on the tests of arithmetic reasoning and numerical operations. These differences in the scores of the men who have chosen flying training tend to confirm previous findings regarding the nature of interest in flying.

TABLE 4.3.—*The value of the stanines and individual tests in the Air-Crew Classification Battery for predicting success in primary pilot training for United States Military Academy cadets, Class of 1946, tested in the fall of 1944 and trained in primary flying schools in the spring of 1945.*

[Subscripts g = graduates, e = eliminees, and t = combined graduates and eliminees.]

| Test | No. of test | $N_t$ | $P_g$ | $M_g$ | $M_e$ | $SD_t$ | $r_{bis}$ |
|---|---|---|---|---|---|---|---|
| Arithmetic Reasoning | CI706C | 368 | 0.709 | 4.93 | 4.78 | 1.87 | 0.05 |
| Dial and Table Reading | CP622-21A | 368 | .709 | 5.19 | 4.51 | 1.82 | .12 |
| Spatial Orientation I | CP501B | 367 | .711 | 5.19 | 4.82 | 1.87 | .12 |
| Spatial Orientation II | CP503B | 367 | .711 | 5.15 | 4.82 | 1.93 | .10 |
| Biographical Data Navigator | CE602D | 367 | .711 | 5.19 | 4.97 | 1.98 | .07 |
| Biographical Data Pilot | CE602D | 367 | .711 | 5.91 | 5.02 | 1.82 | .19 |
| Numerical Operations Front | CI702B | 367 | .711 | 4.93 | 4.90 | 1.91 | .01 |
| Numerical Operations Back | CI702B | 367 | .711 | 4.85 | 4.57 | 1.90 | .09 |
| Reading Comprehension | CI614H | 367 | .711 | 5.10 | 4.76 | 1.95 | .11 |
| Judgment | CI301C | 367 | .711 | 5.38 | 5.02 | 1.91 | .11 |
| General Information | CE505F | 367 | .711 | 5.56 | 4.78 | 1.95 | .24 |
| Instrument Comprehension | CI616C | 368 | .709 | 5.38 | 4.24 | 1.89 | .36 |
| Mechanical Principles | CI903B | 367 | .711 | 5.41 | 4.39 | 1.92 | .35 |
| Mechanical Information | CI905B | 368 | .709 | 5.20 | 4.62 | 1.91 | .18 |
| Speed of Identification | CP610A | 367 | .711 | 5.05 | 4.86 | 1.88 | .06 |
| Rotary Pursuit | CP410B | 361 | .712 | 5.48 | 4.34 | 1.92 | .36 |
| Two Hand Coordination | CM101A | 361 | .712 | 5.34 | 4.34 | 1.87 | .32 |
| Complex Coordination | CM701A | 361 | .712 | 5.40 | 4.36 | 1.87 | .34 |
| Rudder Control | CM120B | 356 | .711 | 5.51 | 4.19 | 2.01 | .39 |
| Discrimination Reaction Time | CP611D | 361 | .712 | 5.37 | 4.68 | 1.91 | .22 |
| Finger Dexterity | CM116A | 361 | .712 | 5.34 | 4.74 | 1.89 | .19 |
| Bombardier Stanine | | 356 | .711 | 7.76 | 6.98 | 1.38 | .34 |
| Navigator Stanine | | 356 | .711 | 8.31 | 7.90 | 1.09 | .32 |
| Bomber-Pilot Stanine | | 356 | .711 | 7.70 | 6.46 | 1.59 | .47 |
| Fighter-Pilot Stanine | | 356 | .711 | 6.89 | 5.38 | 1.80 | .50 |
| Officer Quality | | 354 | .712 | 5.19 | 4.61 | 1.88 | .19 |

In table 4.4 are presented the validity coefficients for primary flying training of the various types of data available from the West Point records. Although these coefficients are uniformly quite low and would be entirely inadequate as a substitute for the air-crew classification tests in predicting success in flying training, there are a few points deserving special mention. For example, it is seen that the military physical-efficiency grade and the military topography-and-graphics grade provide the best predictions available from the West Point records. Grades in history, foreign language, and English have very close to zero relationship with success in flying training. Probably for the same reason as the validity of the military physical-efficiency grade, scores on the dodge run, broad jump, and bar vault are all found to have small but appreciable correlations with success in flying training.

## Implications

This confirmation of the predictive value of the air-crew classification tests is of special importance because it indicates that these tests discriminate within a group which has been carefully selected for officer training and from whom the inferior individuals have been eliminated during the first 2 years of the United States Military Academy course. Discrimination within this type of a group is much more important during peacetime than during the war period. While the war was going on, the demand was for large numbers of pilots, immediately, who had as much skill and ability as possible in flying planes and carrying out the other duties of a military pilot.

In selecting officers for permanent appointments in the Army Air Forces many other matters need to be considered besides skill in flying an airplane. Men must be selected who can fly well during the 5- or 10-year period immediately following their training. It is even more important that they be good group commanders and staff officers 15 or 20 years later. It is of most importance that they contain a large number of men qualified to assume the responsibilities of general officers 20 or 25 years after their

TABLE 4.4.—*The value of USMA academic grades, physical performance tests, and military aptitude ratings for predicting success in primary pilot training for United States Military Academy cadets, Class of 1946, trained in primary flying schools in the spring of 1945*

[Subscripts: g = graduates, e = eliminees, and t = combined graduates and eliminees.]

| *USMA military aptitude ratings* | $N_t$ | $P_g$ | $M_g$ | $M_e$ | $SD_t$ | $r_{bis}$ |
|---|---|---|---|---|---|---|
| First year: Officer Average Rating[1] | 368 | 0.709 | 5.02 | 4.18 | 1.94 | 0.20 |
| Cadet Average Rating | 368 | .709 | 5.20 | 4.76 | 1.97 | .10 |
| Final Composite Rating | 368 | .709 | 5.10 | 4.44 | 1.93 | .21 |
| Second year: Tactical Officer Average Rating | 368 | .709 | 5.20 | 4.77 | 1.78 | .15 |
| Academic Officer Average Rating | 368 | .709 | 5.05 | 4.49 | 1.86 | .18 |
| Cadet Average Rating | 368 | .709 | 5.15 | 4.72 | 1.90 | .13 |
| Final Composite Rating | 368 | .709 | 5.17 | 4.65 | 1.91 | .16 |
| *USMA academic grades* | | | | | | |
| First year: Language | 368 | .709 | 4.72 | 4.68 | 1.80 | .01 |
| Mathematics | 368 | .709 | 4.98 | 4.29 | 1.83 | .23 |
| English | 368 | .709 | 4.93 | 4.58 | 1.90 | .11 |
| Military Topography and Graphics | 368 | .709 | 5.18 | 4.34 | 1.92 | .26 |
| Tactics | 368 | .709 | 5.09 | 4.36 | 1.93 | .23 |
| Military Physical Efficiency | 368 | .709 | 5.54 | 4.65 | 1.96 | .28 |
| Total Proportional Parts | 368 | .709 | 4.97 | 4.29 | 1.80 | .23 |
| Second year: Language | 368 | .709 | 4.80 | 4.52 | 1.77 | .09 |
| Mathematics | 368 | .709 | 5.00 | 4.37 | 1.76 | .21 |
| Physics | 368 | .709 | 5.05 | 4.40 | 1.82 | .21 |
| Military Topography and Graphics | 368 | .709 | 5.18 | 4.36 | 1.89 | .26 |
| Tactics | 368 | .709 | 4.88 | 4.67 | 1.96 | .06 |
| History | 368 | .709 | 4.67 | 4.75 | 1.87 | −.03 |
| Chemistry | 368 | .709 | 4.83 | 4.31 | 1.89 | .17 |
| *Physical performance tests* | | | | | | |
| 300 yard run | 343 | .711 | 5.23 | 5.04 | 1.83 | .06 |
| Dodge run | 344 | .709 | 5.37 | 4.76 | 1.90 | .19 |
| Broad jump | 343 | .711 | 5.30 | 4.67 | 2.05 | .19 |
| Vertical jump | 344 | .712 | 5.24 | 4.99 | 2.04 | .07 |
| Bar vault | 344 | .712 | 5.37 | 4.74 | 1.86 | .20 |
| Rope climb | 344 | .709 | 4.68 | 5.04 | 1.81 | −.11 |
| Sit ups | 344 | .712 | 5.05 | 5.10 | 1.70 | −.02 |
| Chins | 343 | .714 | 5.28 | 4.87 | 1.79 | .14 |
| Dips | 344 | .712 | 5.30 | 5.17 | 1.77 | .04 |
| Soft-ball throw | 343 | .711 | 5.23 | 5.09 | 1.89 | .04 |

[1] These refer to grades in the first year of study at the United States Military Academy.

original flying training was completed. Fortunately, some of the skills required by the two types of duties are similar. However, many of the aptitudes and abilities necessary for a good commander or a good staff officer are quite different from those required of the beginning pilot.

In selecting officers for the peacetime air forces it is very important that the critical requirements for duties during all parts of the officers' anticipated career be defined and weighted in accordance with their relative importance to the effective functioning of the Air Forces. Only on such a basis as this can the Air Forces be assured that it will have suitable personnel in the future.

## STUDY OF WASP TRAINEES[18]

In evaluating the Air-Crew Classification Tests, it is of interest to present the performance of women on these tests. Early in 1944, arrangements were made with the Director of Women Pilots Office, Headquarters Army Air Forces, to administer various tests of aptitude and personality to women taking pilot training in connection with the Women's Airforce Service Pilots training program. During the last week in March the complete Air-Crew Classification Test Battery was administered to all of the WASP's taking training at Avenger Field. This included classes 44-W-3 through 44-W-8. The classes were at different stages in their training, 44-W-3 being almost through advanced training and 44-W-8 having just started primary training 14 March. Each of the entering classes was composed of about 100 students, only about 50 of whom graduated from advanced pilot training. All students were required to have at least 35 hours of previous flying training as a prerequisite for admission to training.

Based on studies of the 480 women tested in these classes, the following findings are of importance. The women receiving pilot training had more formal education than aviation cadets entering pilot training at approximately the same time. Sixty-five percent had received some college training. The group was superior to corresponding groups of aviation cadets in tests of reading comprehension, mathematics, and other academic abilities. The average academic intelligence in this group of women was higher than that in a typical group of men undergoing pilot training in the AAF at the same time. The average scores of the women were lower than those of the aviation cadets on certain tests which were specifically developed for predicting pilot aptitude. Their averages were definitely lower than those for the men on tests of understanding mechanical principles, and of two-hand coordination.

The predictive value of the pilot stanine as obtained from the Classification Test Battery put into effect 1 November 1943 was approximately the same for these classes of women as for typical aviation-cadet classes.

[18] The field work and part of the analysis for this study were carried out by Merrill F. Roff. The final analyses were performed by the staff of the Psychological Section, Headquarters AAF Training Command.

This finding was based upon follow-up and analysis of the results for classes 44-W-6, 7, and 8. In the first two of these classes there were about 20 percent of the women eliminated after classification tests were given and in the last class about 40 percent were eliminated after testing. Students eliminated before the examination dates were not available for testing.

Although the small size of the sample and the large proportion of students who had received a substantial amount of flying training does not permit an accurate statement regarding the relative predictive value of the classification tests for women as compared with those for aviation cadets, it should be noted that the validities follow a similar pattern to that found in the experimental group study. The validity coefficients for the pilot stanine are approximately 0.55. The General Information Test has an even higher validity for this group than was found in the experimental group, the value being 0.56 for 195 cases. The validity coefficients for the Complex Coordination Test, Mechanical Principles Test, and the Rotary Pursuit Test seem to be somewhat lower for this sample than are usually obtained.

In conclusion, although it appears that some revision of the Biographical Data Blank, Mechanical Principles Test, and General Information Test would be desirable if the Air-Crew Classification Test Battery were to be used with women, it was found that the battery without modification provided a satisfactory basis for selecting women for pilot training. Beginning with the class which entered flying training in April of 1944, women entering WASP training were required to meet the same standards on the AAF Qualifying Examination as men who were applicants for aviation cadet training. In June 1944, plans were made to administer the air-crew classification tests to all women entering pilot training in this program. On the basis of the results which had previously been obtained it was believed that there should be no difficulty in obtaining a sufficient number of women who could meet the standards for men. This program was never carried out because of the decision in June 1944 to discontinue recruiting and training women for the WASP program.

## STUDIES IN OPERATIONAL TRAINING AND IN COMBAT THEATERS[18]

### Fighter Pilots

Although the early studies of the effectiveness of selection and classification procedures were carried on in connection with individual training

[18] The studies reported in this section were carried on in the AAF Training Command by: N. E. Miller, S. M. Roshal, W. H. Angoff, J. G. Gleason, C. P. Gershenson; in the Continental Air Forces by: M. P. Crawford, L. B. Ward, R. W. Heyns, F. H. Mitchell, R. T. Sollenberger, R. H. Henneman, H. T. Hadley, W. Webb, D. H. Jenkins, T. W. Smith, R. H. Turner, H. J. Hausman, S. H. Marsh, A. Kolton, L. B. Ward, M. H. Brown, F. Killian; in the combat theaters by the editor of this report and in AERD No. 1 by A. P. Horst, B. Gilmer, C. L. Heathers, L. Berwick, S. W. Cook, W. G. Mollenkopf, L. G. Carpenter, J. S. Harding, J. K. Hemphill, W. M. Wheeler, G. J. Wischner, H. O. Preston, R. R. Borreseur, F. C. Davis, G. H. Shirley, R. B. Miller, R. T. Mitchell, E. Erskine, Jr., R. Glaser, F. N. Knett, J. W. Nakovits; in AERD No. 2 by N. D. Warren, J. R. Brick, W. F. Long, W. W. Grings, J. R. Harsh, C. H. Richardson, A. J. Rinpelle, R. C. Anderson, A. A. Cantfield, G. L. Thomas, P. M. Clancy, H. W. Kirchner; in AERD No. 3 by W. M. Lepley, B. B. Harless, A. H. Brooks, S. M. Roshal, F. M. Moriarty, M. Abern, W. T. Brown, R. E. Johnston, Jr., H. L. Hay, S. A. Peprnice, S. B. Lyerly, R. Merrill, Jr.; and in connection with a special study of fighter pilots in the Pacific theater by W. B. Webb.

in the AAF Training Command schools the orientation of the Aviation Psychology Program was toward the improvement of the effectiveness of the aircrew personnel supplied to combat units. It was desired that this improvement be with respect to those qualities which were most critical in determining the success or failure of these individuals in combat operations.

### *Test Scores*

One of the first studies made of the success of a specific type of aircrew personnel in performing a combat-type job was the study of fixed gunnery performance in transition stations of the AAF Training Command. The early studies of this type were done by Aviation Psychologists from Psychological Research Unit No. 1. This work was continued by Psychological Research Project (Pilot). In table 4.5 are given the average product-moment correlations for all of these studies combined.

It is not surprising to find the Instrument Comprehension Test, the Mechanical Principles Test, and the Complex Coordination Tests at the top of the list. It is unfortunate that the sample is small for the Instrument Comprehension Test due to its relatively late development and addi-

TABLE 4.5.—*Relative value of the Air-Crew Classification Tests for predicting percent hits in air-to-air and air-to-ground firing in fixed-gunnery training among samples of pilots in AAF training command stations*

| Tests | Test No. | Number of individuals in combination | Combined[1] weighted $r$ |
|---|---|---|---|
| Instrument Comprehension | CI616B | 681 | 0.160 |
| Mechanical Principles | CI903A | 1,862 | .158 |
| Complex Coordination | CM701A | 2,385 | .130 |
| Spatial Orientation II | CP503B | 2,385 | .123 |
| Two Hand Coordination | CM101A | 1,662 | .116 |
| Rotary Pursuit | CM803A (CP410B)[2] | 1,862 | .106 |
| Biographical Data Blank—P | CE602D | 1,000 | .100 |
| Spatial Orientation I | CP501B | 2,385 | .100 |
| Discrimination Reaction Time | CP611D (CP611A) | 2,385 | .099 |
| Dial and Table Reading | CP621-22A | 1,862 | .097 |
| Finger Dexterity | CM116A | 2,385 | .084 |
| Rudder Control | CM120B | 1,000 | .070 |
| Mathematics A | CI710A (CI702F) | 1,862 | .070 |
| Reading Comprehension | CI614G (CI614H) | 1,662 | .061 |
| Speed of Identification | CP610A | 1,181 | .060 |
| General Information | CE505E (CE505C, CE505D) | 2,385 | .046 |
| Steadiness Under Pressure | CE206B | 523 | .040 |
| Mathematics B | CI206C (CI706A, CI206B) | 1,662 | .039 |
| Numerical Operations | CI702B | 1,181 | .028 |

[1] Classes included in combination were 43F, 43J, 43K, and 44-1-I through 45-A, Eastern Flying Training Command.

[2] The combined correlations for these tests include a composite of all of the forms indicated. The test number or numbers in parenthesis constitute the smaller proportion of cases.

tion to the Air-Crew Classification Tests. Most of the other values are relatively stable due to the substantial sizes of the samples involved. It is of interest that the correlations for all of the tests are positive. The test of simple arithmetic computations, the Numerical Operations Test, is at the bottom of the list with a correlation which is insignificantly different from zero.

Supplementing these data regarding the relative value of the air-crew classification tests for predicting success in fixed gunnery are studies made by Aviation Psychologists assigned to the Second and especially the Fourth

Air Force. These studies also include correlations with accidents, with flying grades, and with graduation or elimination from fighter pilot training. The data from all studies have been combined and are presented in table 4.6.[14] These samples include approximately 1,800 P-38 pilots and 300 P-61 pilots from the Fourth Air Forces and 350 P-47 pilots from the Second Air Force. When more than one type of data is available for a given group, these types are averaged, but the number of individuals reported is the number of different individuals involved in obtaining the correlations which have been combined, not the sums of the numbers of scores averaged for all individuals.

For all criteria of success combined it is seen that the Rotary Pursuit and Mechanical Principles Tests provide the highest correlations. It is found that the Numerical Operations Test is again at the bottom of the list. This time it has an insignificant negative correlation. All of the remaining values for the combined criteria are positive. The highest correlations obtained with the criterion of percent hits in fixed gunnery and in bombing are for the Mechanical Principles, Complex Coordination, Two-Hand Coordination, Rotary Pursuit, and Discrimination Reaction Time Tests in that order. This is precisely the same order in which these tests are found in the AAF Training Command studies reported in table 4.5. The Instrument Comprehension Test was added too late to have been taken by these Continental Air Force groups. The relatively high coefficients found for the two Spatial Orientation Tests and the Biographical Data Blank (Pilot) were not confirmed by the Continental Air Force Data. Additional samples of fighter pilots are needed to resolve these conflicting findings.

In addition to the tests already mentioned, the tests of Mechanical Information, Reading Comprehension, Speed of Identification, and Mathematics A and B give correlations relatively further from zero for the combined operational training criteria than do the other tests on the list. The first of these, Mechanical Information, was not taken by the pilots studied in the AAF Training Command groups. In general, in these samples, the other tests mentioned give relatively high correlations with the criterion of elimination from fighter-pilot training and relatively lower correlations with the criterion of percent hits both in the AAF Training Command and in the Continental Air Forces studies.

A number of studies were made of data collected in combat theaters regarding fighter pilots. Unfortunately none of the types of records available appeared to be very satisfactory as a criterion of success as a fighter pilot in combat operations. Nevertheless, the correlations of air-crew

[14] As a preliminary step to providing the composite correlations presented in tables 4.7, 4.8, 4.11, 4.13 and 4.15, the data that had been originally reported in terms of critical ratios had to be converted to biserial correlations. When the standard deviation of the total group was not available the standard deviation of the control group was usually used as the best available estimate for it. In all studies involving reevaluation, Flying Evaluation Board cases, Central Medical Establishment cases, etc., the proportion of reevaluation cases was assumed to be 5 percent. In studies involving accidents and casualties, the proportions involved were assumed to be 15 percent and 25 percent respectively. It was assumed that equal proportions of fighter pilots did and did not have at least one air combat victory. The same proportions were assumed for decorated versus nondecorated fighter pilots. In the lead versus non-lead studies, the proportion of lead personnel was assumed to be 25 percent.

TABLE 4.6.—*The relative value of the aircrew classification tests in samples of fighter pilots for predicting percent hits in fixed gunnery and in bombing, accidents, flying grades and graduation or elimination from fighter pilot training because of flying deficiency or unsuitable temperament in operational training*

| Tests | Test No. | Number of individuals | | | | | Combined weighted r | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Hits[1] | Accidents[2] | Flying grades[3] | Graduation Elimination[4] | Total | Hits | Accidents | Flying grades | Graduation Elimination | Total |
| Rotary Pursuit | CM803A | 1,224 | 277 | 415 | 1,471 | 2,020 | 0.063 | 0.140 | 0.180 | 0.126 | 0.112 |
| Mechanical Principles | CI901A | 2,054 | 1,900 | 420 | 1,565 | 2,237 | .070 | .051 | .110 | .165 | .123 |
| Mechanical Information | CI905A | 1,019 | 993 | 84 | 841 | 1,106 | .043 | .035 | —.030 | .092 | .077 |
| Two-Hand Coordination | CM101A | 2,107 | 1,936 | 420 | 1,494 | 2,293 | .065 | .050 | .100 | .098 | .076 |
| Discrimination Reaction Time | CP611D | 2,180 | 2,051 | 432 | 1,634 | 2,409 | .056 | .020 | .001 | .137 | .055 |
| Reading Comprehension | CI614G | 2,054 | 1,875 | 420 | 1,565 | 2,[illegible]5 | .025 | .016 | .114 | .177 | .060 |
| Speed of Identification | CP610A | 2,115 | 2,011 | 432 | 1,635 | 2,348 | .039 | .024 | —.029 | .110 | .055 |
| Mathematics A | CI702E,[5] (CI702F) | 2,179 | 1,986 | 432 | 911 | 2,427 | .050 | —.031 | .093 | .253 | .047 |
| Mathematics B | CI710A, (CI206C) | 2,126 | 1,941 | 432 | 911 | 2,374 | .033 | —.011 | .078 | .183 | .038 |
| Complex Coordination | CM701A | 2,179 | 2,049 | 432 | 1,633 | 2,407 | .065 | —.026 | .043 | .027 | .036 |
| Dial and Table Reading | CP622-21A | 2,126 | 2,010 | 431 | 1,615 | 2,358 | .036 | —.006 | .018 | .072 | .031 |
| Technical Vocabulary—P | CE505C | 1,155 | 1,150 | 96 | 911 | 1,279 | .079 | .034 | —.100 | —.067 | .031 |
| Biographical Data—P | CE602D | 932 | 755 | 314 | | 1,051 | .012 | .050 | .071 | | .030 |
| Technical Vocabulary—N | CE505C | 1,155 | 1,150 | 96 | 911 | 2,040 | .053 | .001 | .110 | .113 | .030 |
| General Information—P | CE505D | 1,015 | 837 | 336 | | 1,139 | .014 | .020 | .153 | | .030 |
| Spatial Orientation II | CP503B | 2,181 | 2,054 | 431 | 1,635 | 2,413 | .017 | .060 | .035 | .031 | .025 |
| Aiming Stress | CE211A | 2,053 | 1,897 | 420 | 1,573 | 2,231 | .032 | .026 | .060 | —.054 | .020 |
| Finger Dexterity | CM116A | 2,179 | 2,053 | 432 | 1,634 | 2,411 | .031 | .035 | —.046 | —.016 | .016 |
| Technical Vocabulary—B | CE505C | 1,155 | 1,150 | 96 | 911 | 2,040 | .015 | .057 | .030 | —.048 | .014 |
| General Information—N | CE505D | 1,015 | 837 | 336 | | 1,139 | .042 | —.040 | .060 | | .011 |
| Spatial Orientation I | CP501B | 2,181 | 2,033 | 432 | 1,635 | 2,412 | —.002 | —.012 | .019 | .061 | .008 |
| Biographical Data—N | CE602D | 932 | 755 | 314 | | 1,051 | —.002 | .020 | —.029 | | .006 |
| Numerical Operation II | CI702B | 2,068 | 2,054 | 432 | 1,635 | 2,300 | —.022 | —.014 | .028 | .115 | .001 |
| Numerical Operation I | CI702B | 2,068 | 2,054 | 431 | 1,615 | 2,100 | —.007 | —.067 | .079 | .036 | —.020 |

[1]Air-to-air, air-to-ground, skip-and dive-bombing scores for P-38 pilots, Fourth Air Force; air-to-air and air-to-ground gunnery scores for P-48 pilots, Fourth Air Force; air-to-air gunnery scores—for P-47 pilots, Second Air Force.
[2]Accidents ascribed to pilot error—P-38 pilots and P-61 pilots—Fourth Air Force.
[3]Mission grades and adversity scores—P-38 pilots, Fourth Air Force.
[4]Reevaluated for flying deficiency by Flying Evaluation Boards—P-38 pilots, Fourth Air Force.
[5]The combined correlations for these tests include a composite of all of the forms indicated. The test number or numbers in parenthesis constitute the smaller proportion of cases.

classification tests with these criterion were computed and the combined correlations are reported in table 4.7. It was previously seen that the correlations with operational-training criteria were generally lower than with fixed-gunnery scores in the AAF Training Command. This may be partially explained on the basis of greater homogeneity of experience in the latter group. The combined correlations for these criteria of combat drop still further — in fact, almost to the vanishing point. However, among the five tests at the top of the list are four which previous studies also found having relatively high validities. These are Discrimination Reaction Time, Rotary Pursuit, Mechanical Principles, and Complex Coordination. The remaining test of the first five is Aiming Stress.

TABLE 4.7.—*The relative value of the Air-Crew Classification Tests in samples of fighter pilots for predicting combat ratings of effectiveness, casualties, accidents, aerial combat victories, decorations, reclassifications, and transfers to other types of duty*

| Tests | Test No. | Data[1] combined | Approx. number of cases in combination | Combined weighted r |
|---|---|---|---|---|
| Aiming Stress | CM211A | 3,5,7,9,11-14 | 2,180 | 0.051 |
| Discrimination Reaction Time | CP611D | 1-14 | 2,700 | .050 |
| Rotary Pursuit | CM803A | 3,5,7,9,11-14 | 2,180 | .047 |
| Mechanical Principles | CI903A, (AC10D6)[a] | 1-14 | 2,700 | .041 |
| Complex Coordination | CM701A | 1-14 | 2,700 | .040 |
| Spatial Orientation I | CP501B | 1-14 | 2,700 | .035 |
| Finger Dexterity | CM116A | 1-14 | 2,700 | .027 |
| Spatial Orientation II | CP503B | 1-14 | 2,700 | .025 |
| Two-Hand Coordination | CM101A | 1,3,5,7,9,11-14 | 2,210 | .023 |
| Dial and Table Reading | CP621-22A, (CP621A, CP622A) | 1-11 | 1,900 | .022 |
| Technical Vocabulary (P) | CE505C | 1-14 | 2,700 | .014 |
| Speed of Identification | CP610A | 1-11 | 2,700 | .011 |
| Mechanical Information | CI905A | 3,5,7,9,11 | 1,380 | .011 |
| Reading Comprehension | CI614C, (AC10D-3) | 1-11 | 1,900 | .008 |
| Mathematics B | CI206B, (CI710A, CI706A) | 1-11 | 1,900 | .005 |
| Numerical Operations | CI702B | 1-11 | 1,900 | −.019 |
| Technical Vocabulary (N) | CE505C | 1-11 | 1,900 | −.020 |
| Mathematics A | CI702E | 1-11 | 1,900 | −.022 |
| Technical Vocabulary (B) | CE505C | 1-11 | 1,900 | −.037 |

[1]Studies included in combination:

| | No. of cases |
|---|---|
| *Initial survey* | |
| 1. Ratings by squadron officers on general over-all effectiveness ETO | 11– 37 |
| 2. Ratings by squadron officers on general over-all effectiveness MTO | 20–169 |
| *AERD No. 2* | |
| r biserials calculated from critical ratio data (See footnote 14). | 47– 65 |
| 3. Personnel removed from flying status vs. matched control | 120–120 |
| 4. Casualties vs. noncasualties, Battery No. 1 | 350–350 |
| 5. Casualties vs. noncasualties, Battery No. 2 | 85– 85 |
| 6. Accidents vs. nonaccidents, Battery No. 1 | 366–366 |
| 7. Accidents vs. nonaccidents, Battery No. 2 | 110–110 |
| 8. Victories vs. nonvictories, Battery No. 1 | 450–450 |
| 9. Victories vs. nonvictories, Battery No. 2 | 73– 73 |
| 10. Decorated vs. nondecorated, Battery No 1 | 158–158 |
| 11. Decorated vs. nondecorated, Battery No. 2 | |
| *AERD No. 3* | |
| 12. Proficiency rating made by Commanding Officer, Operations Officer and Flight Surgeon | 230–305 |
| 13. Hypothetical elimination | 203–263 |
| 14. Retention vs. transfer | 255–357 |

[a]The combined correlations for these tests include a composite of all of the forms indicated. The test number or numbers in parentheses constitute the smaller proportion of cases.

As might have been expected, scores on Numerical Operations showed a small negative correlation as did three other scores — Mathematics A scores and Technical Vocabulary, Navigator and Bombardier scores.

### *Pilot Stanines*

In addition to these studies of the various tests to provide information regarding the best-weighted combination of these scores to predict success as a fighter pilot, a number of studies were carried out to evaluate the effectiveness of the particular combinations of scores included in the pilot stanines which these men obtained at the time of their original classification testing. The results of these studies are summarized in table 4.8.

TABLE 4.8.—*The value of the pilot stanine for predicting various criteria of success as a fighter pilot*

| | Number of individuals in combination | Combined weighted r |
|---|---|---|
| Fixed Gunnery (AAF Training Command) | 2,383 | 0.194 |
| Air-to-Air Gunnery (operational training) | 3,457 | .131 |
| Air-to-Ground Gunnery (operational training) | 2,856 | .091 |
| Skip Bombing (operational training) | 576 | .101 |
| Dive Bombing (operational training) | 886 | .029 |
| Accidents (operational training) | 3,300 | .170 |
| Mission Grades and Comments (operational training) | 628 | .160 |
| Graduation-Elimination (operational training) | 2,798 | .127 |
| Ratings of Effectiveness, Casualties, Accidents, Aerial Combat, Victories, Decorations, Reclassifications, and Transfers to Other Types of Duties (combat) | 1,330 | .041 |

*These data include all of the validity studies reported on each of the criteria indicated, including critical ratios that were converted to biserial correlations utilizing the assumptions reported in footnote 14.

These results indicate a positive relation between pilot stanine and measures of fixed gunnery in the AAF Training Command; air-to-air

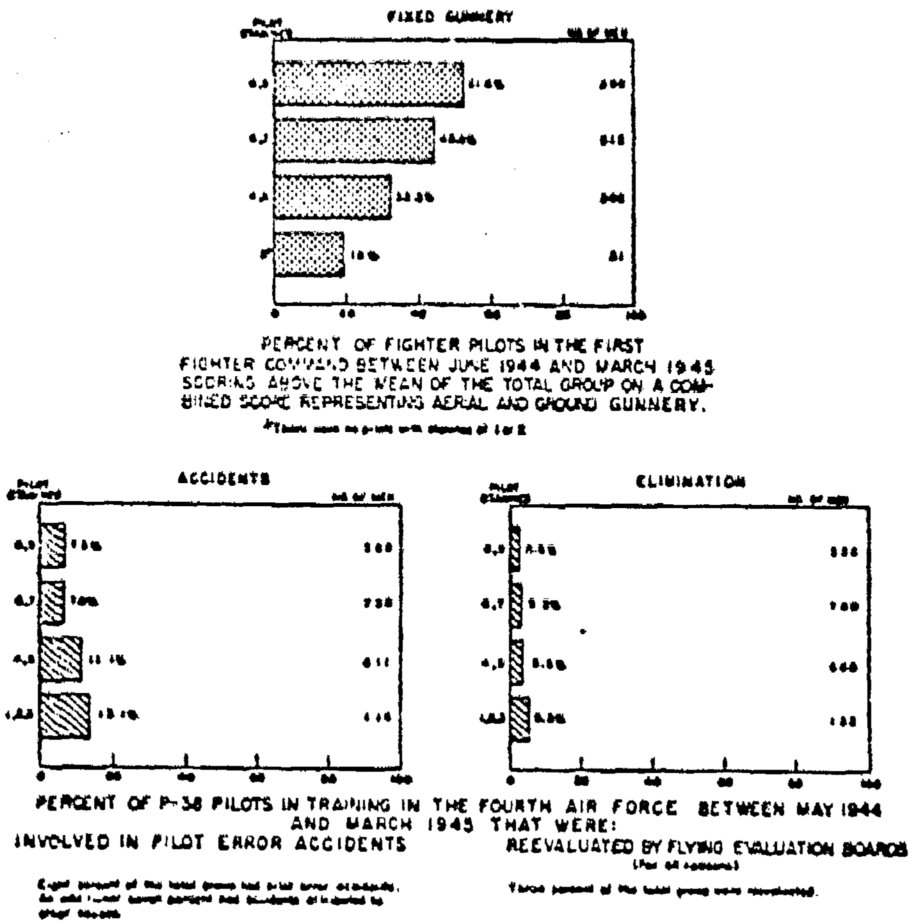



FIGURE 4.17.—Value of pilot stanine for predicting fighter pilot operational training criteria.

gunnery, air-to-ground gunnery, skip bombing, dive bombing, accidents, flying grades, and graduation or elimination in fighter-pilot operational training; and the composite of ratings of effectiveness, casualties, decorations, reclassifications, and transfers to other types of duties in combat. The coefficients, especially those for the latter types of criteria, are considerably attenuated by the action of various extraneous factors. Therefore, it appears that even these very small coefficients can be regarded as giving encouraging confirmation of the validity of the procedures used.

To provide a graphic picture of certain of these relationships, they are presented in figure 4.17. The superiority of the high pilot-stanine groups for fighter-pilot operational training is definitely indicated on these charts.

## Bombardment Crews

In the modern air force, the primary mission of air power is bombardment of enemy targets. Although single-place planes became increasingly effective during the war in their ability to carry appreciable bomb loads for moderate distances, bombardment operations typically involve a large plane manned by a crew composed of several individuals, each trained for a special duty assignment.

A number of studies were carried out to evaluate the effectiveness with which the individual specialists in these crews carried out their part of the bombing mission. It was discovered early that no one individual could be evaluated independently of the other members of the crew. The air crews functioned as teams and success in accomplishing their primary mission was dependent not only on each man's doing his particular job well, but also on the effective coordination of their combined efforts.

### *Bombardiers*

The natural criterion for evaluating the success of bombardment operations is the determination of where the bombs fall in relation to the assigned target. The individual member of the air crew having direct control over this was the bombardier. Accordingly, numerous studies were made comparing test scores and other known facts about the bombardier with bombing results. The bombing results were usually reported in terms of the straight-line distance from the center of the assigned target to the point of impact of the bombs. This was ordinarily referred to as the circular error.

As discussed in the following chapter, it was found that a bombardier's average circular error did not provide a dependable measure of his bombing skill. Many other factors such as the particular bombsight and plane used, the pilot, and the weather conditions were so important as to tend to obscure almost completely individual differences in bombing ability. However, if a criterion of this type is not systematically biased by such extraneous factors, the correlations between the predictive measures and the criterion are merely attenuated by them. In this situation it is merely necessary to increase the size of the sample in order to obtain stable relative correlations between the test scores or other measures and the criteria.

who had received the basic training course. This fundamental study, although based on a very small number of cases tends to cast doubt on the value of the ground trainers which were extensively used in gunnery training.

Studies of the interrelations among the various scores on these trainers also suggested that the trainers were very specific. It must be concluded that the learning of a skill having a superficial resemblance to that for which the individual is to be trained may contribute nothing or even produce habits of interference with respect to proficiency in handling combat-type equipment in combat situations.

## Studies of Types of Planes for Basic and Advanced Pilot Training

In connection with the 5-week "freeze" of pilot training the Psychological Research Project (Pilot) was able to compare the performance in advanced two-engine training, using the TB–25, of two groups of students who had received their basic flying training in different types of planes. A group of 84 students received all 15 weeks of their basic training in a single-engine plane, the AT–6. The other group of 165 students 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 in basic flying schools.

All of these students received their advanced two-engine training at the same school (Moody Field). These two groups of students received their basic flying training at different schools and were, as far as is known, random groups of students. Unfortunately, this part of the experiment was not designed. 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.

After the two groups 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. At the end of each week of advanced training each instructor was required to rank-order his students on 12 aspects of proficiency. These categories taken from the grade slips included taxiing, take-off and climb, stalls, slow flying, single-engine work, landings, formation, night landings, attitude toward work, technique, progress, and judgment. Although the rank-ordering by the instructors was subjective the assignment of equal proportions of each group to each instructor provided a control for one of the chief sources of difficulty in using subjective grades; i. e., the different standards of various instructors.

The comparisons of the two groups were done by means of the differences in the numbers of students in each group whom the instructors rated as being in the top half of their group of students on each of the 12 aspects of proficiency rated. Differences larger than could reasonably be expected as a result of chance were obtained in favor of the group receiving their basic training in the three different types of planes for five of the proficiencies rated. These aspects of proficiency were taxiing, take-off and climb, single-engine work, progress, and technique. By the end of the ninth week all of these differences had decreased and only taxiing showed a sufficiently large difference to rule out the possibility of a chance fluctuation in sampling. Two other large differences were found, one in slow flying and the other in night landings. All of the differences favored the group who had received their basic training on the three planes, one of these being a two-engine plane.

Although as previously indicated, this experiment was not adequately con-controlled to regard the results as conclusive, they strongly suggest that use of a variety of training planes at the basic stage is desirable and particularly that a two-engine plane at this stage is of definite assistance in advanced two-engine training.

A similar study in which all of the conditions were not controlled was made in connection with the use of various types of planes in advanced two-engine training. It was found that students with a total of 10 weeks of two-engine training, all of which was in the higher-powered TB–25, performed better on that airplane in terms of the objective measures used in the training-freeze experiment than students with a total of 15 weeks of two-engine training, 10 of which were in the lower powered AT–10 or UC–78.

### Implications

These studies clearly indicate the need for a large amount of additional research of this type. Carefully designed experiments should precede all major changes in the training course involving training devices or training equipment. If good control of the situation is possible these need not involve large numbers of cases nor great additional expense in order to insure positive results. The key to successful training experiments is scientific design of the experiment to control all of the extraneous factors.

## STUDIES OF METHODS OF TRAINING

### Studies on Aircraft Recognition Training[6]

The most systematic psychological research program on methods of training was carried out by the staff of the Psychological Test Film Unit on methods of teaching the identification of aircraft. This study grew out of differences of opinion regarding the effectiveness of various methods of training in aircraft recognition. In August 1943, following a conference between psychologists in Headquarters Army Air Forces and representatives of the training division in that Headquarters, a formal request was made that aviation psychologists conduct a series of research studies to determine the most effective methods of training in identification of aircraft. This problem was referred to the Psychological Test Film Unit and as a result a comprehensive research program including ten or more studies was carried out.

[6]The experimental research reported in this section was chiefly the work of R. M. Gagne and J. J. Gibson.

### *The Value of Rapid Flash Speeds*

Initially, the Psychological Test Film Unit was requested to obtain experimental evidence regarding several problems related to the flash system of instruction. Accordingly, an experiment was designed to evaluate the efficiency of rapid flash speeds. Certain individuals claimed that flash presentation produces superior learning because it forces the student to learn "total forms." It was argued that exposures as brief as one seventy-fifth of a second were too rapid to permit the moving of the eyes over the picture and therefore the learning would not be dependent on an analysis of features such as had characterized certain earlier methods of aircraft recognition but would depend on the perception of the total form.

Each of four instructors trained one group of students by giving training on all slides at an exposure speed of one second throughout the course. The same instructors trained a second group using one-fifth-second exposures for the first few hours and the moderate flash speed of one-tenth second for the remainder of the course. For the third group, students were started with one-fifth-second exposures and the speed gradually increased to one-fiftieth of a second during the first 2 weeks of the course. The remainder of the course was continued at exposure speeds of one-fiftieth of a second. Each of the 3 experimental groups consisted of about 170 students of aircraft recognition classes in preflight training schools. The results of this experiment are shown in table 6.7.

TABLE 6.7.—*Proficiency in aircraft recognition of three experimental groups as measured by the Aircraft Recognition Proficiency Test and slide examinations at different exposure speeds*

| | Slow trained $N = 173$ | | Intermediate trained $N = 167$ | | Fast trained $N = 177$ | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Motion Picture Test | 63.4 | 13.6 | 61.9 | 12.2 | 62.2 | 13.8 |
| Flash Slides Test (1 sec.) | 16.7 | 2.2 | 16.7 | 2.3 | 17.0 | 2.4 |
| Flash Slides Test (1/10 sec.) | 14.3 | 3.0 | 14.3 | 2.8 | 14.0 | 2.8 |
| Flash Slides Test (1/50 sec.) | 13.0 | 2.9 | 13.6 | 2.9 | 14.1 | 3.3 |

From this study it was concluded that there were no differences in recognition proficiency when students were trained with flash slides at exposures of 1 second, one-tenth second, and one-fiftieth second.

The evidence indicated that slides shown for 1 second are easier to see than those shown for either one-tenth second or one-fiftieth second in that fewer errors are made in the slide test of recognition given at 1-second speed regardless of the type of training of the individuals tested.

Individuals trained at one-fiftieth-second flash speeds are reliably more proficient at recognizing slides shown at this speed than individuals trained at 1-second exposures. However, this superiority does not hold for measure of proficiency at one-tenth-second or 1-second exposures nor for the criterion considered most realistic, that of the motion-picture test called the Aircraft Recognition Proficiency Test.

### *Comparison of Total Forms and Features*

A study was carried out to determine the relative importance of emphasizing the total forms or the features of the airplanes to be identified. Each of three instructors taught two classes by different methods. In both classes slides were presented at flash speeds of one-tenth second. In one group instruction was given on only the total form of each plane and no mention was made of features such as shapes of wings, nose, or tail. In the second group a standard set of distinctive features agreed upon by the instructors was emphasized in the case of each plane presented.

At the end of the 30-hour training period the group which had been encouraged to identify the planes by means of features was slightly superior on both the final slide examination and the motion-picture Aircraft Recognition Proficiency Test. Each group included about 90 students. Although differences between the average scores made by these groups on these tests are not large enough to be significant in the statistical sense, they suggest the ineffectiveness of an exclusive emphasis on total form.

### *The Value of Digit and Counting Slides*

An experiment was designed to test the value of supplementary training in reading digits and counting spots with flash presentation. The flash system of aircraft-recognition training used a series of slides containing 4 to 10 digits which had to be reproduced after a flash exposure and also slides showing from 3 to 30 planes which had to be counted during increasingly brief exposure intervals. This training was claimed by its proponent to increase the general efficiency of perception and to "widen the angle of vision" and to develop the "perception of numerousness."

The results from an experiment using 100 students trained in the use of the counting and digit slides with an equivalent untrained group indicated that training on these slides definitely improved proficiency with respect to the digit and counting slides. However, proficiency in aircraft recognition as measured both by a slide examination and by the Aircraft Recognition Proficiency Test was found to be no greater in the group given this special training than in an equivalent untrained group.

Furthermore, no significant difference in amount of improvement in score on a special perceptual test, Flexibility of Attention, was found between the group given digit and counting slide training and the untrained group. These findings tended to increase doubts concerning the statements that such training "improved the general proficiency of perception."

### *Remembered Shapes and Features of Aircraft*

Following the completion of these experiments which were designed to answer certain questions regarding the value of aspects of the flash system of aircraft recognition, experiments were formulated which could be expected to lead to positive rather than merely negative recommendations. The first of these consisted of a study of the remembered shapes of aircraft as revealed by drawing and by composites constructed from them. A group of 196 students who had just completed the 30-hour course in aircraft recognition in preflight school made free-hand drawings from memory of 8 airplanes selected from those they had learned. This group had received no practice in drawing during this course. Each student drew the three conventional views for each airplane. To standardize descriptive terminology regarding characteristics or features of the plane to be pointed out as possible cues or tips for use by the students in recognition, from six to eight phrases were agreed upon for each plane. These were selected as being significant or important characteristics of that plane. During the training these features were not emphasized but merely pointed out to the students. The sizes of the drawings were kept constant by marks which fixed the length and wing span but not the proportion or shape.

"Adequacy" scores were obtained for each student by giving him a point for each of the characterizing features which were shown on his drawings, taking all three views together. A rating of the draftsmanship exhibited by each of the sets of drawings was also made. The raters were to disregard the correctness of the shape in making this judgment. The students' final course grades based on their ability to identify aircraft from slide photographs were also obtained. For this group of students, the correlations with final grades were found to be 0.61 and 0.37 respectively for the "adequacy" scores and the draftsmanship ratings. The correlation between the draftsmanship ratings and the "adequacy" scores was found to be 0.54. The correlation between the "adequacy" scores and the final course grades became 0.75 when corrected in the usual way for the attenuating effects due to the brevity and lack of consistency of the samples of behavior on which these measures were based.

Composites were obtained by averaging the shapes of the students' drawings. In general, the composites as well as most of the original drawings had the main visual characteristics of the planes they were intended to represent. It may be inferred that differential visual images had been acquired by the students. An interesting point was that studies of the drawings indicated clearly that there were consistent differences between the drawings and the actual shapes they were intended to represent. A number of these differences were exaggerations of those characterizing features of the plane which had been pointed out to the students by the instructors.

This suggests that the verbalized characteristics have modified the images despite the contrary influence of the stimulus pattern and it is therefore reasonable to infer that these features played an important part in differentiating the various memory-images from one another and in giving them significance or meaning. The results of this experiment suggest that training in drawing airplanes from memory would be a useful supplementary training method in courses on aircraft recognition.

### *Effectiveness of Active Response and Reinforcement*

By the early part of 1944 directives from Headquarters Army Air Forces had decreased the emphasis on the flash method of training in AAF schools, particularly progressive speeding up of the exposure intervals. There was increasing emphasis upon training films, photographs, models, and "shadowgraph" presentation of models. Because it was thought that some of the values inherent in the previous procedures might be lost, an experiment was designed to demonstrate that well-established principles of learning operated in this situation and should be followed in planning the use of various training materials.

For this experiment approximately 280 students were divided into 2 groups and presented 20 unfamiliar slides of foreign planes. During each of the 3 presentations, each of the 20 slides was shown for 5 seconds, the name of each plane was announced just before it appeared on the screen, and was repeated while on the screen. The other half of the group was presented slides for the first time in exactly the same way. On the second and third presentations, however, the slides were exposed for 2½ seconds after which the trainees were required to identify each in writing on a numbered answer sheet. Guessing was encouraged. In case no identification could be made, a line had to be drawn through the appropriate answer space.

Immediately thereafter the same slides were exposed again for 2½ seconds and the name of the plane announced. Individuals who had made a correct response placed a check mark beside it. Those who had made either an incorrect response or none at all were required to write down the correct name in the appropriate blank. On the fourth presentation for both groups each slide was exposed for 2½ seconds and the trainees were instructed to write down the name of each plane on appropriately numbered answer sheets. The score on the test was simply the number of planes correctly identified.

For the first group the average score was 10.5 correct identifications. For the group in which active responses and corrections were required the average score was 14.6 correct identifications. This difference was much larger than could reasonably have occurred in samples of this size if no differentiating factor was present.

The results of the experiment were interpreted as a demonstration of the superior effectiveness of any classroom procedure which requires active identifying of responses and permits the confirming or correcting of these responses. The application of this well-established principle of learning to the aircraft-recognition-training situation suggests that mere exhibitions, displays, or motion pictures in which the learner remains passive and is not required to make an active response which is reinforced by confirmation or correction can not be expected to be as effective as a system based on differential reinforcement of responses.

### *A Study of Order of Presentation*

In the courses in aircraft recognition in the preflight schools it was customary to introduce two new planes each day. The procedure used was to introduce two dissimilar planes of the same nationality each day. The reason given for this was to prevent the type of confusion likely to occur if two very similar planes were presented together. Some instructors believed that although this avoided initial confusion, it resulted in greater confusion when planes similar to those already learned were introduced later. The instructors believed that the best way to prevent such confusion was to present

703323—47—13

similar planes in pairs regardless of nationality and to point out clearly those features of the aircraft by which they could be distinguished from one another.

An experiment was designed to test the relative effectivenss of these two procedures. Six classes in aircraft recognition containing about 30 men each were divided into 2 groups, matched for initial knowledge of airplanes on the basis of scores on a pretest of the 30 slides of American aircraft. Three instructors taught these classes, each instructor having one class in each group. Both groups of classes were taught the same 40 planes by the usual methods. The only difference between the two groups was the order of presentation. In one group two dissimilar planes of the same nationality were presented each day. In the second group two similar and confusable planes were presented during each class period without regard to their nationality. The training period of both of these groups was 26 hours.

At the end of this period, both groups of classes were tested by means of a slide examination composed of 45 slides showing the 40 planes presented during the course and in addition were tested with the Aircraft Recognition Proficiency Test (Preflight Level). The analysis of results indicated that on both of these examinations small differences favored the group in which similar planes had been taught together. The small size of the differences suggested that they may have been due to sampling fluctuation.

The results of this experiment suggest that both of these procedures have advantages. As indicated in a previous experiment learning is dependent on correct responses which are reinforced. The presentation of dissimilar planes initially aids in the correct identification on the next presentation which reinforces the response and contributes to learning. It is suggested, however, that after the responses have become fairly well established, the similar and confusable planes should be presented together with emphasis upon their differentiating characteristics.

It would probably be especially desirable to avoid confusion during the early stages of learning by keeping similar planes in separate categories. For example, the P-51 and Me 190 could be kept in discrete categories of American planes and German planes until the correct responses to these planes were fairly well established. Before the student was required to identify them from a mixed list which might contain either he could be given training and practice in differentiating the two similar planes.

***Distance as a Factor in Recognition Training***

During the earlier period of training in the recognition of aircraft, the slides used were for the most part photographic "closeups" of the various aircraft. Identifying backgrounds were eliminated and they eventually came to have no background at all except a clear film representing sky or clouds. The lack of realism in this situation as contrasted with the combat situation in which the aircraft ordinarily had to be identified at distances greater than a half a mile, created some interest in the problem of procedures for providing a natural impression of distance in the slides shown.

One of the first methods tried for creating the impression of greater distance was to move the projector closer to the screen. This was used extensively for some time but later was generally abandoned in the Army Air Forces schools since it was found that the picture was generally perceived not as more distant but merely as a smaller picture. Another method was to produce slides in which the size of the airplane was decreased in relation to the surrounding clear area representing sky or clouds within the picture. Still another method used was to increase the distance at which students viewed the screen or to mark out on the floor of the classroom seating distances which would correspond to the theoretically greater apparent ranges of the aircraft pictured on the screen.

To study the effect of these last two procedures an experiment was designed in which students were asked to estimate the range of three types of slides in one of which the airplane occupied most of the frame, a second type in which the planes averaged five-eighths of the size in the first slide, and a third series in which the plane averaged half of the size of those in the large slides. These slides were viewed at 3 different distances from the screen, 10 feet, 20 feet, and 40 feet by approximately 180 students awaiting entry into preflight schools. Sixty students viewed the slides from each of the distances and made estimates in yards regarding the distance from them that the plane appeared to be. The results are shown in table 6.8.

TABLE 6.8.—*Mean estimates of range in yards as a function of viewing distance and size in the picture*

| Viewing distance | L-slides | | M-slides | | S-slides | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| 10 feet | 206.2 | 213.4 | 315.8 | 224.2 | 413.8 | 280.8 |
| 20 feet | 201.6 | 196.3 | 356.5 | 254.4 | 466.3 | 274.7 |
| 40 feet | 234.5 | 214.3 | 407.2 | 265.7 | 556.0 | 276.9 |

Although the impression of distance produced by the pictures as shown by the estimates made was not a definite and clear one, the mean estimates of range rose in a consistent fashion as the size of the plane was reduced in comparison with the picture frame. On the basis of the sizes used, the distance would be expected to rise in the ratio of 1:1.6:2 if there were a perfect inverse relationship. From the table it is seen that the ratio actually found was 1:1.7:2.2. There is, therefore, an approximately inverse relationship in aircraft recognition slides between the size of the figure within a fixed picture frame and the impression of distance.

On the other hand, it is clear that apparent range did not increase in proportion as the viewing distance increased. If the range were determined by the visual angle principle and nothing else it would be expected to increase in the ratio of 1:2:4. The average ratios actually obtained for the three types of slides were in the ratio of 1:1.1:1.3. These results indicate a tendency for the estimated range to remain constant regardless of the distance of the observer from the screen. The constancy is definitely not perfect but it is at least clear that as the retinal size of a pictured airplane diminishes with viewing distance its apparent range does not increase proportionately.

A problem of even greater practical importance was that of determining the maximum range at which a given fighter or bomber plane can be positively identified in the air. It was impractical to obtain a carefully controlled experimental determination of recognition ranges for various types of planes in the air. Therefore, an experiment employing model planes and distances on the same scale as the models was designed. It was believed that the identifiability of a model seen against a background of sky at a considerable distance was at least a rough approximation to the identifiability of a real plane seen against the same sky at a proportionately greater distance.

In this experiment 307 air-crew trainees who had just completed a 30-hour course in aircraft recognition in the Preflight School (Pilot), Santa Ana Army Air Base, were asked to identify 6 model aircraft in each of 4 attitudes

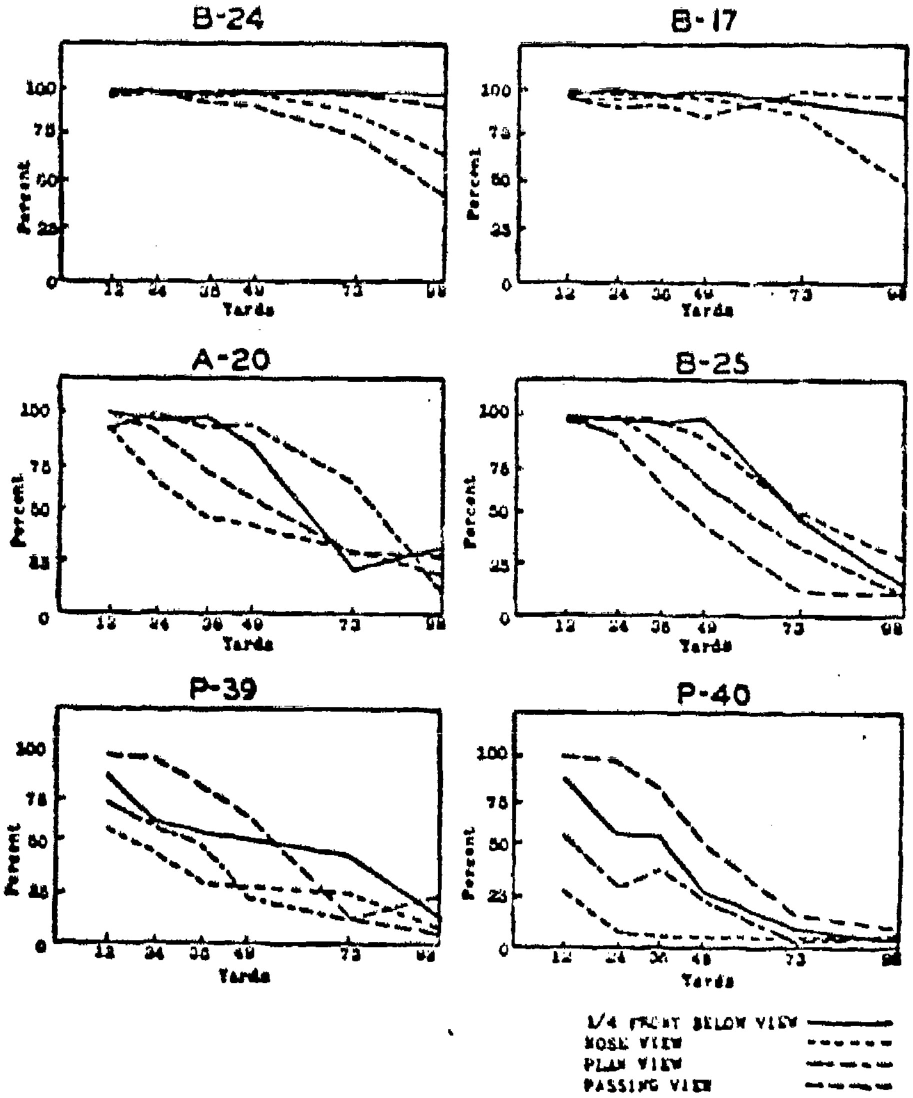



FIGURE 65.—Identifiability of aircraft models in four attitudes at various distances.

when seen against the sky at each of 6 distances varying from 12 to 98 yards. To rule out the factor of knowledge of airplane shapes, very well-known planes were used and observers who made any errors at all at the nearest distance were eliminated from the experiment. The cadets were divided into four groups, one group for each of the four attitudes in which the models were displayed. Individuals began attempting to identify the models at a distance of 98 yards and moved up to successively closer distances. At each distance the six models were displayed. The scale used for the models was 1 to 72, so the equivalent distances ranged from ½ mile to 4 miles. The results are shown in figure 6.5.

These results indicated that the recognizability of a particular aircraft depends to a very large extent upon the attitude in which it is seen. For these planes the most difficult of the four attitudes used was generally the head-on position. The quarter-front-below view was the easiest for the identification of most of these planes. The largest airplanes were found to be more frequently correctly identified at a given distance.

Because of such factors as atmospheric haze, glare, differences in relative brightness, and color contrasts, it seems likely that the recognition range of real aircraft under ordinary conditions is smaller than the ranges which would be estimated from the experiment with models. It will be noted that with the models only 11 percent of the group correctly identified the P-40 from the head-on view at the distance proportional to one mile and only 27 percent made a correct identification at the distance corresponding to half a mile.

### The Motion Picture as a Training Method*

During the war a very large number of motion-picture training films were produced. There were no accepted guiding principles regarding the subjects in which the motion picture could be expected to give better instruction than ordinary methods. Nor was there any body of knowledge regarding the instructional techniques peculiar to motion pictures which give them advantage over other methods of learning. A need was felt for information in this area, and accordingly the AAF Motion Picture Unit at Culver City and the AAF Training Division in New York City requested the Psychological Test Film Unit to conduct some experiments in this area.

An experiment was designed to compare the over-all effectiveness of the training film with that of an illustrated lecture (oral instruction with visual aids) and with the study of an illustrated manual (written instruction with visual aids). There was available for this purpose a training film entitled "Position Firing" which was produced entirely by animated photography in which the subject matter was presented with commentary and animated diagrams in a logical order. The material was, however, also organized around a thread of story providing characterization and humor ("Trigger Joe").

*This work was carried out by J. J. Gibson, R. M. Gagne, L. H. Boria, and C. H. Orvis.

On this same topic a 50-page pocket-size loose-leaf illustrated manual entitled "Get That Fighter," employing advanced visual methods with a minimum of text and a maximum of diagram was available, covering the same subject matter as the film. The diagrams were in color and the execution was extremely skillful. This manual was published by AAF Training Aids Division on 1 November 1943. It was studied by the trainees without discussion or explanation.

To parallel these presentations a half-hour lecture on the same subject matter was organized around a series of 19 lantern slides made from the booklet. The talk was written out for delivery in an informal spoken style, the diagrams on the screen being explained with a pointer. It was then revised, rewritten, and finally memorized for actual delivery. The talk included questioning of the class by the instructor to bring out salient points. As revised and approved it was judged by two experienced college teachers to be an example of excellent instruction.

The subject matter was analyzed into 14 basic points and 25 questions were finally collected for the final examination to test learning of these points. The odd-even reliability coefficient for this examination corrected for full length was 0.63.

The instruction and original testing were given within a 1-hour class period to classes of 40 aviation cadets. Approximately 120 men were included in each of 4 groups. These groups were the film group, the manual group, the lecture group, and the no-training group. This last group was added to provide a base line for determining the actual amount of learning resulting from the procedures used for the other three groups. A few individuals had to be eliminated from each of the groups because they already had some information regarding the position-firing methods and procedures which were the content of the instruction.

The film required only 15 minutes of running time. It proved impossible to give an adequate lecture covering this topic in less than 30 minutes. It also was found that at least 30 minutes were necessary for adequate study of the manual. Since the experiment was to be a comparison of the 3 methods with each working at its best level of effectiveness, it was decided to allow 30 minutes for the lecture and 30 minutes for the supervised study of the manual. The film was shown just once and required 15 minutes. Thirty minutes were allowed for the examination. The results are shown in table 6.9.

TABLE 6.9.—*Mean scores and amount learned immediately after testing*

| *Group* | *N* | *Mean score* | *SD* | Amount learned (*score minus 5.42*) |
|---|---|---|---|---|
| Film (15 min.) | 132 | 17.91 | 3.0 | 12.49 |
| Manual (30 min.) | 101 | 15.43 | 4.2 | 10.01 |
| Lecture (30 min.) | 101 | 15.10 | 4.6 | 9.77 |
| No Training | 122 | 5.42 | 3.1 | |
| MEAN SCORES AND AMOUNT REMEMBERED AFTER 3 MONTHS | | | | |
| | | | | (*score minus 6.96*) |
| Film | 98 | 16.36 | 4.10 | 9.40 |
| Manual | 64 | 12.96 | 5.17 | 6.00 |
| Lecture | 66 | 13.58 | 4.93 | 6.62 |
| No Training | 24 | 6.96 | 4.20 | |

As indicated in the table, the film produced considerably better learning of the material taught despite the fact that it occupied only half the time required by the other methods. On the average, 2.5 more items were learned by the film group. This difference is sufficiently large that it is unreasonable to assume that it was the result of fluctuation in the quality of the students in the various samples. All of the students who could be located and tested were tested with the same examination 2 months later. Here again it is seen that the film group had significantly higher average scores than the groups taught by the other two methods.

To investigate the reasons why the training film produced greater learning than the other two methods, a detailed item analysis of the responses for each group was made. The test questions were grouped according to the one of the 14 points to which they were related. It was found that on 17 of the 25 items the film group was superior to the other group. This superiority was evident for 13 of the 14 points. On one point which had to do with the rules for the direction which the deflection ("lead") should take under certain conditions the other groups were superior to the film group. Examination revealed that these conditions were explained in more detail and more clearly in the lecture and in the manual than in the film. It should be emphasized that the points could have been explained effectively in the film but happened to have been neglected. The lecture and manual were checked to see if their contents were equivalent to that of the film on all points where the film showed superiority. They were judged equivalent, or in some cases better. The treatment was in some instances more detailed since the time allowed was greater.

The items and points on which the film group showed the greatest advantage were carefully studied for clues as to the reasons for the success of the film. These concepts seemed to have the common feature of being "dynamic" in the sense that they dealt with changing events or with the variation of one thing in relation to another. For example, the concept of increasing amounts of deflection with continuously changing angle of attack of the fighter plane and the fact that the amount of deflection must be gradually changed while following the fighter showed 20- to 40-percent superiority for the film group. The film group was also much better in understanding ideas that get their meaning from human actions.

Diagrams and verbal descriptions seem to be much less effective in teaching about the unit of deflection in sighting than the motion picture. In preparing the animated film it was possible to place the viewing position of the camera in the position of the trainee in a learning situation so that the student could see what he would see in operating the gun and place himself in the actual activity rather than as an onlooker. After explaining what is to be done in words and with diagrams in the film, the character "Trigger Joe" scratches his head, and sets about to do it. At this point the camera fades back to where only his hands and the results of his manipulations are visible.

This technique with "Trigger Joe" moving along rather slowly enables the students almost to experience the activity to be learned and at the same time get immediate confirmation or correction from "Trigger Joe" as his thoughts come over the sound track while he is trying to figure out just what it is that he is supposed to do. The student is active and always ahead of him, since "Trigger Joe" gives the appearance of having things come to him only with difficulty.

There is no question but what the incorporating of a thread of a story and the building up the character of "Trigger Joe" give this film real entertainment value. However, it is important to note that in the case of "Position Firing" the humor and dramatization were subordinated to the instruction and were used for the sake of instruction rather than for their own sake.

On the basis of this experiment and other experience of the personnel of the Psychological Test Film Unit, a careful analysis of the special advantages of the motion picture as a training device was made. The motion picture enables the learner to see inside of operating mechanisms through the use of animated photography, to observe completely cut-a-way models, and to emphasize certain parts or processes by color or moving pointers, or other devices. It is also possible by animation or by slow-motion photography to modify the velocity of a series of events so as to make them more easily understood.

Motion pictures can show much better than any other means how one thing varies with another in time. Precise timing, which is of great importance in many skills, can be very effectively presented.

The motion picture is very well suited to showing situation from a subjective point of view, the camera can actually take the position of the performer instead of the usual position of an onlooker in the learning situation and give the student a fairly active rather than a purely passive role in the learning experience. At the same time the student can be faced with choice after choice and be corrected when he chooses the wrong one and confirmed when he chooses the right one. At the same time he can be shown the consequences of the wrong choices which he might have made.

The realism which comes from seeing an enemy fighter "barreling in" on "Trigger Joe" with the latter firing with his sight right on the target (without proper lead) and therefore continuously missing contrasted with the result when he remembers to use the proper lead (the fighter spiraling down in flames) can be very effective. By comparison an instructor suggesting that good results will not be obtained unless certain rules for lead are followed provides a pale and lifeless learning situation.

The motion picture also offers greater opportunities for personalization of abstract ideas by cartooning and caricaturing and personifying or symbolizing electricity, forces, stresses, and similar phenomena which are ordinarily difficult to illustrate directly. Comic emphasis, if properly used, can be a valuable teaching aid and is especially well adapted to use in motion pictures. It has frequently been misused in the development of wartime training films, however.

It is clear that in certain defined situations the motion picture has important advantages. However, it should not be forgotten that it also has disadvanages. These are principally with respect to such things as personal relationship and guidance which a good teacher can bring into the classroom situation. Properly used along the lines suggested by experimentation and experience, the training film can be a very effective aid to instruction.

### Studies in Flexible-Gunnery Training[7]

Surveys made by personnel of the psychological research group at the Central School for Flexible Gunnery indicated that gunners arriving in combat theaters were much less proficient than the gunners being graduated from the schools in the AAF Training Command. It therefore appeared desirable to initiate an improved program for continuation training of gunners in the training air forces and to evaluate the effectiveness of this program as compared with that of the existing program at a typical training station. Accordingly, an experimental training program was developed and approximately half of a group of 400 gunners in training at Pueblo Army Air Base were instructed under the new program while the remaining half used the program previously in effect at that base.

In planning the experimental program a number of principles based on psychological and educational experimental findings were used. The five principles given special emphasis in planning this program were:

1. *Learning the whole job in context.*—Two general courses were developed to bring the various parts of the gunner's job into relation with each other. One of these was on the care of equipment and the other on the use of equipment. In these courses operational exercises which were designed to require the gunner to perceive his job as a whole and to practice specific skills in proper context and sequence were introduced.

2. *Motivation.*—Men who achieved perfect performance on specific phase checks were excused from further drill on these phase checks during the period in question. It was also believed that the new operational exercises increased motivation because of their greater realism.

3. *Spaced Practice.*—Instead of having all gunnery-training practice concentrated in a block of several hours, the practice was broken up into smaller units and spread over a longer period of time.

4. *Knowledge of results.*—Gunners were informed immediately of errors in the phase checks and phase-check answer sheets were used by review instructors to aid the gunner on specific weaknesses. Progress charts were also used to maintain the interest of the gunner in his skill.

5. *Learning by doing.*—The new program stressed actual performance on the job by the student.

Suitable criteria of success in the primary task of a gunner were not available to provide an adequate evaluation of the results of this experiment. It

[7] The studies reported in this section were the work of R. N. Hobbs, W. B. Schrader, M. J. Hahn, A. C. Jensen, W. S. Gregory, R. W. Russell, F. J. Harris, J. F. Lawrence, and J. A. Valentine.

was found, however, that in the experimental group, which made frequent use of the turret phase check during the course, perfect scores were made by almost all (97 percent) of the individuals on this check at the end. Only 38 percent of the individuals trained under the old program made perfect scores on this phase check upon completion of combat-crew training.

This study at Pueblo resulted in a general reorganization of gunnery training in all of the training air forces. Based on the results of the project and on experience gained in planning the experimental program a manual was prepared entitled "Standard Flexible-Gunnery Program for B–17 and B–24 Combat-Crew Training Stations, Instructors Guide and Lesson Plans." All training air forces were directed by Headquarters Army Air Forces to revise their gunnery progams in accordance with procedures outlined in this manual.

As mentioned earlier, it had been found that gunnery proficiencies were not being maintained adequately in combat groups. In January 1945, officers of the psychological research group at the Central School for Flexible Gunnery were requested to conduct a study in one of the combat groups in the Eighth Air Force. Following initial proficiency testing, each of the four squadrons was trained under a different training program. At the end of one month's training all the gunners remaining in each of the four groups were retested to determine progress made under each system of training.

The four squadrons were trained as indicated below:

*Squadron 708.*—Each gunner flew three gun-camera missions and was required to practice until he could pass all phase checks with no errors and make a certain minimum qualifying score on the Jam-Handy (E–14) Trainer. He was also required to get a specified number of hits on the Burst Control Range. When these levels of proficiency had been met the gunners were excluded from further practice until the final test.

*Squadron 709.*—These gunners flew one gun-camera mission and practiced on the phase checks and the Jam-Handy Trainers 3 hours a week throughout the training period.

*Squadron 710.*—These gunners received the traditional Eighth Air Force Gunnery-Training Course. This consisted of 1 hour a week practice on the turret trainer or on the Jam-Handy Trainer, plus 3 hours per month of lectures.

*Squadron 711.*—These gunners received no formal training at all beyond that inherent in the taking of the tests used.

Although various factors made an adequate evaluation of the results of this experiment impossible, it is clear that Squadrons 708 and 709 made better records at the end of the month on the phase checks than the other two squadrons. The results suggest that the members of the squadron in which gunners were excused after making perfect scores on the phase checks were motivated to obtain these perfect scores as soon as possible. However, they then rapidly forgot what they had learned so that by the end of the month their scores were no better than those which continued a regular training program of 3 hours a week during the month. The scores of all three squadrons which had some types of training were very similar on the Jam-Handy Trainer. The fourth squadron which received no training was definitely inferior on this trainer at the end of the month.

As a result of this experiment, a training program for gunners in the Eighth Air Force was proposed based upon proficiency standards. Each gunner was to be required to take the proficiency tests every 2 weeks. If he were found to be below the qualifying standards with respect to any of these proficiency measures, he was to be required to attend regular practice sessions until he could again qualify on this aspect of his proficiency as a gunner.

This experiment attracted considerable attention in the Eighth Air Force and is believed to have been very influential in improving gunnery training in the groups of that Air Force.

## SUMMARY AND CONCLUSIONS

The findings reported in this chapter indicate the types of training problems typical of military training schools. These studies emphasize the serious deficiencies both in and out of the military service. The primary problem of any program of education or training is to determine and specify the desired outcomes of the training program. It is believed that most of the defects in educational or training programs are related not to inefficient learning situations, but to having individuals learn the wrong things. In the military situation the fundamental basis for determining what shall be learned is a thorough analysis of the operating and combat requirements of the assignments for which the individuals are being trained.

Many studies were carried out by personnel of the Aviation Psychology Program in an effort to gain more accurate information regarding combat requirements. Of special importance are those requirements which are critical in the sense that a significant number of operational missions have succeeded or failed because of the presence or absence of this element or quality. There can be no substitute for human values and human judgments in determining the relative importance of the many requirements of the specific assignment for which an individual is to be trained. Once the requirements of the assignment have been defined in detail, it is presumably possible to proceed, using rational and scientific procedures, to an exact solution of instructional problems.

It is likewise clear that the first technique which is necessary is a procedure for evaluating the results achieved in terms of the defined job requirements. Valid measures of success in fulfilling requirements must be available. Much time has been spent on this problem by personnel of the Aviation Psychology Program within the past few years and important progress has been made.

With respect to the problems and procedures of training itself including selection and training of instructors, curriculum content, training devices and equipment, and training methods, a large number of studies have been conducted. Unfortunately, very few of the results obtained can be regarded as crucial because they had in most cases to depend on partial and internal analyses for the evaluation of the results. Probably the most important contributions which have been made in this area are those related to the definition of the problem, the exploration of procedures for evaluating success, and the important experience gained in designing and conducting experiments within an operating organization.

There have also been a number of important negative findings such as the lack of value of general ratings which are not based on systematic observation and the specificity of many types of skills. On the positive side there has been a tendency to find many training situations in which knowledge of well established psychological and educational principles seem to govern the results. In these situations it would be possible for almost any expert having experience with and knowledge of these principles to point out serious flaws in the training procedures in use and suggest revisions which would greatly increase the efficiency of training procedures.

CHAPTER SEVEN

# Research on Problems Regarding Operational Procedures[1]

## CAUSES OF MISSION FAILURES

Studies of the critical requirements for success in combat of the type discussed in the section on the development of training curricula are important for evaluating operating procedures as well as selection and training procedures. The survey of reasons for mission failures in the Eighth, Ninth, Twelfth, and Fifteenth Air Forces in the winter of 1943 and 1944 indicated not only the importance of selecting and training superior personnel for essential assignments, but also the need for a more accurate evaluation of the effectiveness of lead crews in combat operations. The report included recommendations regarding the number of combat missions which pilots could fly under various conditions of operations existing at that time without serious impairment of operating effectiveness. It was also recommended that research on procedures for bombing-through-overcast be conducted by aviation psychologists.

With the development of a psychological research program in the redistribution stations of the AAF Personnel Distribution Command, it was possible to continue these studies of reasons for mission failures by interrogating personnel passing through the redistribution stations who had completed combat tours in the various theaters. Accordingly, in early 1945, questionnaires were devised to obtain explanatory information concerning the extent of the mission failures, the causes of such failure, and certain control factors such as overseas air force, type of aircraft involved, formation position, and purpose of the mission. In the first questionnaire the individuals were asked to describe the mission which they believed was least successful.

A second form was prepared based on the experience with the preliminary questionnaire. This form included more specific questions which were designed to secure uniformity in the amount of detail included in the descriptions. Bombardiers and navigators were interrogated on separate forms and each officer was requested to report on personnel errors peculiar to his aircrew position and encouraged to describe errors of his own that led to an unsuccessful mission.

[1]The research reported in this chapter is mainly the work of C. W. Harris, M. J. Williams, J. F. Kamman, and the editor of this report.

After some experience with this questionnaire, a third form was prepared for use with all air-crew positions. On this form the individual was asked to select a mission on which he was unsuccessful in accomplishing his mission because of errors that he made. He was also asked what he or others could have done to prevent or correct his error. In an attempt to broaden the type of incidents reported, some of the individuals were asked to describe mission failures which occurred at the beginning of the combat tour and others were asked to describe mission failures occurring near the end of their combat tours. Another innovation was to ask the individual what specific training before or after going overseas would have helped to prevent or avoid his error. These 3 questionnaires were administered to approximately 250, 500, and 1,500 members of returned air-crews respectively.

The importance of personnel errors as a cause for mission failures was indicated by the responses to the first questionnaire. In this questionnaire individuals were asked to select the one mission which they believed was the least successful. Their reports indicated that 78 percent of these errors were due to participating personnel, 14 percent due to briefing and planning errors of other personnel, and only 6 percent were due to equipment failures, weather conditions, and other circumstances beyond the control of the personnel involved. The remaining 2 percent were caused by miscellaneous factors. It was also found from this analysis that nearly all of the errors of participating personnel reported were errors of lead-crew personnel.

The percentages given above are based on subjective reports and cannot be regarded as providing valid objective evidence of the relative importance of such factors as equipment and weather as compared with personnel. However, they do indicate that personnel errors are very much in the minds of these men and are fairly frequent in occurrence.

Since the air-crew personnel were not given a definition for "least-successful missions" it is of interest to analyze the type of mission described. In 76 percent of the missions described as least successful by these men, there was failure to bomb the assigned target. Only 2 percent reported on missions in which the assigned target was successfully bombe The criteria of failure for the remaining 22 percent of the missions could not be ascertained.

## ANALYSIS OF BOMBING ERRORS

The results obtained from bombardiers using the second and third questionnaires have been combined and are reported in table 7.1 on the following page.

It will be noted that these bombardiers report that one of their principal errors was failure to identify the target soon enough or accurately. These errors accounted for approximately 20 percent of the mission failure reported. A typical example is quoted below as submitted by a lead bombardier of the Eighth Air Force:

TABLE 7.1.—*Reports by 202 Bombardiers in the Eighth and Fifteenth Air Forces of errors made by them which resulted in mission failures*

| | *Error category* | *Frequency in percent* |
|---|---|---|
| 1. | Errors of target identification and pilotage | 20 |
| 2. | Errors in use of bombsight or related equipment | 20 |
| 3. | Errors of decision and judgment | 15 |
| 4. | Errors due to unsatisfactory coordination within own air crew | 15 |
| 5. | Errors due to failure to inspect and maintain equipment | 11 |
| 6. | Errors due to failure to check switches, extended vision, etc. | 7 |
| 7. | Errors due to unsatisfactory coordination with personnel of other crews | 5 |
| 8. | Accidental release or failure to observe bomb-release signal | 4 |
| 9. | Errors due to distractions and divided attention | 3 |
| | Total | 100 |

The day was perfect for good (visual) bombing, but I failed to identify the right airfield to bomb. There were two grass airfields about 4 miles apart, and in my haste I picked up the wrong one in the sight. Outside of that everything went off normal. I hit the wrong airfield with a good pattern.

Another very frequent cause of mission failures was incorrect use of the bombsight or related equipment. This also accounted for approximately 20 percent of the mission failures reported. A fairly typical example of this sort of error is the one reported below by a Fifteenth Air Force bombardier:

There was a mixup at the IP (initial point of the bomb run) and our box was forced to swing around coming in 90° to briefed heading. By the time we were lined up on target, crosshairs were already over target when indices had crossed. Did 180° turn and picked up target of opportunity. Telescope motor was left running and I hit optic clutch. It failed to completely engage without my knowing it. Attempted synchronization but bombs dropped 4 miles short.

Another common cause of mission failures consisted of errors of decision and judgment on the part of the bombardier. Such errors accounted for approximately 15 percent of the mission failures. A typical example is quoted from a Fifteenth-Air-Force lead bombardier below:

The mission in question was flown over the Brenner Pass and as squadron was too close to the squadron in front I turned out over the mountains, and in doing same, gave the target enough time to be completely obscured by smoke from previous squadron's bombs; however, upon coming back or making our initial run over the target we lost one ship and had several others badly damaged, so again I turned off, missing some of the flak, and bombed a marshalling yard in the Pass. The mission for my part was unsuccessful in that a target of opportunity was not specifically specified for that particular mission.

Failure of the crew members to work together smoothly as a team was also a frequent cause for mission failures. Errors of this type were found to account for 15 percent of the mission failures reported. A typical report by a Fifteenth Air Force lead bombardier is quoted below:

Bombing small marshalling yard. Group broke into box formation before IP and each box started their own run. IP to target was 90 seconds on run. C-1 (automatic pilot) was ineffective. Colonel called to center PDI (pilot direction indicator). C-1 failed second time and then called for PDI run. Thirty seconds remained to kill drift on course and bombs hit 500 feet west of target.

I should have taken charge of plane as soon as C-1 failed and made run PDI.

Failure to inspect and maintain equipment accounted for 12 percent of the mission failures. This included failures both in preflighting equipment and also in preventing bombsight and bomb-bay doors from freezing. A typical example reported by a lead bombardier of the Eighth Air Force is quoted below:

> First part of bomb run was made over 10/10 clouds using P.F.F. (radar) procedure. Up to 2 minutes before "bombs away" there was no chance to check pointing ourselves on the ground. The target broke clear and the bombardier made a 20° correction on the secondary. Bombardier got the target into his sight and would have had at least good results if bombs had dropped. The bombsight release mechanism was out of order and bombardier salvoed 10 seconds after indices had passed.
>
> Bombardier could have put less faith in radar operator and picked up target sooner. He failed to make a complete preflight check of bombsight and could have taken less time to salvo.

A specific error which accounted for 7 percent of the mission failures was failure to check the switches, extended vision, and other levers, causing the bombs to fail to release or to be dropped prematurely. An example of this type of error is provided by a lead bombardier of one of the squadrons of the Fifteenth Air Force:

> I had been using extended vision on first part of run. Failing to check to see if it was all rolled out I went ahead and dropped my bombs. The other ships in the box dropped on me. The entire box's bombs hit approximately 3 miles short of target on course, in the middle of a field.
>
> I should have been more cautious to check the extended vision after using it. Poor procedure on my part.

An appreciable number of missions failed because of lack of satisfactory coordination with other air crews. This was usually between the lead and deputy-lead crews or in some instances with other groups or squadrons. An example typical of the 5 percent of the missions which were failures for this reason is quoted from a bombardier in the Fifteenth Air Force who was flying as deputy lead:

> I was flying deputy lead and followed through the rate on the bombsight. Lead ship had a malfunction. My bombs dropped when my indices met and the rest of the formation dropped their bombs on me. With the lead ship's malfunction he was off on course and the bombs missed the target. Lead ship did not drop his bombs because of malfunction.
>
> I should not have dropped my bombs.

Another very specific type of error which accounted for 4 percent of the mission failures was the accidental release of bombs by brushing against the toggle switch, rolling the telescope index up without releasing the trigger or in some similar manner. Some of these errors in this category were failures to release the bombs on the signal because the bombardier was not alert or did not realize the formation was in the target area. A typical example quoted from the report of a bombardier who was not flying in a lead position on the Fifteenth Air Force mission reported is given below:

> My flak suit fell off on bomb run, disconnecting oxygen tube—in connecting same accidentally toggled out bombs—two other planes released on our bomb drop. Could have been more careful in putting on flak suit. Navigator who helped put on flak suit did not hook snaps properly.

Although excitement and distractions on the bomb run contributed to many of the errors previously mentioned, they were also of predominate importance in about 2 percent of the mission failures reported. The following paragraph is quoted from the report of the lead bombardier of one of the Fifteenth Air Force squadrons:

> Visibility was CAVU (ceiling and visibility unlimited). No flak nor fighters were encountered. The pilot (flying his first lead) had difficulty in his spacing, consequently set up a collision course with the lead box. I kept pulling my head out of the sight to make sure we wouldn't drop on the other box and in the excitement I rolled out on a marshalling yard somewhat similar in appearance but just 5 miles short of the briefed target. Ironically I even missed the yards I had meant to hit. An error in the lead box's altimeter forced us to increase our altitude above briefed to stay above them. In the excitement caused by the collision course this change in altitude went unnoticed thus causing the pattern of bombs to fall over the target. This was my first mission in lead position. I could have and should have held my bombs and called for a 360° turn, and made a run on the target without the confusion of a collision course with the other box.

Although it is evident that a number of these errors reflect deficiencies in the selection and training of personnel, it is also apparent that many of the failures could have been remedied by developing improved operating procedures. Reading over these stories of mission failures does not give one a sense of a smooth well-oiled machine operating with occasional lapses. It is believed that a systematic research and development program on operating procedures could contribute substantially to the effectiveness of combat operations.

## ANALYSIS OF NAVIGATION ERRORS

A similar analysis has been made for the reports of 243 navigators who replied to the second and third forms of the questionnaire. Table 7.2 presents the results of the analysis of these navigators' reports of their own errors which resulted in mission failures in the Eighth and Fifteenth Air Forces.

TABLE 7.2.—*Reports by 243 navigators of errors made by them which resulted in mission failures in the Eighth and Fifteenth Air Forces*

| | *Error category* | *Frequency in percent* |
|---|---|---|
| 1. | Errors due to unsatisfactory coordination within own air crew | 19 |
| 2. | Errors of decisions and judgment | 15 |
| 3. | Errors in using navigational equipment and procedures | 15 |
| 4. | Errors due to failure to do continuous navigation | 12 |
| 5. | Errors in target identification and pilotage | 11 |
| 6. | Errors due to unsatisfactory coordination with personnel of other crews | 10 |
| 7. | Errors due to failure to inspect and maintain equipment | 9 |
| 8. | Errors due to inadequate preparation for the specific mission | 7 |
| 9. | Errors due to distraction and divided attention | 2 |
| | Total | 100 |

Although differing in certain details because of the different jobs of the bombardiers and navigators, the general picture from this analysis is similar to that obtained for the bombardiers. The navigators report their largest number of errors to have been due to unsatisfactory coordination within their own air crew. This cause is given for 19 percent of the mission failures reported. Below are reproduced two reports, one by a navigator in the Eighth Air Force who was squadron lead and the second by a navigator in the same air force who was wing lead:

Upon reaching briefed IP (initial point for bomb run), Group Leader aborted without any signal or R/T (Radio Telephone) message. Approximately 25–30 miles past IP, Low Squadron returned to make second run alone on primary. Target area clear, visibility 25–30 miles. IP was made on second run. Started bomb run to target—fuel dump in a wooded area—wooded areas only check points on run. Wind used was good—course checked for about 10 miles of 25-mile run. At this point, about 6 minutes prior to bombs away, Air Commander called and became excited when his call was not acknowledged. Meantime I had been checking the run, and got away from nose section to answer A/C (Air Commander). While I did, bombardier called for a correction. After correction I lost the position and missed target by 3 to 4 miles. We had a strictly PFF (radar) secondary, so we bombed a target of opportunity successfully. I should have refused or *ignored* A/C's call as to time left to target!

* * * * * * *

Both the bombardier and myself were unfamiliar with the use of PFF (radar). A new mickey (radar) man was flying with us and he made a mistake in his fix. I took his word instead of following my D. R. (Dead-Reckoning). Had about 8/10 cloud cover with haze. Meager to moderate flak. Bombed similar target about 11 miles from assigned target. I should have relied upon my own work here. The mickey man should have been more careful in identification. All of us too nervous. Should be more team coordination in States. Crew members do not know enough about other jobs when they are sent overseas, especially the pilot in regards to bombing and navigation.

Also as in the case of the bombardier, errors of decision and judgment accounted for a large number of mission failures. The navigators reported that 15 percent of their errors causing missions failures were of this type. A number of reports referred to not enough confidence, or too much confidence in dead-reckoning and pilotage results, or similarly, not enough confidence or too much confidence with radar and other types of radio-navigation aids. One of the reports made by a lead navigator of the Eighth Air Force illustrating a situation in which this navigator thought the navigator leading made an error in judgment by relying too much on pilotage is reproduced below:

Bombing railway yards at Koblenz. Mission went well up until 40 miles from the IP. At that time, the instruments in the lead plane became inoperative. They called on the deputy to take over. During the mixup due to pilot fault the lead navigator became lost just before the IP, missing it about 10 miles to the left. Instead of getting fixes he sat back and waited until the Rhine river came up, and then turned south toward the target. Flak had been hitting us for about 15 minutes due to being off course to the left. The run was made 90 degrees to what it should have been, and the bomb runs very short, due to ground haze. Bombs missed.

There were also a substantial number of errors found due to ineffectiveness of navigational equipment and inaccurate work. Some of these errors were due to lack of adequate training on some of the special navigation aids in use in the European and Mediterranean Theaters. However, quite a few of the mission failures (15 percent) were due as in the case of the following example, to poor technical navigation work. This story was reported by the deputy-lead navigator of an Eighth Air Force Group:

The navigational error occurred when we were approaching the IP. Due to the fault of the navigator's failure to check the drift, we went several miles past the IP. When we turned for the bomb run, the navigator failed to anticipate the wind on the run; therefore we had to keep correcting to the right. When the mickey picked up the target, he found that the wind had shifted, and that we would have to turn at right angles to the bomb run. We were so close to the target that when we were still in a bank the bomb sight released the bombs and therefore threw the bombs way left of the target. I should have had the mickey give me a fix more often so that I could keep better informed as to the nature of the direction of the wind. The leaders of the other two squadrons could have given the leader a check on his position more often, since we were dropping in group formation. Also it shows a lack of cooperation between the bombardier, navigator, and mickey operators.

To a much greater extent than the bombardier the navigator must keep continuously at work if he is to be able to meet the demands on knowledge of position which arise in an emergency such as being hit by flak and having to drop out of formation. However, many missions went by when the navigator carried through his work and never had any occasion to use it. It was therefore very easy for navigators who were not in the lead ship to let their navigation slide. There was also a tendency for lead navigators to fail to carry out continuous dead-reckoning navigation because they believed that they could rely on the radar operator to keep track of their position. These navigators reported that 12 percent of their errors were due to failure to do continuous navigation work. Below is given a report of one of the lead navigators of a group in the Eighth Air Force:

As it was 10/10 cloud below us, radar was used for navigation. Got careless about check points, and we were hit by flak which knocked out 9 planes from a group of 36. They had to return to base. This was in the first 100 miles of our journey. Our bombs hit primary but the bombs of the group were less effective because we were nine planes short when the target was reached. D. R. navigator in nose was hit and had to suffer through rest of mission. I should have been more careful about rechecking on my check points.

Although target identification was primarily the job of the bombardier, it was the responsibility of the navigator to direct the formation to the initial point of the bomb run, and get the formation started along the proper course toward the target. It was also usually customary for him to assist the bombardier in identification of the target area and also of the target itself. Errors in target identification or pilotage accounted for 11 percent of the failures reported by this group of navigators. The following account by one of the navigators in a formation in which the lead navigator missed the initial point illustrates this type of error:

Navigational error did not occur until nearly 15 minutes before the IP. The error wasn't too great, had it been done in undefended territory; but it amounted to over-shooting a visual IP and having to turn into an extremely strong head wind and spend 8 to 10 extra minutes in intense flak. Above it all the lead plane lost the target and had to cut out another group to get out. I should have kept my head in the few minutes when I knew we were at the IP and called the lead ship. I might mention that this was my second mission. The lead was depending too much on PFF equipment and D.R. with visual conditions existing. It had been emphasized that the IP must be made good.

On a short bomb run it is not possible for the bombardier to get a great deal of assistance from bombardiers of the other planes in the formation. However, it should be possible for the navigator to get a large amount of help from other navigators in his formation. The reports by these navigators make it very clear that in many instances, maximum use was not made of this available help with consequent mission failures. The navigators report that 10 percent of their missions failed primarily because of unsatisfactory coordination with personnel of other crews. The example given below is reported by a navigator in the Eighth Air Force who was not flying a lead position:

On this mission the radar set of the lead crew went out and the lead navigator had only D. R. He did not have an accurate wind and went off course badly. Navigators from other planes in the formation told him he was off course, but the Command Pilot would not let him correct. When the formation got flak from a town 20 miles left of course the C.P. allowed him to correct. He went back to course and then off again because his fluxgate compass went out. C. P. at this point let him relinquish the lead but the Deputy Lead could not get to primary and identified wrong secondary. Primary and Secondary were both missed but an important target of opportunity was smashed. The Command Pilot should have listened to the navigator both when he wanted to correct his course and let Deputy Lead take over as soon as radar set went out.

Particularly when bombing missions were carried on over a complete cloud cover the navigator had to rely very heavily on various types of navigational equipment. Most navigators knew so little about this equipment that they were lost if it failed to function properly. The 9 percent of errors due to failures to inspect and maintain equipment could probably have been substantially reduced if more adequate checking procedures and more thorough training in maintenance of the equipment had been given the navigators. The following account which involved a failure similar to the one in the preceding report illustrates this and also other errors of navigation. This lead navigator in the Eighth Air Force reported, as did many of the other navigators, several different types of errors, all occurring on the same mission.

Course was lost because lead plane had 30° error in fluxgate compass which was not discovered. The compass was OK during first 2 hours in air. All mickey sets were out within the first two hours and due to heavy cloud layer, all navigation had to be dead-reckoning. Formation began to leave course about halfway to target and missed target by 25 minutes although no one realized this at the time. Following D. R. and Flight Plan, formation went over three extra-heavy flak areas, lost bomber stream, and was under fighter attack for 20 minutes without escort protection. We wing ships could not give lead plane any useful navigation help. Why, I don't know. Lead D. R. navigator should have checked his compass against the pilot's compass and allowed for the error. Also wing navigators D. R. following the formation should have had more accurate D. R. positions with their good equipment.

The preparation of the bombardier for a particular mission consists largely of a thorough study of the target area and a check on his equipment. The navigator has a somewhat greater responsibility. He must keep informed as to boundaries of enemy territories, plot courses to alternate targets, have a thorough knowledge of emergency landing fields, be familiar with various forms of radio procedures, have a thorough knowledge of flak area and terrains in the vicinity of the route to be covered, and work out cooperative procedures with other members of the crew. It was reported that 7 percent of the mission failures were due to some defect of preparation for the specific mission. Below a lead navigator of the Fifteenth Air Force reports a mission failure due to inadequate preparation of a new navigator who was in charge of pilotage in the lead plane:

> Our plane led the group to the target which was on the Brenner Pass in Northern Italy. Bombs were dropped, but target was missed. Inadequate briefing of pilotage navigator on my part. He started our turn directly over IP instead of a bit early. Tall mountains on either side of briefed axis of attack obscured bombardier's vision of target until it was too late. We did not give him a long enough run. I should have given more thorough briefing of new pilotage navigator and reminder of duty coming up to IP.

The reader will have recognized in some of the preceding accounts that as in the case of the bombardiers, distraction, apprehension, and excitement were contributing factors to many of the types of errors listed. As in the bombardier reports, about 2 percent of the mission failures were indicated as having been predominantly influenced by this type of distraction. A lead navigator of the Eighth Air Force reports the following example of this type of error:

> After turning at IP, I looked from my radar set into the bomb bays and found them full of smoke. Called pilot and bombardier and then attempted visually to determine what was causing smoke. Assuming plane was on fire, I became more interested in doing something about situation than attending to my bomb run. Result, we toggled bombs on lead plane's release which proved inaccurate. I should have done my job after notifying pilot of situation in bomb bays.

Although as in the case of the bombardiers, it appears that a number of these errors might have been prevented by better selection and training, an even larger proportion of mission failures due to errors by navigators appear to be such that improvement in operating procedures would eliminate many of them. It appears that one of the most serious errors made during the past war was the failure to carry on a research and development program for procedures for using new equipment under operational conditions to parallel the research and development on the production of the equipment itself. A closer integration of the research programs on the development of new equipment and the selection and training of personnel to operate this equipment according to carefully designed operating procedures should greatly increase the effectiveness of combat operations.

## ANALYSIS OF ERRORS BY BOMBER PILOTS

The reports obtained from 153 bomber pilots using the third form of the questionnaire were analyzed in the same way as were the reports of bombardiers and navigators. Although there was some similarity to the types of errors previously reported, the analysis as shown in table 7.3 indicates that the problem of the bomber pilot is quite different from that of the bombardier and navigator.

*TABLE 7.3.—Reports by 153 bomber pilots of errors made by them in the Eighth and Fifteenth Air Forces which resulted in mission failures*

| Error category | Frequency in percent |
|---|---|
| 1. Errors in technique in operating the engine and controls | 34 |
| 2. Errors of decision and judgment | 27 |
| 3. Errors due to inadequate knowledge regarding equipment and procedure | 14 |
| 4. Errors due to failure to inspect and maintain equipment | 10 |
| 5. Errors due to inadequate preparation for the specific mission | 7 |
| 6. Errors due to unsatisfactory coordination within own air crew | 5 |
| 7. Errors due to unsatisfactory coordination with personnel of other crews | 3 |
| Total | 100 |

By far the most frequent type of error made by the bomber pilots were errors of technique in operation of the plane. Many of the reports related to the difficulty in operating controls so as to fly good formation. This was especially true of the B–24 pilots. There were also many errors reported regarding faulty operation of the engines. These errors accounted for approximately 34 percent of the mission failures reported by these bomber pilots. An example of this type of error is contained in the report given below made by a B–24 pilot from the Eighth Air Force:

> Due to lack of formation practice, I was not in formation at time of bombing. There was no opposition other than flak. I was at least 200 yards out of formation. I realize that one ship would not make too much difference but there were several other ships also out of formation. If the lead ship hit the target, which it did, the ships that were out of formation would break up the pattern enough to cause a highly unsuccessful mission. I failed to set a constant air speed that would keep me in formation and when I would get in formation I would come up too fast, I could not judge the difference in speed of the other ships. I should have had more training in formation before ever attempting to fly in combat. I had the absolute minimum requirement—20 hours.

The other principal type of error recorded by these bomber pilots was errors of decision and judgment. This type of error accounted for 27 percent of the mission failures. The types of decisions concerning which errors were made included errors as to whether to fly through clouds and weather or around it, whether to leave the formation when a malfunction occurred or when the plane was hit, or which alternate target to bomb when the primary target could not be reached. These errors accounted for 27 percent of the failures reported by these bomber pilots. A decision regarding weather which proved costly is reported by a lead pilot from the Eighth Air Force below:

> Decision to take a 63-ship formation thru high overcast. Resulted in breaking formation up and caused us to get bounced by enemy aircraft and the loss of several ships. I should have gone around front with formation instead of through it.

A rather substantial number, approximately 14 percent, of the errors were reported by these bomber pilots to be caused by inadequate knowledge regarding equipment and procedures. The pilots seemed especially to lack information regarding emergency procedures in handling malfunctioning equipment, rate of fuel consumption, and standard operating procedures in the groups and Air Forces in which they were flying. A report by a B–24 pilot on his first combat mission is fairly typical of these errors which account for 14 percent of the mission failures. The pilot whose story follows is from the Fifteenth Air Force:

We were about one-third of the way to the target. My engineer came up and said the fuel was half gone, and he thought we had a leak. I went back and checked the gauges, and the engineer was correct on the reading. Believing that we would not have enough fuel to reach the target and back, I turned back to the base. When we landed we still had half the fuel so obviously the gauges were incorrect. I should have known that even with only half our fuel at the point we turned back to the base, we could have made it to the target and home. We had gained our altitude so the pull was almost over. I found out on the next mission that we use very little gas returning from a target.

As in the case of the air-crew members previously discussed, a fairly large number of errors due to failure to inspect and maintain equipment were reported. The most common error was the failure to make an adequate pre-flight check of the equipment. These, together with various failures to do adequate inspection and maintenance work during flight, accounted for 10 percent of the mission failures. A fairly typical report by a B–17 pilot from the Eighth Air Force is quoted below:

Interphone out. Made run on target and dropped short. Should have checked on ground and had a radio man repair interphone. Co-pilot's first mission also bombardier's first lead. With interphone working, I could have given bombardier and navigator more confidence.

It was reported that a small number of errors were due to inadequate preparation for the specific mission. This group of errors, accounting for about 7 percent of the total, consisted principally in failure to give the crew adequate instruction and training prior to the mission. This is illustrated by the report of a B–24 pilot from the Fifteenth Air Force as quoted below:

Our ship went over the target. Intense flak was encountered. As pilot of ship I failed to make certain that nose turret gunner and bombardier knew their toggling procedure. As a result, nose gunner called bombs away late and our bombs dropped several thousand feet over the target. I should have checked to make sure all crew members knew their duties. All crew members could assume far more responsibility. Too many crew members go for the ride and "glory" with the attitude that the pilot will catch hell for any mistakes—and he always does.

It is of considerable interest to note that this group of bomber pilots who are responsible for coordinating the work of the various air crew members in their own planes and also coordinating with personnel in other crews report a very much smaller proportion of their mission failures to be due to unsatisfactory coordination. They indicate only 5 percent and 3 percent of the mission failures to have been due to unsatisfactory coordination within their own air crew and with personnel of other crews respectively. The most frequent type of failure these bomber pilots reported was the failure to coordinate with the navigator and bombardier regarding the position of the plane. In dealing with other crews, the principal types of errors mentioned were failure to notify the lead plane of an error in position or the failure of the lead plane to notify wing ships of the proposed course of action. Samples of these types of errors are given in the two reports which follow quoted from an Eighth Air Force B–24 and B–17 pilot respectively:

> Supposed to hit rail center in Luxembourg. Lead ship dropped early and other planes in formation dropped at the same time. Failure to coordinate with navigator caused wing ships pilots to order bombs dropped when almost 30 miles from target. Pilots should check with navigators on bomb run at least every 1 or 2 minutes.
>
> * * * * * * *
>
> Oil cooler went out and lost all the oil. Couldn't feather the prop and couldn't stay in formation. I failed to feather the prop soon enough. Failed to let my wing men know I was leaving the formation until we had dropped back quite a long way.

Although lack of training in four-engine equipment and in formation flying and lack of specific knowledge which should have been obtained in training are more evident here than for the bombardier and navigator, the lack of standard operating procedures is very evident in these reports of errors also. As mentioned in the preceding section, a close integration of new developments and equipment with training and standard operating procedures would have anticipated many of these errors and resulted in a substantial increase in the effectiveness of operations.

As indicated earlier, a special analysis was made of errors occurring during the first 10 missions and errors occurring in missions after the tenth for these heavy-bombardment operational groups. These analyses did not reveal any differences which were significant in these groups for bombardiers, navigators, or bomber pilots.

## ANALYSIS OF ERRORS OF FIGHTER PILOTS

In contrast to the analyses of reports from air-crew personnel in heavy-bombardment operations, the analysis of reports of errors by fighter pilots indicated definite differences in the errors reported for early missions as compared with those for later missions. Errors related to flying techniques and inadequate knowledge regarding equipment were very much more numerous in the reports of errors for early missions. The reports for missions beyond the twentieth included errors of decision and judgment and errors of naviga-

TABLE 7.4.—*Reports by 157 fighter pilots of errors made by them in the Eighth, Ninth, and Fifteenth Air Forces which resulted in mission failures*

| *Error category* | *Frequency in percent* | | |
|---|---|---|---|
| | Missions 1 to 20 ($N = 82$) | Missions 21 and beyond ($N = 75$) | Total ($N = 157$) |
| 1. Errors in decision and judgment | 16 | 45 | 30 |
| 2. Errors in technique in operating controls and equipment | 30 | 8 | 20 |
| 3. Errors in navigational procedures | 7 | 23 | 15 |
| 4. Errors due to inadequate knowledge regarding equipment | 13 | 5 | 15 |
| 5. Errors due to failure to make routine checks and inspections | 30 | 12 | 15 |
| 6. Errors due to unsatisfactory coordination with other pilots | 4 | 7 | 5 |
| Totals | 100 | 100 | 100 |

tion to a much greater extent than did the earlier missions. For both groups a number of failures to make routine checks and inspections was reported and neither group reported many errors due to unsatisfactory coordination with other pilots. These results are shown in table 7.4 which provides an analysis of the errors reported for early missions, later missions, and for the total group.

An example of the type of error which was more typical of the experienced fighter pilot is given by a P-51 pilot in the Eighth Air Force in the paragraph quoted below:

Little ack-ack was encountered and we were making several passes at a number of live engines near Heidelberg. I became too bold and eager and wanted to see my results so on one good pass I pulled up wide and dipped my wing to see results. This gave a 20-mm. crew a perfect target. I had to return to base.

An example of the kind of error in flying technique made especially by the pilots on their first few missions is given in the following paragraph as reported by a P-38 pilot from the Eighth Air Force:

Leader and myself peeled off at two Jerries 5,000 feet below who were attacking two P-38's. I became too eager—picked up too much speed on the dive—when the leader leveled out behind the Jerry I passed them both, losing all effectiveness and becoming a target on my own. Later climbed safely back to altitude and rejoined the group. I should not have used quite so much throttle on the dive.

Pilots on their first few missions have less responsibility for navigation and are less likely to get off by themselves and have to rely on their own navigational procedures. A number of the fighter pilots describing navigational errors reported on reconnaissance missions. A typical report by an F-6 pilot from the Ninth Air Force is quoted below:

In planning the legs I misread the distance scale on the plotter which doubled the time of flights on two legs of a triangular course. I couldn't have seen anything because of the weather anyhow, but as a result of the mistake, I had to find my location after the letdown. No harm was done because I was very familiar with the territory. I should have taken more time figuring time of flight on the legs of the course.

As in the case of bomber pilots, many of the fighter pilots reported inadequate knowledge regarding equipment and as mentioned previously this was much more true of pilots during their first few missions. A typical statement is quoted from a P-47 pilot from the Ninth Air Force:

Cylinder head temperature gauge became unserviceable. Aborted and returned to base before entering enemy territory. In aborting, command of flight was taken over by inexperienced leader. I should have known how to tell condition of engine with various instruments inoperative.

Routine checks and inspections were frequently neglected by these pilots. A typical report of a P-47 pilot from the Ninth Air Force who was in such a hurry that he couldn't bother to check is quoted below:

Took off with engine cold. Thought something was wrong with oil pressure. Returned to field. It was merely a hurry-up mission and I was too excited to check my instruments properly. I should have stayed on the ground a little longer or just gone on and let her warm up. I was off the ground. That was the main thing.

Errors due to unsatisfactory coordination with other pilots were generally reported as failures to understand instructions or failures of a wing man and leader to stay together. A fairly typical example reported by a P-51 pilot from the Eighth Air Force follows:

Flying through an overcast, I did not have the confidence in my element leader I should have had. I though that he was in a spiral and left him. However, he continued on with squadron and I was lost. I spent about 3 hours trying to find my squadron. I should have trusted the instrument flying of my element leader.

## IMPLICATIONS AND CONCLUSIONS

It is clear from these analyses that much improvement in operations could be gained by systematic tabulations of reasons for mission failures and the development of operational procedures, routines, classification procedures, and special training programs. Effective work of this type demands thorough knowledge of operating procedures which can only be gained by close association with operating units over an extended period.

The psychological branches of the Central Medical Establishments in combat air forces, which were authorized in the summer of 1945, were given this type of research as one of their principal functions. It is believed that these organizations could have been very effective in providing technical assistance to operating personnel in conducting systematic surveys and in performing actual controlled experimentation in connection with combat operations.

The myth that each separate operation is so important that its success cannot be jeopardized for the sake of obtaining accurate information on which to base future operations has been relegated along with the horse cavalry and the battleship to the limbo of lost causes. The effectiveness of future military operations must be a matter of continual, systematic, and thorough scientific observation and experimentation.

# CHAPTER EIGHT

# Studies of Individual Reactions to Combat

## INTRODUCTION

One of the most fundamental of all military problems is the morale of the personnel involved. Discussions of morale in military groups frequently imply that morale is a vague quality of a large group which in some unspecified way affects the outcome of battles and wars. In an effort to provide a more specific and definite basis for understanding the origin, nature, and effect on operations of morale in the military situation, a series of studies were carried out on the reactions of individuals to the combat situation. It is believed that only through intensive studies of individuals is it possible to obtain satisfactory insight into this important factor.

A preliminary survey of individual reactions to combat which provided background for later studies in this area was carried out as part of the survey of air-crew personnel in the Eighth, Ninth, Twelfth, and Fifteenth Air Forces in the winter of 1943–1944. The report of this survey emphasized the high losses which were typical of air-force operations at that time.

Bombardment operations in which the attrition rate was approximately 5 percent for each bombing mission certainly represent the most hazardous military operations which have been conducted over a sustained period. Mathematical calculations indicate that for an attrition rate of 5 percent applied to the remaining group in the case of each successive mission, only 277 men out of a thousand can be expected to remain after 25 missions. This theoretically expected proportion of men completing 25 missions is slightly higher than was found in a group of something more than 1,300 air-crew personnel checked through from initial assignment to heavy bombardment groups in the Eighth Air Force to termination of duty with these groups. Fortunately, such severe attrition was not typical of all types of combat operations in the Air Forces. In many units an attrition rate of 1 percent for each mission was more typical. Under such conditions 605 of the initial group of 1,000 men could be expected to complete 50 missions.

However, whether we consider operations in which an individual's chances are as low as one in four or a little better than one in two of finishing his tour of operational duty, it was definitely a hazardous business.

In the face of this it was encouraging to find that morale in these groups was generally high and that there was very little actual breakdown of personnel. A number of instances were observed where groups after losing half of the planes they sent on a particular mission nevertheless sent out a full group on the mission the following day.

On the basis of this survey of a large number of groups of various types engaged in combat operations in the European and Mediterranean Theaters, it was concluded that the principal factors in keeping these men flying and fighting in addition to the more immediate motivation of finishing up and getting home, were the leadership of the group and the character and temperament of the individuals.

## Leadership

Informal observations indicated that groups with better bombing accuracy records, lower proportions of abortive planes, and a smaller number of cases of anxiety reaction to combat stress were those in which the commanders and flight surgeons were superior leaders. The superior commander appeared to know his men and know their jobs to a much greater degree than did the average commander. He also took an active interest in the welfare of his men but did not allow this to extend to pampering or relaxing the requirements that combat demands be met promptly and efficiently. The superior flight surgeons appeared to be strong, steady, masculine, and aggressive personalities who took a real interest in the welfare of the men. However, as in the case of the superior commander, this sincere interest in the individuals in the combat unit was oriented around the fact that the group had a job to do and the primary consideration was getting this job done.

The good group commander and flight surgeon took a good deal of interest in the welfare of their personnel with reference to such things as food, living conditions, recreation, medical care, promotions, and awards. However, these factors did not seem in themselves of primary importance in keeping the men flying and fighting. The human being is very adaptable and if necessary he can exist and function quite efficiently in conditions far more primitive than those to which the people of this country have become accustomed.

The primary motivating force which more than anything else kept these men flying and fighting was that they were members of a group in which flying and fighting was the only accepted way of behaving. The air-crew combat personnel were closely knit together. First, because they flew, and second, because they fought. In combat operations they lived together and had little contact with people outside the group. They were usually too far away, too busy, and found transportation too difficult to make any friends outside the group during the relatively few months or weeks that they were in the group. Under these circumstances, the individual identified himself very closely with the group. He took great pride in his membership in the group and wanted to be told that the job they were doing was a very important one and that their operations were effective. He was especially sure that the personnel and special operational procedures of his group were superior to those of other groups. The task of creating and maintaining these group attitudes fell directly on the group leaders. The superior leaders had the complete respect of those under them so that these individuals had the fullest confidence in their decisions.

Administrative procedures which were found to be used effectively by superior commanders and flight surgeons in maintaining the efficiency of their units included:

1. Proper provisions for rest and recreation.
2. Development of a belief in the immediate and ultimate objectives of the group and the effectiveness of the operations.
3. Effective use of the strong leaders among the air-crew personnel.
4. Provision of an opportunity to deemphasize a vivid emotional experience by talking it out under conditions which tend to make the experience seem more commonplace and natural.
5. Shifting individuals who had developed strong emotional reactions in specific situations to duties in a different environment involving flying if possible.
6. Making awards on the basis of the effectiveness of the operations of the unit rather than on an automatic basis.

## Character and Temperament

The discussions and observations of this survey indicated that character and temperament were of great importance in determining how the individual would adjust to the combat situation. A person of good character is one who identifies himself as a member of the group and accepts the general attitudes and responsibilities of group members. He is a team player. A person lacking in character does not identify himself with the group and refuses to adopt the group point of view. He is an individualist. In very simple terms, character is merely a matter of whether the individual puts the welfare of the group or his own personal welfare first.

Although the origins of the quality of an individual's character were not clearly defined, observations in these groups suggested that it was not easily altered by the time the individual attained military age. It was observed that the concept of membership in a social group was fundamental and applied for a wide variety of types of groups. Individuals appeared to vary considerably in the nature of the groups to which they had an intimate sense of belonging. The crew, the squadron, the group, the air force, the Allied military forces, and the peoples of the Allied nations had various degrees of identification for various individuals. It appeared to take a good deal of imagination and perspective to feel any close attachment to the last group mentioned and only a minority of the air-crew members appeared to achieve a genuine feeling of close affiliation with these larger groups.

The ties to the larger groups appeared to be those which led men and women who did not expect to be drafted to volunteer for service and it was known that these numbers were relatively small. Fortunately, loyalty to the crew or squadron of which the individual was a member appeared to provide practically as effective a basis for getting the job done, if the characters of the men were good and their leader had developed appropriate attitudes in the men, as the higher loyalties to which were attached the ideals and principles for which the war was being fought.

It appeared that in addition to character the other fundamental way in which people differ, which kept some fighting while others stopped or became ineffective, was temperament. The principal aspect of temperament in this regard was the extent to which the individual had a fundamental predisposition to develop anxiety when under stress. Anxiety as used here refers to the tendency to become agitated over a situation involving personal danger or discomfort. A chronic reaction of anxiety to a given situation was found to be fairly specific to the situation in which it was originally acquired (or in which the danger was originally perceived). Anxiety appeared to be acquired in much the same way as are such things as specific reactions of attraction and repugnance by associating the place and circumstance of a vivid emotional experience with that experience. Thus, a bombardier who turned around after completing his bombing run to see his navigator breathe his last gasp was reported to show signs of emotional agitation only on the bombing run and to appear to be perfectly cool and collected at all other times. In another case a copilot who showed anxiety in the cockpit due to previous experiences seemed very calm and efficient when he rode as a tail gunner.

On the other hand, there is also a tendency for anxiety to generalize or spread to a lot of other superficially similar situations. One of the most commonly observed examples of this is the tendency for anxiety, which is produced in a situation in which there is a lot of gunfire, to spread so that all kinds of loud noises invoke anxiety symptoms.

Severe anxiety was observed to be incapacitating in its effect, and persons who had developed a substantial amount of anxiety were usually not only of little assistance to the rest of the group, but actually were definitely detrimental on a mission. Although practically all air-crew members showed some signs of being nervous and "keyed up" while on operational duty, only about 3 or 4 percent of the air-crew personnel in these theaters broke down, while on combat duty in a group, to the extent that they were sent before a Central Medical Board.

The typical situations which tended to produce anxiety in these men were those involving fear of all types. In addition to fear of injury to themselves, there was fear of injury to friends and fear of failing to do their duties effectively. The tendency of the situation to produce anxiety is directly related to the individual's estimate of the likelihood that the feared event will actually occur. It is also a direct function of the evaluation placed by the individual on whatever is being endangered. As soon as the individual's subjective estimate of the likelihood of a feared event's happening gets large it is necessary for the individual to devalue the relative importance of the thing threatened if he is to prevent an excessive amount of anxiety from developing. In combat operations such as those visited it is essential that each individual be able to devalue the lives of all of the individuals in the group, including himself, in relation to the importance of the work being done if losses are to be prevented from causing excessive anxiety. If the results being obtained are believed to be poor the situation becomes very difficult to rationalize.

All of the factors concerning the individual such as his fatigue, attitudes, estimate of the danger, evaluation of the mission, and previous anxiety experiences tend to affect the individual to a greater or less degree according to the extent of predisposition toward anxiety which is characteristic of his temperament.

## Other Factors Affecting Morale

It was observed that one of the principal sources of motivation in keeping these air-crew men flying and fighting was the desire to complete their full share of the air forces' combat missions and return to the United States. Several factors tended to be detrimental to their attitudes toward combat. One of these was the reports from back home that indicated there were many people who were not making any real contribution to the war effort. Another was doubts concerning the interest in their welfare and the competence of officers in higher headquarters. The third was the persistent report that officers returning to the United States with combat experience were not being adequately utilized. They reported that letters from returned air-crew officers stated that they were meeting a very cool reception in the training stations in the United States. The attitude seemed to be, "Yes, we know you're a hero, but please don't bother us. We have all the answers and we'd like to have you take over a special job that won't interfere with what we're doing."

Although as stated previously, a very small percentage of air-crew personnel in these theaters were sent before the Central Medical Boards, the policies of these boards concerning diagnosis and disposition had a very important effect on the attitudes of air-crew personnel. The report of this survey contained detailed discussions of the procedures used in the theaters on disposition of breakdown cases and on factors related to length of combat tours and utilization in the United States after completion of the first tour of combat duty.

This report was reproduced and distributed to all training and personnel policy-making groups in Headquarters Army Air Forces and in the major air forces and commands, along with the following recommendations: (1) "It is recommended that a simple uniform policy for the diagnosis and disposition of breakdown cases in air-crew personnel be formulated on the basis of experience in the various theaters and published for use in all theaters," (2) "It is recommended that a definite uniform policy on length of combat tour and the rotation of personnel with combat experience to the Zone of the Interior be issued to all theaters. This policy should be based on the need for personnel with combat experience in the Zone of the Interior, and experience in the various theaters concerning the effect on air-crew personnel of various types of combat operations against enemy opposition of various amounts and qualities under the types of conditions prevalent in a particular theater," and (3) "It is recommended that the psychological research personnel in the AAF Redistribution Center direct its efforts primarily toward the problem of determining the type of duty in which the returning air-crew personnel with combat experience can make the most effective contribution to the general war effort."

This preliminary survey was followed by more detailed investigations of individual reactions to combat which were planned and carried out by the aviation psychology group in Headquarters AAF Personnel Distribution Command and subordinate stations. These detailed studies are reported in the sections which follow.

## A STUDY REGARDING FEAR AND COURAGE IN AERIAL COMBAT[1]

Early in 1944 a questionnaire was developed in the Psychological Division of Headquarters AAF Personnel Distribution Command regarding the frequency, symptoms, causes, and effects of fear in aerial combat and its control. This questionnaire utilized techniques suggested by the work of Dollard[2] who interrogated 400 American veterans of the Spanish Civil War on their fears as ground combat troops.

The preliminary form of the interrogation was given to 102 officers and 103 enlisted men who had served in air-crew positions in combat and were being processed through redistribution stations in March and April of 1944. Interviews with a number of these individuals and an analysis of the results from the preliminary administration provided the basis for the revised questionnaire which was administered between 15 May 1944, and 22 July 1944, at all three of the redistribution stations of the AAF Personnel Distribution Command.

The final questionnaire consisted of 159 multiple-choice questions of which the first 23 were background questions of rank, age, duty, and facts about the combat tour. At this time 1,985 officers and 2,519 enlisted men who had participated in aerial combat completed the questionnaire. Approximately three-quarters of these groups were from the European and Mediterranean Theaters and the remaining quarter were from the Pacific and Asiatic Theaters.

The reports of this group indicated the severity of the combat in which they had been participating. Although nearly 300 of the total group served in single-place fighter planes, approximately half reported that individuals flying with them in the same plane were killed or wounded. About one-fifth of the group had been wounded in action and a similar number had been injured in connection with aircraft. More than a third of them reported they had lost more than 10 pounds in weight during their combat tours.

[1]The studies reported in this section were primarily the work of L. F. Shaffer, N. E. Miller, and R. Pearson.

[2]John Dollard *Fear in Battle*. New Haven: Yale University Press, 1943.

Of the officers, 61 percent were pilots and the remainder were approximately equally divided between navigators and bombardiers. About 95 percent of the officers were lieutenants and captains, with the majority being first lieutenants. Close to 50 percent were married. The median age was 24 years. Of these officers 60 percent were in heavy-bombardment aircraft, 21 percent in medium- and light-bombardment planes, 16 percent in fighter planes, and 3 percent in other types of aircraft. The median number of missions flown was 35.

All but 2 percent of the enlisted air-crew members were gunners. This 2 percent included radio operators and photographers. Only 11 percent were gunners without another specialty and the other 87 percent were also mechanics, armorers, or radio operators. Almost all of these men were staff sergeants or technical sergeants, less than 4 percent being below these grades. The median age was also 24 years and 33 percent were married. Heavy-bombardment aircraft accounted for 83 percent and the remainder flew in medium or light bombers.

Some of the most significant results of the study are summarized below. In general, the officers and enlisted men gave similar answers, significant differences being exceptional. In one or two instances where the differences appeared negligible the figures for officers and enlisted men have been combined. The principal generalizations and the data supporting them were as follows:

1. *Almost all of these flyers reported they were afraid on combat missions.*

| | O | EM |
|---|---|---|
| | *Percent* | *Percent* |
| Were afraid on first mission | 85 | 83 |
| Were never afraid | 1 | 1 |
| Were afraid 1 to 3 times | 18 | 15 |
| Were afraid one-quarter to three-quarters of the time | 48 | 42 |
| Were afraid every or almost every mission | 33 | 42 |

2. *They reported that the fear they experienced in aerial combat was very strong.*

The officers and men were asked to compare the fears experienced in combat with any other fears they had had in their lives. Their answers were:

| | O | EM |
|---|---|---|
| | *Percent* | *Percent* |
| Very much stronger than any other fear I ever had | 37 | 44 |
| Somewhat stronger than any other fear | 17 | 24 |
| About the same as the strongest fear I ever felt in another situation | 24 | 21 |
| Weaker than some other fears | 14 | 10 |
| Never experienced any fear in combat | 1 | 1 |
| No answer | 7 | 0 |

3. *This group indicated that combat experiences frequently had noticeable effects on them during the missions.*

The officers and men reported that they felt these signs of fear "often" or "sometimes," while in the air, flying combat missions. The typical physio-

703323—47—18

logical signs of emotion predominated, but substantial numbers reported more exclusively psychological symptoms such as a feeling of unreality, irritability, or an inability to concentrate. . . .

| | Percent feeling it often | | Percent feeling it sometimes | |
|---|---|---|---|---|
| | O | EM | O | EM |
| A pounding heart and rapid pulse | 28 | 32 | 85 | 86 |
| Feeling that your muscles are very tense | 24 | 35 | 80 | 85 |
| Being easily irritated, angry, or "sore" | 20 | 33 | 81 | 80 |
| Dryness of the throat or mouth | 31 | 39 | 81 | 79 |
| "Nervous perspiration" or "cold sweat" | 23 | 28 | 79 | 79 |
| "Butterflies" in the stomach | 24 | 32 | 78 | 74 |
| Feeling of unreality; that this couldn't be happening to you | 15 | 24 | 65 | 73 |
| Having to urinate (pass water) very frequently | 22 | 28 | 60 | 70 |
| Trembling | 7 | 15 | 56 | 72 |
| Feeling confused or "rattled" | 2 | 3 | [illegible] | [illegible] |
| Feeling weak or faint | 2 | 5 | 33 | 49 |
| Right after mission, not being able to remember details of what happened | 4 | 6 | 37 | 40 |
| Feeling sick to the stomach | 3 | 6 | 39 | 46 |
| Not being able to concentrate | 2 | 3 | 37 | 34 |
| Wetting or soiling your pants | 1 | 1 | 7 | 4 |

*4. They also reported that combat experiences had noticeable effects after the mission was over.*

Some effects of fear do not occur when the danger is present, but may be felt later. The group reported the following symptoms before or after missions. They are almost identical with those that when persisting are diagnosed and treated as operational fatigue or combat strain.

| | Often | | Sometimes | |
|---|---|---|---|---|
| | O | EM | O | EM |
| | Percent | Percent | Percent | Percent |
| Feeling all tired out | 36 | 42 | 92 | 92 |
| Feeling restless, not being able to sit still | 34 | 42 | 88 | 90 |
| Feeling generally "blue" and depressed | 12 | 17 | 79 | 81 |
| Being "jumpy" at loud or sudden sounds | 21 | 31 | 71 | 80 |
| Not getting along with people, getting angry or grouchy too easily | 8 | 13 | 63 | 67 |
| Loss of appetite | 19 | 24 | 42 | 64 |
| Not wanting to fly the next mission | 11 | 13 | 66 | 59 |
| Being bothered by thoughts of combat coming up when you don't want them to | 11 | 15 | 57 | 65 |
| Bad dreams | 9 | 17 | 53 | 62 |
| Having shaky hands or knees, or muscular twitches | 7 | 14 | 48 | 64 |
| Suddenly feeling afraid without any special reason for it | 6 | 11 | 49 | 60 |
| Wanting company, can't stand it to be alone | 9 | 12 | 48 | 55 |
| Not wanting to be with people, wanting to be left alone | 9 | 14 | 47 | 53 |
| Having to urinate (pass water) too frequently | 10 | 15 | 43 | 51 |
| Feeling more afraid of other things, such as of riding in an automobile | 10 | 14 | 45 | 43 |
| Hating the sight of an airplane | 5 | 8 | 27 | 37 |

*5. Fear was reported as strongest when approaching and over the target on a specific mission.*

With regard to specific missions, the officers divided into those who felt their strongest fear just before taking off (19%) and those who were most afraid at or near the objective (48%). The enlisted men showed a different pattern, with fear increasing at each phase of the mission, to a peak near or at the target (54%).

A feeling of relief was reported after leaving the objective (33%), on reaching friendly territory (35%), and on arriving at one's base (18%).

6. ***Apprehension regarding the next mission usually increased with later missions.***

The last missions were reported as more feared (56%) than the first missions (15%), or the middle missions of the tour (14%). The others said that it made no difference.

| | *O* | *EM* |
|---|---|---|
| | *Percent* | *Percent* |
| Became more afraid as more missions flown | 41 | 57 |
| Became less afraid | 28 | 20 |
| There was no difference | 31 | 23 |

A few had their *first* qualm of combat fear before reaching the theater; more (19%) had it before their first mission; but most men had their first fear when on a mission (59%).

If a mission was tough, more men were afraid (46%) when they knew it would be tough before they started, than when they were surprised by unexpected opposition (21%). To the others it made no difference.

Flyers reported that they had more fear (43%) when they knew well in advance that a mission was scheduled, than when they were told suddenly to fly a mission (18%).

The majority (O's 64%, EM 53%) said that it was more fearful to go on a mission than to sweat it out waiting for friends. Eighteen percent of each said that it made no difference and 17 percent of the officers and 28 percent of the enlisted men said they had more fear waiting at the base for friends.

7. *These flyers reported their strongest fear on the first mission was concerning their ability to carry through the job, but on later missions this was surpassed by fear of being killed or wounded.*

| | On the *first* mission | On *later* missions |
|---|---|---|
| | *O* | *EM* |
| | *Percent* | *Percent* |
| Fear of being killed | 14 | 35 |
| Fear of being wounded, crippled, or disfigured | 11 | 25 |
| Fear of failure — not being able to do the job | 32 | 11 |
| Fear of being captured by the enemy | 5 | 14 |
| Fear of being a coward — that "they couldn't take it" | 17 | 6 |
| None of these | 21 | 9 |

8. *These men reported that they were not more afraid of wounds in one part of the body than another.*

Of these groups, 50 percent say that they had no more fear of one wound than another. Of specified wounds, eyes (22%), abdomen (9%), brain (6%), and genitals (5%) rank in that order.

9. *They reported that they feared flak more than enemy planes and that weather and malfunctions were of less concern.*

In response to a question regarding the relative concern they replied:

| | O | EM |
|---|---|---|
| | *Percent* | *Percent* |
| Feared flak most | 55 | 75 |
| Feared planes most | 34 | 19 |
| No difference | 11 | 6 |

Of the group 59 percent stated that they feared enemy action most, against 18 percent for weather, and 12 percent for malfunction.

10. *These flyers stated that having their plane on fire was a powerful stimulus for fear.*

More feared having the plane on fire (71%), than being slightly wounded (2%), having an engine go out (7%), or having the controls damaged (6%).

11. *These groups reported that fear sometimes improved their efficiency.* Mild fear was stated to increase the efficiency of many (50%), making them more accurate in their work, and to decrease the efficiency of only 9 percent. Even strong fear increased the efficiency of 37 percent, and had an adverse effect on 29 percent. Most men reported that they fought harder and worked better after overcoming a strong fear (O's 56%, EM 64%). Relatively few were so much affected by fear that they "lost their heads" (Officers, 80 percent never, 13 percent once, 7 percent a few times; EM, 73 percent never, 17 percent once, 10 percent more than once).

12. *Fear was reported to have been especially increased by situations where no counteraction could be taken and by lack of definite information as well as by seeing an actual source of danger.*

Factors that were reported to have *increased* fear to the greatest extent are given below. Being in danger when one can't fight back, being idle, or being insecure of the future, are notable.

| *Factors increasing fear* | O | EM |
|---|---|---|
| | *Percent* | *Percent* |
| Being fired on when you have no chance to shoot back | 81 | 88 |
| Someone reporting an enemy plane that you can't see | 79 | 81 |
| Seeing enemy tracers | 76 | 73 |
| Having something go wrong with your oxygen equipment | 62 | 74 |
| Feeling that you have been in so long that the law of averages is bound to catch up with you | 47 | 65 |
| Not knowing how many missions you will have to fly before being sent home | 47 | 65 |
| Having to be idle or inactive while flying a combat mission | 60 | 64 |
| Before your first mission hearing reports about the dangers and losses on missions | 57 | 55 |
| Knowing that you have to achieve an objective "at all costs" | 49 | 51 |

13. *The officers and men were sympathetic rather than reproachful toward the few individuals who "couldn't take it."*

About one-third of the group "thought of" asking to be grounded because of fear (O's 27%, EM 42%), but only a few did ask to be grounded for fear of flying (O's 3%, EM4%).

Combat flyers had lenient attitudes toward the man who quit because of fear. The majority "would not judge him as it might happen to anyone." Officers were a little less lenient than enlisted men.

| | If he quits after a *few* missions | | If he quits after *many* missions | |
|---|---|---|---|---|
| | *O* | *EM* | *O* | *EM* |
| | *Percent* | *Percent* | *Percent* | *Percent* |
| He is a traitor and should be shot | 0 | 0 | 2 | 1 |
| He is a coward and should be punished severely | 3 | 4 | 2 | 1 |
| He is a weakling who should not have been in air crew | 30 | 17 | 4 | 3 |
| One should feel sorry for him | 7 | 5 | 12 | 13 |
| One should not judge him, as it might happen to anyone | 52 | 60 | 60 | 61 |
| One should admire him, as it takes more guts to quit than to go on | 5 | 14 | 16 | 20 |
| No answer | 3 | 0 | 4 | 1 |

14. *Confidence, morale, effective activity, and an example of composure and confidence by others were regarded as effective in reducing fear.*

Factors that were reported as most effective in reducing fear are given in order below. Having confidence, keeping busy, taking action to meet dangers, and seeing others calm, are outstanding in promoting coolness under fire.

| *Factors reducing fear* | *O* | *EM* |
|---|---|---|
| | *Percent* | *Percent* |
| Having confidence in your equipment | 95 | 91 |
| Having confidence in your crew | 93 | 91 |
| Having confidence in the technical ability of your immediate superior or commanding officer | 87 | 88 |
| Keeping busy all the time while in the air | 90 | 85 |
| Seeing or hearing other men acting calmly in dangerous situations | 82 | 79 |
| Concentrating on the job you have to do | 74 | 77 |
| Knowing that you will be sent back home after completing a definite number of missions | 72 | 74 |
| Taking evasive action | 74 | 70 |
| Having a CO who does everything possible to look out for the food, shelter and comforts of his outfit | 48 | 70 |
| Shooting at an enemy plane | 60 | 77 |
| Hearing others make wise-cracks on the interphone | 63 | 68 |
| Cracking jokes on the interphone | 60 | 71 |
| Making a good hit on the target | 61 | 65 |
| Telling yourself to calm down | 63 | 62 |
| Praying | 58 | 67 |
| Feeling that you were lucky and would not get hit | 60 | 60 |
| Seeing your own tracers | 52 | 64 |

The following events were reported as not having had a significant effect on fear; they neither increased nor decreased it by any appreciable amount:

Having visiting generals fly with your outfit; getting commendation or citations; having losses replaced promptly; singing, shouting, or cussing; having a clear idea of what we were fighting for in the war; knowing how important your objective was; hearing conflicting rumors about enemy strength and our losses.

15. *Reading letters from home and other activities which entertained them and helped them think about matters unrelated to combat were reported as the best types of relaxation.*

In seeking relief from tension between missions, "calming their nerves," airmen reported greatest value for these activities, in the order given: letters from home, getting away, sleep, music, and the movies. At the bottom of the list, as least favored, are athletics, work, hobbies, and radio programs other than music.

The percents show those who indicated either first, second, or third choice among the nineteen alternatives.

| | O | EM |
|---|---|---|
| | Percent | Percent |
| Reading letters from home | 37 | 33 |
| Getting away from the field for a change of scenery | 24 | 27 |
| Sleeping | 26 | 22 |
| Listening to music | 25 | 22 |
| Seeing movies or a show | 22 | 21 |
| Playing cards or shooting craps | 15 | 22 |
| Having a drink | 20 | 30 |
| Joking about your combat experiences | 21 | 17 |
| Reading a book | 17 | 15 |
| Writing letters | 14 | 18 |
| Talking with friends, not about the war | 16 | 14 |
| Praying, reading the Bible, or going to religious services | 11 | 15 |
| Getting plastered | 11 | 13 |
| Having a date with a girl (Some of the Pacific returnees wrote in "What girl?") | 10 | 12 |
| Athletics | 13 | 6 |
| Talking about your combat experiences | 9 | 5 |
| Helping the crews on the line | 4 | 8 |
| Working on a hobby | 4 | 3 |
| Listening to radio programs other than music | 1 | 4 |

Talking about their fears at the present time makes no difference to most (53%), makes many feel better (33%), but makes a few feel worse (14%).

This group stated that when on a mission, men should not talk about their fears (68%).

However, between missions, they said it was best to admit one's fears and talk about them (59%).

The group stated that a policy of having a definite number of missions in a combat tour would help. The question and responses were as follows:

Which helps most to reduce fears in combat?

| | O | EM |
|---|---|---|
| | Percent | Percent |
| Having each man fly a definite number of missions or hours | 72 | 75 |
| Having each man fly until, in the judgment of his commander and flight surgeon, he begins to show signs of combat strain | 17 | 17 |
| It makes no difference | 7 | 7 |
| No answer | 4 | 1 |

16. *The three reasons reported by these men as of most importance in keeping them going suggest that they were principally concerned with doing their share in the group effort.*

The question was asked, "What are the three strongest reasons that carried you through combat missions?" The total list, with the percent who indicated each alternative as either first, second or third choice, follows:

| | O | EM |
|---|---|---|
| | Percent | Percent |
| Because it is my duty to my country | 60 | 57 |
| Not to let my outfit or crew down | 54 | 50 |
| To complete my quota of missions or hours and return to the United States | 45 | 45 |
| To show that I did my part in the war | 35 | 38 |
| To make the world a better place to live in | 22 | 27 |
| For the love of flying | 24 | 19 |
| So that my family and loved ones will approve | 18 | 15 |
| To get revenge for friends who have been killed | 10 | 18 |
| So that I wouldn't be considered a coward | 13 | 13 |
| Because I hate the enemy so much | 4 | 8 |
| To keep my flying pay | 2 | 6 |
| To get commendations or citations | 3 | 2 |
| To be able to get a better job after the war | 2 | 2 |

In the above list, it is of interest to see that certain motives were reported as very weak, including hatred, citations, pay, and self-advancement.

To supplement the data from the 1944 study, two additional samples answered part of the questions of the original questionnaire. The first group included 700 officers and 880 enlisted men who filled out the blank in redistribution stations between 13 March and 14 April 1945. The second sample was collected shortly after the victory in Europe and included 1,088 officers and 1,181 enlisted men who responded to the questionnaire between 14 May and 16 June 1945. These groups of flying personnel differed little from the 1944 group in age, rank, marital status, and similar data. Their combat experiences were a little less severe as indicated by fewer wounded, fewer who lost many friends, a smaller average loss of body weight reported, and fewer who reported enemy opposition on every mission.

The 1945 groups reported a frequency of fear equal to or greater than the 1944 group, although other evidence indicated that combat was less dangerous in the later period. They also reported that their fears were stronger in comparison with other fears than did the 1944 group. This emphasizes the fact that fear is a relative matter having a rather complex basis. With respect to most of the other items, the 1945 group gave very similar responses to those obtained from the 1944 group. The findings of this study on fear and courage had practical implications for the training of air crews and for the utilization of personnel in combat operations. The results of this study were made available to both policy-making and training and operating groups in the Army Air Forces in order that their procedures could be based on a maximum of information regarding this situation.

This study indicated that the most important factors in reducing fear are confidence in equipment, in fellow crew members, and in leaders. Although good equipment is a material problem, confidence in this equipment is psychological. It was therefore recommended that deliberate training procedures be set up to create confidence. The development of confidence in both fellow-crew members and in leaders is the responsibility of the leaders. It was therefore recommended that officers be given a more effective training in principles of sound leadership. It was also recommended that the principles of "activity" and a "composed and confident example" should be maximally utilized in indoctrinating new crews into combat operations. An effective procedure used by some combat groups was to send a person with a considerable amount of combat experience who could be depended upon to provide a good example along with a new crew on their first combat mission. Other implications of this important study are discussed in subsequent sections.

## RESEARCH ON ANXIETY REACTION TO COMBAT STRESS*

### Introduction

As indicated in the previous section, practically all of the men who participated for a substantial period of time in aerial combat reported evidence

*The studies described in this section were carried out by L. F. Shaffer, G. Forlano, C. W. Crannell, F. K. McDowell, W. G. Mollenkopf, A. S. Switzer, F. Wickert, H. E. Kirkman, D. E. Super, M. F. Ross, L. Smile, G. L. Heathers, W. S. Gressey, J. J. Johnston, L. L. Martin, A. S. Levine, B. Morton, D. D. Raylesberg, S. W. Bijou, W. H. Lucio, R. M. Rust, J. P. Madeira, J. B. Beck, P. W. McReynolds, L. J. Hibbs, F. Kitch, G. J. Wischner, J. B. Rotter, R. D. Gillman, A. L. Irion, G. R. Pascal, and B. Willerman.

of fatigue, restlessness, and depression following these intense emotional experiences which were typical of the combat missions. Since this group was selected as being fairly normal, these findings suggest that excessive emotional reactions arising from frustration and conflict in any life situation can be expected to induce an anxiety in which the individual shows the usual symptoms of the psychoneurotic.

A primary function of the redistribution stations was to determine the fitness of returnees for immediate duty in the United States. This decision was the responsibility of the medical division and was based on a physical and a psychiatric examination. Approximately 5 to 10 percent of flying officers and 15 to 20 percent of enlisted flying personnel were found on psychiatric examination to exhibit a fairly well-defined pattern of psychological disturbances. This pattern was first called "operational fatigue" and later referred to as "anxiety reaction." Typical symptoms included tension, rapid heart rate, sweating, irregular sleep, vivid battle dreams, loss of appetite, loss of weight, irritability, inability to concentrate, depression, and low motivation.

Those individuals who were diagnosed by the psychiatrists as showing symptoms of anxiety reactions were sent to AAF convalescent hospitals with the expectation that a few weeks of treatment would be sufficient to enable these individuals to qualify for immediate active duty. Unfortunately, the heavy load at the typical redistribution station prevented the psychiatric interview from lasting more than 5 to 10 minutes in most instances.

Although this short time seriously limited the psychiatrist's investigation, it was believed that the typical interview which usually included inquiries as to sleeping habits, including insomnia and battle dreams, and into eating habits, motivation for flying, anxiety over home conditions or postwar plans, and general attitude of the patient were reasonably effective. The diagnoses were derived as much from the visual observation of symptoms such as flushing, sweating and tension, and interpretation of the patient's manner of response, as from the contents of the verbal reply. As previously mentioned, the purpose of this screening was to select those individuals who had acquired somewhat stronger symptoms than the average returnee, and who it was believed would benefit from a few weeks in an AAF convalescent hospital.

Using the diagnoses of the psychiatrists regarding the presence of these symptoms, a series of research studies were carried on by the aviation psychologists in the redistribution stations of the AAF Personnel Distribution Command.

## Studies of Air-Crew Classification Test Scores

One group of studies attempted to discover tests which might be used at the time of original selection. The scores made at the time of original testing in the AAF classification centers on the air-crew classification tests were compared for: approximately 600 normal bombardiers and 300 bombardiers diagnosed as having anxiety-reaction symptoms; samples of returned navigators including about the same number of individuals of each type; approximately 500 normal bomber pilots and 200 sent to AAF convalescent hospitals by psychiatrists; and for a small group of 67 fighter pilots diagnosed as normal, and 25 diagnosed as anxiety-reaction cases.

No clear-cut pattern emerged from these analyses although there is a scattering of coefficients significant at both the 5-percent and 1-percent levels of significance. Some of these coefficients are opposite in sign to those which might have been expected. Differences between the means of the groups significant at the 1-percent level are found for the Mathematics and Reasoning Test for both bombardiers and navigators. For the bombardiers, differences large enough to reach this level of significance were also found for the Mechanical Information, Mechanical Principles, and Reading Comprehension Tests.

A difference for the Numerical Operations Test favoring groups with low scores on this test, significant at the 1-percent level, was also found for bombardiers. It is of interest to note that this same test showed lower scores for the normal group for an additional sample of bombardiers, one of navigators, and one of pilots, each of which was sufficiently large to reach the 5-percent level of significance. These results are especially interesting in view of the results reported on the negative correlations of this test with combat criteria in chapter 4.

The only other difference reaching this level of significance which was based on a substantial number of cases was found on the Finger Dexterity Test. For one sample of bomber pilots, the normal group was found to have significantly lower scores than the group diagnosed as having anxiety reactions. However, the results obtained from three other samples of bomber pilots favored the normal group. It therefore seems probable that this, as well as some of the differences previously mentioned, were attributable to sampling fluctuations.

The findings with regard to the Aiming Stress Test and Steadiness Under Pressure Test were of special interest because in devising these tests it had been hoped that they might be predictive of ability to remain cool and composed in the combat situation. The differences found for the two groups on these tests were negligible in size. It had been previously established that these tests had no predictive value for primary pilot training. On the basis of this earlier finding regarding these tests lack of validity for primary pilot training, they had been eliminated from the battery.

### Studies of New Tests

Unfortunately, the individuals on which these analyses were based were tested with the early batteries of the air-crew classification tests so that practically none of them had taken either the General Information Test or the Biographical Data Blank. These tests were regarded as the most promising in the Air-Crew Classification Test Battery for measuring personality and temperament factors and it is therefore hoped that later analyses of records on these tests in relation to combat can be obtained.

. an effort to develop forms of these two tests which would be more directly applicable to the combat problem, items from various experimental tests were selected and a number of new items written to comprise special forms of the Biographical Data Blank and the General Information Tests. These tests were completed in July 1945 and were administered to a large number of returnees in the summer of 1945. Unfortunately, the percent of returnees being diagnosed as having anxiety reactions became insignificantly small in the redistribution stations at about this time. It may still be possible to carry out some analyses, using these tests if a substitute criterion such as the individual's own report of his symptoms can be obtained from the records.

Using data from the regular air-crew classification test form of the Biographical Data Blank and also including data from the new blank for those items which were common to both forms, a preliminary analysis was made for 100 normal and 50 anxiety reaction cases. Although the item analysis revealed a few differences of moderate size, especially for certain questions regarding participation in athletics, the sample was much too small to provide stable findings.

Exploratory studies involving approximately 100 cases each were carried out for 3 types of tests which are of considerable interest. The studies of a free-response group-form of the Rorschach Test in which standard Rorschach plates were projected on the screen and the returnees wrote what they saw, 2 minutes being allowed for each slide, failed to discriminate between the two groups of normal officers and officers with anxiety reactions. Furthermore, an item analysis of the responses from the normal and anxiety-reaction cases reveals no consistent trend on which a new interpretation of the test could be based.

A study of the performance of psychiatric groups on the conventional Rorschach tests was carried out at the AAF Convalescent Hospital at Fort George Wright. In this study 3 groups of 84, 60, and 36 patients respectively, the first two of which had diagnoses of severe anxiety reaction and the last, other types of psychiatric diagnoses, were compared with norms provided for the test by Klopfer and Kelley[4]. These scores were also compared with the results from testing approximately 300 unselected aviation students who were later sent into pilot training. This latter group was tested in the summer of 1943 at Psychological Research Unit No. 1, Nashville.

The test results for the convalescent patients would be interpreted as evidence of maladjustment when judged by the Klopfer-and-Kelley data for normal subjects. However, these samples of patients with psychiatric diagnoses have scores very similar to those of the aviation students. These aviation students judged by Klopfer-and-Kelley norms would also appear to be severely maladjusted, in spite of the fact that they had passed the physical and psychiatric examination to qualify them for air-crew training. On the basis of these findings, the most reasonable conclusion appears to be that

[4] B. Klopfer and D. M. Kelley. The Rorschach Technique. Yonkers on Hudson: World Book Co., 1942.

the separate Rorschach categories are not satisfactory indicators of maladjustment and that the Klopfer-and-Kelley norms are not representative of a properly selected normal group.

The Shipley-Hartford Retreat Scale for measuring intellectual impairment, consisting of a recognition-vocabulary test and a test of abstract problems failed to discriminate between 61 normal officer returnees and 33 officers with anxiety reactions. The biserial correlation coefficient found was 0.11, which was too small to be significant for a group of this size.

A similar study was carried out at the AAF Convalescent Hospital at Cochran Field. In this study it was found that the biserial correlations between the ratio score on the Shipley-Hartford Retreat Scale and the groups of 217 patients admitted for psychiatric reasons and 228 patients admitted for nonpsychiatric reasons was 0.13. This value is very similar to that found in the other study but because of the larger number of cases exceeded the 5-percent level of significance.

At the AAF Convalescent Hospital at Santa Ana a somewhat similar type of test called the Efficiency of Mental Application Test, DE602A, was developed. This test consisted of three complex nonlanguage tests including a maze test, a spatial-relations test, and a number-pattern test, and two vocabulary tests, one multiple choice and the other free answer. The scores in terms of number of correct responses on nonverbal tests were modified by applying an error-rate and a variability-rate correction in an effort to obtain a combined measure of speed, accuracy, and continuity of application. The final score on the complex test was divided by the vocabulary score and this ratio multiplied by 100 to yield the EMA score. A study of a sample of 150 convalescent patients indicated that the EMA score of a group of patients with psychiatric diagnoses was lower than the mean of the group with nonpsychiatric diagnoses. The difference was statistically significant at the 1-percent level of confidence. Thus this type of test tended to show a consistent, if small, degree of discrimination between psychiatric and normal groups.

The third area covered in these exploratory studies was the relation of anxiety reaction to the Flicker Fusion Test. In this preliminary tryout, a General Radio "strobotac" capable of flashing a light from 600 to 14,500 cycles per minute and increasing or decreasing the frequency until fusion or flicker was reported was used. A difference between 50 normals and 50 returnees diagnosed as having severe anxiety reactions which was significant at the 1-percent level was obtained. The positive findings from this group led to the development of plans for an experiment using a much larger number of cases. Unfortunately, this was not carried out because of the rapid changes in the program that occurred following the victory in Japan.

Studies were made to determine whether any of the tests used in the Instructor Aptitude Test Battery differentiated between normal and anxiety-reaction cases. For pilots, bombardiers, and navigators these comparisons were made using 100 normals and between 50 and 100 anxiety-reaction cases.

The only difference for any of the intellectual or achievement-type tests for these groups which reached the 5-percent level of significance was for the Arithmetic Reasoning Test with the navigator groups. The difference in scores favored the normal group.

For a large sample of approximately 500 normal gunners and 450 gunners diagnosed as having anxiety reactions, small differences were found for all of the aptitude and achievement tests given to gunners which, because of the size of the group, could not reasonably be attributed to chance. The differences found were of the order of about one-fifth of the standard deviation of the group of normals. For comparative purposes a part of the anxiety-reaction group was tested after from 2 to 6 weeks in the AAF convalescent hospitals. Scores of this group did not differ significantly from those of the anxiety-reaction cases tested in the redistribution stations.

A new type of projective test which provided considerable more structuring of the response situation and yielded promising results was the Incomplete Sentence Test, DE303A. This test was adapted from a test reported by N. Cameron. It consisted of a sheet containing 40 incomplete sentences, usually of two or three words such as "My greatest fear.....................," "I regret......................," "I am very..................... ," "The Army......................," and "Most girls .................. ." The individual was instructed to: "Complete these sentences to express your real feelings." The test required about 20 minutes and was given as a group test.

Two methods of scoring were used. In one, the psychologists examined all of the responses and rated the tests on a four-point scale indicating no, mild, moderate, and severe disturbance or maladjustment. The other method of scoring consisted of giving a value from +3 to —3 to each response by comparing it with scoring examples. Thus the range from healthy to maladjusted responses for the item, "Other people ............... ," is illustrated by: "are swell," "are O. K.," "get along with me," "are different," "have their worries too," "talk too much," and "laugh at me."

The consistency of rating for the clinical method was 0.68 and the agreement for the item-scoring method between two different scorers was 0.85 and 0.89 on two samples. These coefficients were based on 148, 200, and 50 cases each. For the first group, the clinical ratings of the test responses as a whole agreed with the physician's diagnosis of psychiatric or nonpsychiatric disturbance at the time of admittance to the extent of a biserial correlation coefficient of 0.40. The item-scoring method yielded a triserial correlation coefficient of 0.61 with the psychologist's evaluation of severity of disturbance during the initial interview for the sample of 200 cases.

### Personal Adjustment Inventories

A large number of studies were carried out in AAF Classification Centers, Convalescent Hospitals, and Redistribution Stations to determine the accuracy with which questionnaires of various types would predict the diagnoses of flight surgeons and pyschiatrists. The primary purpose of most of these studies was to develop an effective screening technique which would be useful in determining which cases should be given a thorough psychiatric interview during the physical examination. In the case of hospital patients, it was intended to assist the physician in providing information regarding his new patients. It was also hoped that such studies would contribute to an understanding of the nature of anxiety reaction and other psychiatric conditions. Such an inventory might also be useful as a partial check on the patient's readiness to leave the hospital if it were readministered at the time he was being considered for final disposition.

The form most extensively used in the Redistribution Stations and Convalescent Hospitals was the Personal Inventory, DE201. This inventory was based on the Personal Inventory constructed by Dr. Walter C. Shipley for the Office of Scientific Research and Development. The original 145-item form was edited to fit the returnee situation and reduced to 120 items in the fall of 1943 at Headquarters AAF Personnel Distribution Command. An abbreviated form containing 45 items was later prepared. A biserial correlation coefficient of 0.49 was obtained between the scores on the abbreviated form of the Personal Inventory and diagnosis of normal or anxiety reaction. This coefficient was obtained from 2 samples totaling 1,796 cases tested from January to March 1945 at the AAF Redistribution Station at Santa Ana. One hundred and twenty-four of these individuals were diagnosed as having anxiety reactions.

It should be noted that the method of estimating the biserial correlation coefficients on a sample that includes all cases, both normal and with anxiety reactions, tested during a specified interval of time provides a more acceptable procedure than calculating the coefficients from a sample composed of a specified number of anxiety-reaction cases and a partial sample of controls selected from all of the control cases available. The percentages in the two groups are relevant in this case because the assumption of normality for the total group is made. This assumption is of course unreasonable if the actual number of cases used in the 2 categories are approximately equal, but one represents only a small fraction such as 10 percent of the available cases for that group. This latter type of coefficient, however, has some meaning especially in terms of comparisons with similarly obtained coefficients for other tests. This limitation should, however, be kept in mind in interpreting such coefficients when reported.

The Personal Inventory used a forced-choice type of item. The individual was to select the one of two choices which most correctly described him. These items were such as the following: "I am more nervous ......................," "I am more easy-going ................. ..........."

Another type of inventory used was the Cornell Selectee Index, Form N.[8] This inventory consisted of a series of 64 questions dealing with typical symptoms of personality disturbances, physical complaints, and social mal-

[8]A. Weider, B. Mittelmann, D. Wechsler, and H. G. Wolf, The Cornell Selectee Index. *J. Amer. Med. Ass.*, 1944, *124*, 224–28.

adjustment such as, "Are you often misunderstood?", and "Do you have pain or pressure in the head?". The individual indicated presence or absence of the problem by circling a "yes" or "no."

Three inventories were especially constructed for use in convalescent hospitals. These included the Convalescent Personal Inventory which consisted of a comprehensive list of anxiety reaction symptoms obtained from the literature and from statements taken from patient interviews. This inventory was constructed at the AAF Convalescent Hospital at Miami Beach.

The Questionnaire Regarding Present Reactions and the Inventory of Psychological Problems were developed at the AAF Convalescent Hospitals at Fort Thomas and Bowman Field, respectively. The items for the first of these were developed after a survey of problems reported by patients during the initial evaluation. The items dealt with problems in the following general areas: nervousness, tension, worries and fears, sleeping and dreaming, physical complaints, memory and concentration, sociability, emotional expression and control, and feelings of well-being. The items were answered as either true or false with regard to the individual.

The Inventory of Psychological Problems was a revision of this questionnaire. It consisted of 7 items from each of 10 general areas. These included: sleep and dreams, nervousness, smoking and drinking, socialization, memory, concentration, worries and fears, physical health, depression, and decision and plans. The individual answered each item by circling a letter from A to E on a 5-point frequency scale running from "Never" to "Always." The second section of the inventory contained 10 items, each referring to 1 of the 10 general areas into which the specific items fell. The patient denoted the severity of his problems in each of the 10 areas by circling a number on a 1-to-9 rating scale on which 1 indicated "none" and 9 "very severe."

All of these various inventories contained items which required the patient to state the presence, frequency, or severity of a symptom or problem as it applied to him. It was found that fairly good predictions of the physician's diagnosis of psychiatric complaint could be based on these responses. This is of course not very surprising since the physician's diagnosis was to a large extent also based on the patient's report of his symptoms. In general, critical scores could be established on these inventories which would select between 60 and 70 percent of psychiatric patients and include only about 10 to 20 percent of nonpsychiatric patients. This level of differentation provided a useful preliminary screening.

Some information regarding the nature of anxiety reactions is provided by a comparison of the responses of a group of approximately 80 patients admitted to Bowman Field with a nonpsychiatric diagnosis with the responses of a group of approximately 45 patients who were admitted with psychiatric diagnoses. The largest differences in the ratings of the severity of problems are found for these groups on ratings on sleep and dreams, the next largest differences in order are for memory, depression, worries and fears, concentration, decisions and plans, nervousness, and smoking and drinking. Socialization has a smaller difference for these groups and physical health shows only a slight difference in the ratings for the two groups.

It would be very valuable from the point of view of a more adequate understanding of the causes of and the factors predisposing toward anxiety reactions if a number of items of a biographical or objective nature could be found which would discriminate between the two groups of normals and anxiety reaction cases. Although adequate exploration of this area was not carried out, the negative nature of most of the findings considered along with the overwhelming prevalence of the symptoms of anxiety reaction in combat returnees would suggest that the effect of the predisposing factors might have been slight in comparison with the effect of the combat stress to which the individual was exposed. The degree of combat stress must be interpreted in terms of the total situation in which it occurs as indicated by the responses of the returnees regarding factors which produced increases and decreases in the individual's anxiety.

It is believed that the findings from these studies should be of tremendous practical value in developing efficient mass-screening devices for use in connection with the physical examination for flight surgeons and psychiatrists. Further valuable information should be provided when the analyses of the comprehensive interview conducted by four specially trained flight surgeons is analyzed in comparison with the comprehensive inventory which was administered to the same group of several hundred fighter pilots. These data are in the files of the School of Aviation Medicine. These surveys also provide much basic information concerning the nature of the individual's reaction to combat which should be of great value in developing plans and procedures for possible future combat operations. The broader implications for the understanding of the psychoneurotic individual are obvious and should be of great value in this general field.

## RESEARCH RELATED TO COUNSELING AND THERAPY*

### General Studies

A number of studies of the attitudes, interests, and backgrounds of patients in Convalescent Hospitals were carried out. The studies of interests and attitudes were of considerable value in directing individual and group counseling sessions. In general, these surveys revealed that patients had favorable attitudes toward the Convalescent Hospital program and felt they were benefitting from hospitalization. Various surveys of patient interest were made for use by the Convalescent Services Division. These surveys permitted evaluation of the activity program by determining how well the curriculum corresponded to the dominant interests expressed by the patients. In this way it provided a basis for modifying the curriculum as well as the methods of instruction in the interests of greater effectiveness.

*The work described in this section was done by S. W. Bijou, D. H. Lawrence, A. S. Levine, D. D. Raylesberg, B. Morton, G. L. Heathers, D. E. Super, B. Willerman, G. R. Pascal, R. N. Hobbs, and A. L. Irion.

A sociological questionnaire was developed at the AAF Convalescent Hospital at St. Petersburg. It was found that for a group of 170 patients for whom the prognosis was indicated to be "excellent" or "good" as compared with approximately 70 patients for whom the prognoses were "fair" and "poor," differences in percentages responding to the various choices significant at the 5-percent level or better were obtained on 20 of the 146 items. Similarly, differences significant at the 5-percent level were found for 22 items between a group of approximately 80 diagnosed as "mild" as compared with approximately 180 diagnosed as "moderate" and "severe" anxiety-reaction patients. However, only three of the items were found significant at this level in both of the groups. This, together with the fact that the physician's diagnosis of severity and estimates of prognosis were based to some extent on information of these same types obtained directly from the patients, offered little encouragement for the development of a useful predictive instrument using this type of background data and this criterion.

## The Group Counseling Program

In the AAF convalescent hospitals a program of group counseling or personal-adjustment group conferences as they were called, was developed. Personal-adjustment group conferences were defined as a series of talks and discussions conducted to assist the patient to understand and revise unwholesome attitudes which generally accompanied the mental or physical illness responsible for hospitalization. These groups varied in size from 5 to 30 and were composed of all patients except those with severe emotional disturbances. They usually met twice a week for 3 or 4 weeks.

The principles underlying the planning of these conferences are listed below:

1. Discussion by and with the patient of his problem leads to an objectification and clarification of them and makes them more accessible to solution.

2. The verbal expression of emotion relieves tensions, to some extent, and prepares the way for rational control of emotional reactions.

3. The discovery that other patients with similar experiences have similar problems decreases the anxiety associated with being "mentally ill."

4. Identification with a group provides emotional support.

5. Intellectual understanding of psychological problems tends to decrease the overevaluation of their significance.

6. The assimilation of sound psychological knowledge tends to result in more rational solutions to problems.

The topics selected for inclusion in the group conferences included psychological principles involved in personal adjustment, topics relating to morale, and topics relating to return to duty or to civilian life. Although the contents of the conferences at various convalescent hospitals and within the same hospital varied a great deal, the objectives remained relatively constant. For this reason, a series of objectives was provided for the discussion leaders as follows:

1. Emphasize that behavior is the result of composite influences, especially the experiences of an individual.

2. Emphasize that behavior develops to satisfy certain needs of the individual.

3. Show that we are continually acquiring new habits in civilian and in military life to meet changing needs in changing situations.

4. Allay undue fears arising from misconceptions of the nature of emotional behavior, and stress the normalcy of emotional responses under varying conditions.

5. Demonstrate the constructive use of emotions in military life.

6. Stress the "naturalness" of having particular problems in certain situations.

7. Indicate proper ways to channel any "bitterness" or "grievances."

8. Present an understanding of the American civilians' contribution to the war effort, and try to reestablish, if necessary, identification with the nation.

9. Point out that confidence in old skills can be maintained since they are not lost through disuse and that new skills are relatively easy to learn.

10. Analyze some of the attitudes and movements that may arise and which may be detrimental to the nation and the war effort, such as unfounded suspicion of our Allies, unfounded gripes against civilians, and prejudice against racial and religious groups.

### Development of Objective Records for Group Counseling

In establishing this program it was found that the body of established information regarding group counseling techniques was extremely meager. Most of the discussions which could be found appeared to be based on pure speculation, or at best, observations of an informal and uncontrolled nature by those with experience in group-counseling work. It was therefore believed of great importance to develop a systematic, objective, and quantitative method of analyzing group-counseling behavior. Accordingly, a project was initiated at the AAF Convalescent Hospital at Cochran Field to explore the possibility of developing a technique for evaluating the procedures in this area.

The procedure tried out consisted of obtaining a verbatim account of each group-counseling meeting. Using the transcripts of these meetings, statements made by group leaders and patients were studied and categorized with respect to their psychological significance. The categories used for describing the group leaders' statements included: (1) directing, (2) accepting, (3) rejecting, (4) reflecting (recognizing or restating), (5) interpreting (explaining the dynamics in a specific situation), (6) explaining (general statements, definitions, and broad outlines), (7) continuing (mild encouragement to proceed), (8) reassuring (supporting and exonerating), (9) minor rapport-building (conversational).

In categorizing the statements of the patients, 6 positive and 10 negative categories were used. The positive categories included: (1) insight, (2)

703333—47—16

abreaction (experiences told with some emotion), (3) rapport (conversational or approving the leader), (4) acceptance of interpretation, (5) understanding, (6) reorientation (indicating new viewpoints).

The 10 negative categories included: (1) rationalization, (2) projection (assigning own thoughts and actions to others), (3) displacement (shifting emphasis to unimportant detail to avoid stress or unacceptable motives), (4) passive-dependence, (5) hostile-aggressive, (6) obsessive-compulsive, (7) depressive, (8) guilt, (9) anxiety, and (10) traumatic.

After some exploratory study of other methods, it was decided to weight statements according to the number of lines they occupied in the transcript. All statements were categorized independently by pairs of psychologists and final classifications were based on discussions and agreements between the two judges. Graphs and bar charts were prepared showing the behavior of the patients, individually and collectively, and the group-leaders' procedures by session and for all sessions in a given series.

These techniques were applied to a series of sessions at the AAF Convalescent Hospital at Cochran Field. The analysis appeared to be very revealing. For example, the leader talked for about two-thirds of the time during the first session and for about 80 percent of the time during the second session. However, in the third session he talked for only one-third of the time and in the remaining sessions the time was fairly equally divided between patients' statements and the leader's statements. In the first session, a large portion of the group leader's time was taken up with *explaining*, the next largest with *directing*, then came *reflecting* patients' feelings, reassuring, and interpreting. Negative factors were very prominent in the patients' responses during the first session.

During the second session most of the patients' responses were positive, the larger proportion of them indicating that rapport was increasing. During this session there were also some exchanges of battle experiences among the patients which were of an abreactive nature. The hostile-aggressive responses characteristic of the first session decreased very markedly. During the last two or three sessions more of the patients' responses showed insight, acceptance of interpretation, understanding, and reorientation.

## Conclusions

This exploratory work appears to have provided a basis for greatly improving available procedures for evaluating group-counseling work. These procedures were developed too late to use in evaluating the general group-counseling program of the AAF convalescent hospitals and it is therefore necessary to rely entirely on subjective reports. Since attendance at these personal-adjustment group conferences was voluntary, the fact that they were well attended at the convalescent hospitals suggests that the individuals found them of value. Surveys of patients' attitudes in certain of the hospitals also indicated a very favorable opinion regarding these conferences among the patients. It therefore appears likely that these conferences made an important contribution to the individual adjustments of these groups of men. It is also believed that some of the procedures and research techniques developed will be found of considerable values in similar situations.

## STUDIES OF THE ATTITUDES AND PREFERENCES OF COMBAT RETURNEES[1]

In the questionnaire regarding fear and courage in aerial combat, certain questions were included regarding returnee attitudes toward a second tour of combat. Preliminary analyses of these data indicated that a large proportion of combat returnees were not very interested in volunteering for a second tour of combat duty. In the summer of 1944, the Assistant Chief of Air Staff, Training, in Headquarters Army Air Forces needed information regarding the proportion of combat returnees who would be available for a second tour of combat duty in order to make training plans for the following year or two. Accordingly, personnel of the Aviation Psychology Program were requested through the Air Surgeon to make a study of returned combat personnel. At that time training officers were particularly interested in the problem of combat gunners, but the study was extended to include other specialties.

From the sample of 1,990 air-crew officers and 2,474 enlisted aerial gunners previously described in connection with the study of fear and courage, the first systematic tabulation of responses was obtained.

One question in the anonymous interrogation of fear and courage was, "What is your present attitude toward returning to combat duty?" This was followed by six alternatives, which are stated, below together with the percents of officers and of enlisted men giving each response.

| *Response* | *Officers* | *Enlisted* |
|---|---|---|
| | *Percent* | *Percent* |
| Eager to return to combat duty immediately | 4 | 3 |
| Plan to volunteer for combat duty within 6 months or so | 8 | 7 |
| Would have no objections to returning to combat after a reasonable rest (9 to 12 months) in the United States | 28 | 23 |
| Ready to return to combat after all trained men still in the United States have been sent. | 40 | 19 |
| Desire to continue noncombat flying but could not return to combat duty | 16 | 26 |
| Would prefer nonflying duty in this country | 4 | 22 |

It is clear that very few of the officers and men were eager to return to combat duty immediately and only very small proportions, 12 percent and 10 percent of the officers and enlisted men respectively, planned to volunteer for combat duty within 6 months or so. Twenty percent of the officers and almost half the enlisted men either didn't want to fly at all or said they could not return to combat duty.

To study the attitudes of returned personnel who had been assigned to duty in the four Continental Air Forces, a detachment of aviation psychologists from the AAF Training Command administered a questionnaire to 2,659 returned gunners in August and September 1944. At a later date

[1] The studies reported in this section were the work of L. F. Shaffer, R. Pearson, F. Wickert, D. E. Super, W. S. Gregory, C. W. Brown, and M. T. Hollinshead.

similar information was obtained from 528 returned flying officers. In this study the flying personnel were asked to put their names on the questionnaires.

This sample had been in the United States for a longer time, 6 months on the average, than those questioned at the Redistribution Stations. They were assigned to duties in this country, and perhaps had had a greater opportunity to readjust since return from combat. In age, rank, missions, theaters, and number wounded they did not differ significantly from the redistribution station samples.

Their attitudes toward returning to combat were reported as follows:

| | *Officers* | *Gunners* |
|---|---|---|
| | *Percent* | *Percent* |
| Desired to return | 29 | 15 |
| Indifferent toward returning | 31 | 16 |
| Desired not to return | 21 | 46 |
| Desired not to return "under any circumstances" | 28 | 22 |

Attitude toward volunteering:

| | *Officers* | *Gunners* |
|---|---|---|
| | *Percent* | *Percent* |
| Planned to volunteer immediately | 10 | 5 |
| Planned to volunteer after six months or less | 31 | 10 |
| Did not plan to volunteer | 58 | 81 |

Although the time interval was longer, it appears that a larger proportion of this group planned to volunteer than planned in the group of returnees questioned immediately following their return to the United States.

One additional item revealed an interesting attitude. The gunners were asked, "*If* you were given a choice between assignment to a second tour of duty as a gunner or assignment as a basic soldier, which would you prefer?" A second tour as a gunner was chosen by only 42 percent, against 58 percent for becoming a basic soldier.

Of the gunners in the continental air forces who were still on flying status, only 60 percent had a desire to continue to fly, while 40 percent were indifferent or opposed. It is likely that those who did not even want to fly were antagonistic to a second combat tour.

The readministration of the questionnaire on fear and courage in March and April 1945 and May and June of 1945 provided an opportunity to observe the trend in attitudes toward return to combat. The responses of the group in March and April were very similar to the group in May and June 1945. However, they both differed substantially from the samples obtained in the summer of 1944. The proportion of officers indicating that they preferred not to fly at all or who felt that they could not return to combat increased from 20 percent to approximately 40 percent, and the similar responses for gunners increased from approximately half to more than two-thirds of the group. The percentages of those eager to return or planning to volunteer decreased to 10 and 6 percent for the officers and enlisted men respectively.

Analyses were made to determine which factors were most highly associated with the attitudes of these officers and men regarding return to combat. These analyses indicated that the subjective reports of these men regarding the frequency and strength of their fear during their first combat tour was most closely related to willingness to return. Of officers who reported they were afraid 3 times or less, 90 percent showed favorable attitudes toward return, while of those afraid on every mission only 72 percent were favorable. The enlisted gunners showed similar relationships. An interesting finding is that flyers who were wounded or injured indicated that they were as willing to return to combat as those who had not been wounded or injured.

| | Percent favorable | |
|---|---|---|
| | Officers | Enlisted |
| Wounded | 77 | 52 |
| Not wounded | 81 | 52 |
| Injured | 80 | 49 |
| Not injured | 81 | 52 |

No relationship was found between attitudes toward return to combat and age among officers. The 1944 study indicated that the youngest gunners, those under 21 years of age, were most eager but this was not confirmed in the 1945 study. The 1944 and 1945 studies indicated that pilots were somewhat more willing to return than bombardiers or navigators and that those who flew in fighter aircraft were markedly more willing to volunteer than crews of bombardment planes. It was also found that unmarried officers and men were more favorable toward return than the married ones. In most of the categories, the differences in percentages willing to volunteer are approximately 10 percent in favor of unmarried groups.

Using responses to items of the Air-Crew Officer Blank, which was a part of the Instructor Selection Test Battery, studies were made of attitudes toward assignments in the United States. The samples used for this analysis consisted of 400 each of bombardiers, navigators, bomber pilots, and fighter pilots. These groups were processed at the AAF Redistribution Station at Miami Beach in November and December of 1944. The responses to a group of 6 items which the returnees responded to by indicating one of various degrees of like or dislike are tabulated below in terms of the percentages of each of the four groups indicating "like" or "like very much" for the assignment described:

| | Bombardiers | Navigators | Bomber pilots | Fighter pilots |
|---|---|---|---|---|
| | Percent | Percent | Percent | Percent |
| Considerable flying in noncombat aircraft | 47 | 51 | 69 | 33 |
| Training in new combat planes | 20 | 16 | 41 | 65 |
| Operations or communications officer with flight pay | 56 | 56 | 61 | 41 |
| Instructor | 68 | 68 | 43 | 43 |
| Assistant air inspector | 46 | 34 | 40 | 20 |
| Administrative or tactical ( er in United States with flight pay | 54 | 54 | 29 | 18 |

These responses indicated that bombardiers and navigators would prefer assignments as instructors and would least like training in new combat planes. Bomber pilots would prefer an assignment which would give them considerable flying in noncombat aircraft, while the fighter pilots are most eager for training in new combat planes.

Less than one-fifth of the flyers said they would take an assignment under "definitely bad" or "rather bad" conditions in order to be promoted. More than three-quarters said they would prefer an assignment where the chances for promotion were "normal" or even "below average," provided that assignment was "average" or "suited" them. The distribution of responses to this item was as follows:

| | *Bombardiers* | *Navigators* | *Bomber pilots* | *Fighter pilots* |
|---|---|---|---|---|
| | *Percent* | *Percent* | *Percent* | *Percent* |
| Prefer an average assignment with whatever chance of promotion is normal in that work | 43 | 44 | 45 | 51 |
| Would take a situation that suited me even though my chance for promotion was below average | 29 | 31 | 30 | 27 |
| Would give up all thought of promotion for the rest of the war for a sufficiently desirable assignment | 16 | 16 | 12 | 7 |
| Would take an assignment under rather bad conditions in order to be promoted | 7 | 7 | 8 | 9 |
| Would take an assignment under definitely bad conditions in order to be promoted | 5 | 2 | 5 | 6 |

The returnees were asked their attitudes toward being assigned where it would be possible to have their wives or families with them in fairly comfortable quarters. As can be seen from the following distribution of responses, more than two-thirds felt this was an important consideration:

| | *Bombardiers* | *Navigators* | *Bomber pilots* | *Fighter pilots* |
|---|---|---|---|---|
| | *Percent* | *Percent* | *Percent* | *Percent* |
| The most important consideration | 22 | 19 | 15 | 18 |
| Quite important *or* very important | 49 | 50 | 63 | 53 |
| Not at all important *or* relatively unimportant *or* I have none | 29 | 31 | 22 | 29 |

About two-thirds of the returnees said that assignment in the United States rather than in the combat zone was a strong influence in their attitudes toward an instructor assignment. They indicated that this factor was of more importance to them than any of the other factors listed. The opportunity to keep up to date on their specialty was also reported to be an important consideration.

The results of these analyses of studies of attitudes and preferences for assignment in the United States and overseas were made available to the personnel and training officers as a basis for developing policies and plans. The findings were of great value in providing a sound basis for developing plans for training and for the utilization of personnel with combat experience.

## Summary and Conclusions

A number of studies were conducted to provide a more specific and definite basis for understanding the origin, nature, and effect on operations of morale in the military situation. The general approach was to make an intensive analysis of the reactions of individuals to the combat situation. In general, the morale in operating units was found to be excellent in spite of the extremely hazardous character of the combat operations in which many of them were engaged.

An analysis of the most important factors which kept these men flying and fighting was made and the importance of leadership in the combat situation was emphasized. Administrative procedures found to be effective in operating units were reported. It was indicated that the men regarded the most important factors in reducing fear to be confidence in equipment, in fellow crew members, and in leaders.

A very small percentage of men engaged in air-crew combat operations found it impossible to continue and had to be removed from combat flying assignments. On the other hand, almost all personnel who had been engaged in combat operations for a substantial period showed definite evidence of fatigue, restlessness, and depression. Since these air-crew men were originally selected as having normal emotional behavior, these findings suggest that excessive emotional reactions arising from frustration and conflict in any life situation can be expected to induce an anxiety in which the individual shows the usual symptoms of the psychoneurotic.

Screening procedures were developed by aviation psychologists in the AAF Personnel Distribution Command which selected between 60 and 70 percent of individuals admitted to hospitals as psychiatric patients and included only 10 to 20 percent of those admitted as nonpsychiatric patients. On the basis of the research data accumulated, it should be possible to develop a very valuable screening questionnaire for use in connection with large-scale medical examining procedures.

An interesting group-counseling program was developed by aviation psychologists in the AAF Convalescent Hospitals. On the basis of subjective observations and opinions and the relatively high percentage of attendance in these voluntary groups, it was believed that this program was of substantial value to the patients. To provide a basis for improving the methods of evaluating the effectiveness of group-counseling procedures, a technique of objective recording and analysis was developed at the AAF Convalescent Hospital at Cochran Field. This procedure was not developed soon enough for use in evaluating the group-counseling program in the AAF Convalescent Hospitals. It appears to provide a valuable research technique for evaluating future activities of this type.

The series of studies on attitudes and preferences of combat returnees provided much valuable information for use in developing plans for the utilization of personnel with combat experience.

It is believed that further analyses of data which have been collected along with findings reported in this chapter should provide a much sounder understanding of individual reactions to combat than has previously been available.

# Part III

# General Contributions to Aviation Psychology

# CHAPTER NINE

# General Contributions to the Theory and Knowledge of Individual Differences and Trait Differences

## INTRODUCTION

In this and the following chapters an attempt will be made to list and discuss briefly some of the more general findings growing out of the work of the Aviation Psychology Program. Although it will not be possible in the space here to give all of the available data concerning the points discussed, an attempt will be made to provide enough of the experimental data to illustrate the basis for the conclusions reached. Certain of the conclusions are based on informal evidence and experiences rather than systematic observation or controlled experimentation. Therefore, some of the statements would be better described as hypotheses than as conclusions. However, in no cases are casually developed hypotheses presented.

The intent is to provide a clear statement of certain conclusions based on the experience during the war in the Aviation Psychology Program in the hope that they will stimulate research in those areas where definitive information is lacking. The statements made in this part as in previous parts are generally shared by a number of individuals having had the common experiences in the Aviation Psychology Program in the Army Air Forces. Responsibility for all statements is taken by the editor of this report. This chapter will discuss the theory and knowledge of individual differences and trait differences under the headings of aptitudes, motivational factors, and factors of individual adjustment.

## THE NATURE OF INDIVIDUAL AND TRAIT DIFFERENCES IN APTITUDE

### Introduction

The large number of ways in which individuals differ is demonstrated by the many adjectives and trait names which have grown up in modern languages to describe people. The fact that these words are extensively used in conversation and in writing strongly suggests that the ordinary person believes he is capable of detecting differences in these traits in the people with whom he associates. However, it is only within the past 50 to 75 years that systematic attempts have been made to obtain accurate measures of these individual and trait differences.

The pioneer work of Binet and Terman demonstrated that it was possible to describe individuals in terms of their ability to do a variety of selected short standardized tasks. It appeared to be especially useful to relate the level of the most difficult tasks they could do to the age of children just able to perform these tasks. This led to the intelligence quotient or "IQ" which began to be extensively used about 1916.

The First World War accelerated the wide use of a second type of measure of general intellectual ability. This was typified by the Army Alpha Test. This test was a printed test which could be administered to large groups by examiners having only a small amount of training in testing procedures. The extensive use of these tests in the Army popularized them to such an extent that they were widely used in the years which followed.

This led to an overemphasis of the traits which were the most conspicuous factors in these tests. These were principally the traits associated with success in scholastic work of a general type. A few leaders in the field conducted research studies during this period in an effort to define and provide measures for some of the more important aptitudes. The principal contributions in this area were those of Kelley, Thurstone, Thorndike, and Spearman.

However, in 1941 the established knowledge regarding the correlations between the various tests of aptitudes regarded as basic were based on only a few samples of a few hundred cases each. Of even greater practical importance, practically nothing was known about the general significance or predictive value of these tests except for the special situation of predicting grades in typical academic courses.

The world emergency in 1941 created a need for the rapid training of many men in various technical specialties. This brought with it the necessity for some way of classifying candidates for training in these various types of schools. Work on the solution of these problems provided an opportunity to make large-scale studies of the practical significance of a wide variety of aptitude measures for several types of duty assignments which differed markedly in their requirements for success.

Thus, whereas the First World War demonstrated that the well-known differences in the individual's general intellectual ability could be measured by a test such as the Army Alpha, the Second World War has shown that the special aptitudes and characteristics needed for success in particular assignments can be measured by appropriate objective tests using apparatus or printed-test booklets. The classification testing program also provided an opportunity for studying the interrelations of a substantial number of aptitude tests on unprecedentedly large groups.

The compressed "life cycle" of men selected for special training in the Air Forces from the time of individual analysis and classification, through training, and on to performing the job for which he was selected and trained consisted of a period of only 2 or 3 years. During this period the necessities of war gave military authorities a much greater degree of control of the individual than would have ordinarily been possible. These circumstances, together with the very large staff, including a substantial number of the country's best psychologists, made it possible to make important contributions to this general field during the war program.

### Trait Theory[1]

The number of basic uncorrelated aptitude factors underlying the acquisition of general human abilities is probably very much larger than most experts had supposed. As mentioned above, at the time of World War I, attention was focussed on the one trait called general intelligence and there was a tendency among psychologists to oversimplify all individual differences and categorize people almost entirely with regard to this one trait in thinking of their mental aptitudes. Following this, a few psychologists began to think of possibly three or four traits.

In a study reported by T. L. Kelley[2] in 1935, a number of eminent psychologists were asked to rate the various occupational groups listed in the census reports with regard to the traits which they regarded as important for individual success in the specific occupations. Very few of the raters felt that it was necessary to use more than 8 or 10 traits to describe the trait requirements of the various occupations.

The experience in developing tests in the Aviation Psychology Program suggests that rather than the 10, or even 20, traits which many psychologists regarded as the maximum number of important independent traits only a few years ago, the number is almost certainly 50 or 100, or even larger.

Examination of the intercorrelations of the tests in the Air-Crew Classification Test Battery which are reported in appendix B indicates that a substantial number of traits are required to explain the individual differences included in the 20 tests of this battery. The more extensive tables of intercorrelations included in the research report on Printed Classification Tests[3] and on the AAF Qualifying Examination[4] emphasized to an even greater extent the large number of independent traits which are necessary to provide an adequate description and prediction of individual and trait differences.

### Trait Measurement

It is possible to develop reliable measures for a large number of traits which are practically uncorrelated with each other in groups of adults with relatively homogeneous backgrounds. As mentioned in the preceding section, the tests of the Air-Crew Classification Battery had relatively low intercorrelations. In a typical sample of men tested at the AAF Classification Centers only about 2 of approximately 200 intercorrelations were as high as 0.50. Approximately half of the intercorrelations were below 0.20. These intercorrelations are presented in table 9.1.

---

[1] As used in this report, a trait refers to any distinguishing quality or characteristic of an individual.

[2] T. L. Kelley. *Essential Traits of Mental Life*: Cambridge. Harvard University Press, 1935.

[3] J. P. Guilford and J. I. Lacey, eds. *Printed Classification Tests.* AAF Aviation Psychology Program Research Reports, No. 5. Washington: Government Printing Office, 1947.

[4] F. B. Davis, ed. *The AAF Qualifying Examination.* AAF Aviation Psychology Program Research Reports, No. 6. Washington: Government Printing Office, 1947.

TABLE 9.1.—*Intercorrelations of September 1944 Air-Crew Classification Test Battery (N = 8,158)*

[Decimals before correlation coefficients omitted]

| Test | Test No. | Dial and Table Reading | Biographical Data (P) | Biographical Data (N) | Spatial Orientation I | Spatial Orientation II | Reading Comprehension | Instrument Comprehension | Mechanical Principles | Speed of Identification | Numerical Operations I | Numerical Operations II | Mechanical Information | General Information | Practical Judgment | Arithmetic Reasoning | Rotary Pursuit | Rudder Control | Finger Dexterity | Complex Coordination | Two-Hand Coordination | Discrimination Reaction Time |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dial and Table Reading | CP621-22A | | 03 | 17 | 40 | 35 | 41 | 37 | 21 | 32 | 52 | 55 | 03 | 15 | 27 | 49 | 14 | 05 | 23 | 37 | 22 | 45 |
| Biographical Data (P) | CE602D | 03 | | 20 | 05 | 13 | 01 | 19 | 26 | 13 | -13 | -11 | 35 | 36 | 07 | 03 | 13 | 27 | 08 | 18 | 21 | 10 |
| Biographical Data (N) | CE602D | 17 | 20 | | 09 | 04 | 13 | 09 | 00 | 03 | 14 | 12 | -07 | 07 | 07 | 14 | 05 | 00 | 09 | 07 | 02 | 16 |
| Spatial Orientation I | CP501B | 40 | 05 | 09 | | 39 | 18 | 28 | 15 | 46 | 23 | 23 | 01 | 15 | 14 | 16 | 16 | 05 | 19 | 32 | 17 | 30 |
| Spatial Orientation II | CP503B | 35 | 13 | 04 | 39 | | 33 | 37 | 39 | 36 | 10 | 17 | 18 | 23 | 28 | 30 | 13 | 14 | 13 | 29 | 19 | 31 |
| Reading Comprehension | CI614H | 41 | 01 | 13 | 18 | 33 | | 33 | 31 | 28 | 29 | 35 | 18 | 27 | 41 | 52 | 04 | 02 | 10 | 20 | 09 | 33 |
| Instrument Comprehension | CI616C | 37 | 19 | 09 | 28 | 37 | 33 | | 38 | 28 | 17 | 22 | 22 | 37 | 26 | 30 | 20 | 23 | 15 | 36 | 25 | 36 |
| Mechanical Principles | CI903B | 21 | 26 | 00 | 15 | 39 | 31 | 38 | | 17 | -01 | 05 | 47 | 35 | 33 | 34 | 20 | 34 | 07 | 31 | 33 | 24 |
| Speed of Identification | CP610A | 32 | 13 | 05 | 46 | 36 | 19 | 28 | 17 | | 15 | 17 | 07 | 23 | 14 | 12 | 15 | 05 | 10 | 28 | 14 | 26 |
| Numerical Operations I | CI702B | 52 | -13 | 14 | 23 | 10 | 29 | 17 | -01 | 15 | | 67 | -13 | 07 | 13 | 40 | 03 | -07 | 15 | 16 | 04 | 27 |
| Numerical Operations II | CI702B | 55 | -11 | 12 | 23 | 17 | 35 | 22 | 05 | 17 | 67 | | -06 | 05 | 20 | 30 | 04 | -05 | 14 | 12 | 07 | 31 |
| Mechanical Information | CI905B | 03 | 35 | -07 | 01 | 18 | 18 | 22 | 47 | 07 | -13 | -06 | | 45 | 25 | 13 | 05 | 28 | 04 | 15 | 29 | 03 |
| General Information | CE505F | 15 | 35 | 07 | 15 | 23 | 27 | 37 | 35 | 23 | 07 | 05 | 45 | | 22 | 14 | 16 | 39 | 08 | 23 | 19 | 16 |
| Practical Judgment | CI301C | 27 | 07 | 07 | 14 | 28 | 41 | 26 | 33 | 14 | 13 | 20 | 25 | 22 | | 34 | 03 | 11 | 05 | 20 | 14 | 24 |
| Arithmetic Reasoning | CI206C | 49 | 00 | 14 | 16 | 30 | 52 | 30 | 34 | 12 | 40 | 30 | 11 | 14 | 34 | | 01 | 02 | 09 | 19 | 11 | 21 |
| Rotary Pursuit | CP410B | 14 | 13 | 05 | 16 | 13 | 04 | 20 | 20 | 15 | 03 | 04 | 03 | 16 | 07 | 01 | | 35 | 30 | 39 | 33 | 23 |
| Rudder Control | CM120B | 05 | 27 | 00 | 05 | 14 | 02 | 23 | 34 | 05 | -07 | -05 | 24 | 19 | 13 | 02 | 35 | | 11 | 19 | 40 | 15 |
| Finger Dexterity | CM116A | 23 | 08 | 05 | 19 | 13 | 10 | 15 | 07 | 20 | 15 | 14 | 04 | 08 | 03 | 07 | 30 | 11 | | 29 | 23 | 27 |
| Complex Coordination | CM701A | 37 | 18 | 07 | 30 | 29 | 20 | 36 | 31 | 28 | 16 | 12 | 15 | 22 | 20 | 19 | 39 | 19 | 29 | | 49 | 41 |
| Two-Hand Coordination | CM101A | 22 | 21 | 02 | 17 | 19 | 09 | 25 | 33 | 14 | 04 | 07 | 29 | 19 | 14 | 11 | 33 | 40 | 23 | 49 | | 27 |
| Discrimination Reaction Time | CP611D | 45 | 10 | 16 | 30 | 31 | 33 | 36 | 25 | 26 | 27 | 31 | 05 | 16 | 24 | 31 | 23 | 15 | 27 | 41 | 27 | |
| Mean | | 5.7 | 5.1 | 5.1 | 5.0 | 5.0 | 5.3 | 5.0 | 5.2 | 5.3 | 5.0 | 5.1 | 5.1 | 5.2 | 5.1 | 5.4 | 5.0 | 5.6 | 5.0 | 5.4 | 5.0 | 5.2 |
| Standard Deviation | | 2.1 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.1 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |

It is interesting that the two correlations exceeding 0.50 are between variables which are frequently thought of as being quite different. One of the correlations is between the Numerical Operations Test and the Dial and Table Reading Test. The first of these tests is a simple arithmetic test including addition, subtraction, multiplication, and division of small numbers. The answers printed on the answer sheets are merely checked as right or wrong. On the other hand, the Dial and Table Reading Test involved no calculations at all. The only use of numbers is to enter tables with them or to use them as limits in performing interpolation in reading the values indicated by pointers on dial faces.

The other pair of tests having a correlation of above 0.50 includes Arithmetic Reasoning and Reading Comprehension. The Reading Comprehension Test involved no quantitative problems. The task consists simply of reading paragraphs and answering questions about the material in the paragraphs which require understanding and interpretation. The Arithmetic Reasoning Test involves only very simple vocabulary and the sentences are not long or involved but are short statements of conditions of the problems. Both of these latter tests have correlations of 0.40 with the Dial and Table Reading Test which superficially appears not to require any complex thought processes at all, but merely the identification of certain symbols in a simple clerical fashion.

In the group of more than 8,000 men on whom these intercorrelations are based, Finger Dexterity scores were not correlated with scores on any of the other test variables as highly as 0.30 and Rotary Pursuit Test scores had correlations with the other variables which were all below 0.40.

An analysis of the functions measured by the various apparatus tests provides some interesting points. For example, the correlation between the Finger Dexterity Test and Mechanical Principles Test is 0.07. However, both Mechanical Principles and Finger Dexterity have correlations with the other apparatus tests which are very nearly as high as those between most of the other apparatus tests.

The Instrument Comprehension Test is another of the printed classification tests which has a relatively substantial correlation with most of the apparatus tests. The largest of these correlations are with the Complex Coordination Test and the Discrimination Reaction Time Test. These are both 0.36. These correlations are within 0.02 of the largest correlations between Instrument Comprehension and any of the printed tests.

The Dial and Table Reading Test also has substantial correlations with the tests of Complex Coordination and Discrimination Reaction Time. The correlation with the Complex Coordination Test is 0.36, and that for Discrimination Reaction Time Test is 0.45.

It does not appear that the correlations of these apparatus tests with printed tests had anything to do with the comprehension of the directions, since the correlations between the Reading Comprehension Test and Rotary Pursuit, Rudder Control, Finger Dexterity, and Two-Hand Coordination were 0.04, 0.02, 0.10, and 0.09 respectively. The correlations between the Reading Comprehension Test and Complex Coordination and Discrimination Reaction Time Tests were 0.20 and 0.33 respectively.

There are a number of other interesting points within the tables which space does not permit discussion of here. As an aid to formulating hypotheses regarding the nature of the common variance for the tests studied, a large number of studies using various techniques of factor analysis were carried out. The results of these studies are reported in detail in the research reports in this series on the AAF Qualifying Examination and on Printed Classification Tests.

The principal point gained from studying these intercorrelations is not so much the hypotheses for explaining the small amounts of common variance as it is the obviously large amount of unique variance in each of the tests in the battery.

## The Importance of Traits

Differential patterns in terms of aptitude requirements exist for various types of activities and these patterns have great practical value for the guidance and placement of individuals. One of the outstanding findings of the work in the Aviation Psychology Program was the very high degree of specificity among the requirements for various types of activities. Individuals who failed to solo in primary pilot training because of extreme ineptness were later graduated at the top of their class in navigation training. Similarly, men who did very well on one type of gunnery trainer did very poorly on another which superficially appeared to be quite similar.

This independence in the requirements critical for success in various types of activities makes possible substantial gains in the efficiency in the use of human resources by identifying the requirements and classifying individuals in accordance with their scores on tests measuring the required traits. The very substantial gains which can be made in the average ability for each trait if there are two, three, or more activities for which the requirements are independent are illustrated in the introductory discussion of the classification problem in chapter 4.

It need hardly be stated that individual differences of great practical significance exist with respect to ability in practically all types of skilled work. Few jobs exist which can be performed equally well by practically all of the participants. Of great importance was the finding that some of these individual and trait differences persist in spite of long periods of training and experience in the activity. It was found that the man who did poorly on the aptitude tests used in computing the pilot stanine not only did poorly in primary flight training, but those few who were not eliminated tended to continue to do poorly in basic, advanced, and operational training and also in combat.

Another valuable finding was the correlation between the original test scores of men who took the aptitude tests and were classified as bombardiers and navigators and the scores made by these men when retested after completing their training and a tour of duty in combat. These correlations for the applicants for pilot training after return to the United States indicated that there was a substantial amount of similarity in the order with respect to aptitude for pilot training on retesting to that which they were given at the time of their original classification testing.

It was especially interesting to find that certain tests, such as Arithmetic Reasoning, Vocabulary, and Numerical Operations, which were known to be substantially correlated with criteria of success in academic work had practically zero or even slightly negative correlations with measures of success in such activities as learning to fly, to bomb, or to shoot a gun. On the other hand, this finding did not result in the removal of these tests from the Air-Crew Classification Battery, since it was also found that these same tests were substantially correlated with success in learning navigation.

In conclusion, the important point to be noted is that not only were a large number of independent traits identified, but that these traits were found to be of definite practical importance in classifying individuals for the types of activities in which they could render the largest relative contribution to the total efforts of the group.

## THE NATURE AND SIGNIFICANCE OF MOTIVATIONAL FACTORS

### Introduction

Probably most psychologists believe that the next most important factor after aptitude in determining an individual's success in a specific activity is his motivation or desire to succeed in that activity. The Air-Crew Classification Program provided an unusually good opportunity to observe the effects of motivational factors both in training work and in later job assignments.

Among the important questions which were studied are such matters as:

1. Is motivation something which can be readily imparted to the individual in a specific situation?

2. Are there fundamental bases for motivations, and if so, how can these be defined and identified?

3. How stable and permanent are statements of preferences for various types of activities, and how important are they for subsequent job performance?

### The Stability of Stated Preferences

Stated preferences for various types of activities frequently lack a sound basis and can be modified by relatively short periods of indoctrination. As was pointed out in chapter 4, more than 90 percent of the men taking the air-crew classification tests wished to have their aptitudes, as determined by the tests, given some degree of consideration in making their final assignment regardless of their stated preferences. It was also found in interviews

723323—47—17

that these men recognized the importance of the findings regarding their aptitudes and were usually willing to accept advice regarding the most suitable assignments for them.

As indicated in the discussion of instructor selection, the preferences and attitudes regarding assignment to duty as an instructor as given at the redistribution station had practically no agreement with the statements regarding their liking for this assignment after having been on the job for several weeks. On the other hand, their attitudes toward the job remained fairly stable during the period that they were assigned to this job.

## The Origins of Motivation

There do exist deep-rooted and broad interests and values which can be strong sources of motivation for the individual. There has long been a difference of opinion regarding the relative extent to which motivation is something which the new person needs to be given or something with which he either is or is not naturally, or at least rather permanently, endowed.

It was indicated in earlier chapters that surveys of combat operations and particularly of morale in these combat units clearly indicated that the leadership of command and staff personnel were of fundamental importance in determining the motivation of the groups and the men under them. Similarly, it was observed in the training schools that an enthusiastic instructor could carry along cadets whose basic motivation was very weak for the particular type of activity.

In addition to these specific outside motivating factors, another important consideration in determining motivation is the general motivation of an individual to succeed at all activities. Even broader than this is the individual's character as discussed in the preceding chapter. This tendency to work for the group rather than for the individual is fundamental to effective group action.

There was ample evidence, however, that broad interests and values which can be utilized in increasing motivation for a specific type of assignment were present in these individuals and could be utilized to improve the effectiveness of the individuals' efforts in one of the specific fields if he were assigned to it.

An exploratory study of the origin of preferences for various types of duties was made using approximately 100 high-school boys attending summer school at the Central High School, Washington, D. C.[6] Most of the boys were 17 years old. This group was asked to rate 14 job descriptions using a 5-point scale from "would dislike this job" to "would want very much to apply for this job." The jobs, some of which were nonexistent, were carefully structured so that a measure of the relative importance of rank, flying versus ground jobs, postwar value of experience, and combat duty versus noncombat duty, could be obtained. In this group the order of importance of these factors in determining the choices between various jobs was flying,

[6] This study was carried out by R. L. Thorndike.

combat, rank, and finally postwar value of experience. It is interesting that these boys stated that being an enlisted tail gunner on a Flying Fortress was more attractive to them than being an officer on engineering duty.

It was found that there was a definite correlation between the aptitude score made by the individual for a particular job and the strength of interest reported for that job. This correlation was rather small, presumably because the individuals had only a vague notion regarding their relative aptitudes for various types of activities and also lacked adequate knowledge of the requirements of the activity itself.

An example of changing preferences is provided by a group of several hundred individuals who were classified for navigator training and sent to navigator preflight school. Although some of the individuals had originally wanted to be pilots or bombardiers, while in preflight school the large majority of the group became enthusiastic about becoming navigators. However, in order to fill quotas caused by changing requirements in the numbers needed for the various air-crew positions, it was found necessary to send the entire class from the preflight school directly into advanced bombardier training. In bombardier training this group was a source of more difficulty than any other bombardier class on record. They were sure that they wanted to be anything but bombardiers.

Two or three weeks before this group was scheduled to graduate from bombardier training, a request was received for a substantial number of individuals having both bombardier and navigator training. Accordingly, an aviation psychologist was sent to this particular bombardier school to screen individuals on the basis of aptitudes and preferences and select a group to receive navigator training as officers at the completion of their bombardier training. Much to everyone's surprise, it was discovered that very few individuals in the group now wanted navigator training.

The analysis of the responses to items on the Biographical Data Blank provided one of the best sources for identifying those deep-rooted and broad interests and values which provide a sound basis for preferences. For example, it was found that individuals who had been interested in model-airplane building were more successful in flying training than those who had not had such interests. In reporting their reasons for becoming an aviation cadet, the large majority of these men stated that they expected to use the flight training after they left the Army. A substantial proportion gave as their reason that they felt they were best suited for this type of duty.

Although it is clear that preferences do have a fundamental basis, it is also obvious that these bases need to be clarified for the individual. It was usually found that their broad extent made it possible to reconcile them with the individual's aptitudes.

### The Prediction of Motivation

Measures of motivational factors can be used to improve the predictions of success obtained from aptitude test scores. The individual's statements

In table 4.9 are shown the correlations for a sample of approximately 10,000 bombardiers between their scores on the various air-crew classification tests and their average circular errors. It is seen that only four tests have negative correlations with this criterion and none of these negative values exceeds —0.005. Since the standard error of most of these correlation coefficients is 0.01, it is clear that they cannot be regarded as a chance distribution of values. For this type of bombing it appears that the traits measured by the tests of Numerical Operations, Complex Coordination, Two-Hand Coordination, Biographical Data (Navigator and Pilot scores), Rudder Control, Dial and Table Reading, Instrument Comprehension I and II, Spatial Orientation I, Rotary Pursuit, and Discrimination Reaction Time are all significantly correlated with bombing accuracy scores.

TABLE 4.9.—*The relative value of the Air-Crew Classification Tests for predicting average circular error in samples of bombardiers in AAF training command stations*

| Tests | Test No. | Number of individuals in combination[1] | Combined weighted r |
|---|---|---|---|
| Numerical Operations | CI702B | 3,947 | 0.053 |
| Complex Coordination | CM701A | 10,342 | .048 |
| Two-Hand Coordination | CM101A | 10,000 | .036 |
| Biographical Data (N) | CE602D | 10,000 | .036 |
| Biographical Data (P) | CE602D | 10,000 | .031 |
| Speed of Identification | CP610A | 4,075 | .028 |
| Rudder Control | CM120B | 7,941 | .028 |
| Dial and Table Reading | CP622-21A | 10,764 | .026 |
| Instrument Comprehension II | CI616B | 7,941 | .026 |
| Spatial Orientation I | CP501 | 11,672 | .024 |
| Rotary Pursuit | CM803A | 10,000 | .024 |
| Instrument Comprehension I | CI615B | 7,941 | .024 |
| Discrimination Reaction Time | CP611D | 10,000 | .023 |
| Aiming Stress | CM211A | 2,059 | .019 |
| Finger Dexterity | CM116A | 10,430 | .017 |
| General Information (P) | CE602D | 10,000 | .011 |
| Mathematics B | CI206C | 10,000 | .004 |
| General Information (N) | CE505D | 2,059 | .000 |
| Mechanical Principles | CI903A | 10,393 | —.001 |
| Mathematics A | CI702F | 10,962 | —.001 |
| Reading Comprehension | CI614G | 10,32[illegible] | —.003 |
| Spatial Orientation II | CP503B | 10,000 | —.005 |

[1] Bombardier Classes 44-09 through 45-13 provided the largest number of cases in the composite. Earlier data on Classes 42-11 through 43-1 are represented in smaller proportions in those tests which appeared in identical or similar form on both the early and later batteries. Both new aviation cadets and reclassified pilots were included.

The job of bombardier in training is not very similar to his job in combat. The average bombing circular errors in the AAF Training Command schools are less than half the size of those in combat. The problems of target identification are practically nonexistent for bombs dropped at the training schools, combat conditions frequently necessitate short bombing runs which emphasize speed much more than the uniform runs of the training situation, more demands are made on flexibility of mental processes and resourcefulness by the complex combat situation as compared to training, and new factors of distraction and apprehension are introduced into the combat situation which are not present in training.

No studies were made of the correlations between bombing circular errors in operational training and the air-crew classification tests. The operational-training situation is more like the combat situation in the respect that the bombardier always works with the same air crew but differs from combat in most of the other important respects listed above.

A number of studies were made comparing air crew classification test scores with the circular errors of bombs dropped in actual combat. These circular errors were obtained by measuring the distance on strike photographs from the assigned target and the point of impact of the bombs. In combat operations, bombing was usually done in formations. Under these circumstances, the bombardier in the lead plane of the formation was responsible for identifying the target and establishing the course and dropping angle. The bombardiers in the remaining planes of the formation usually watched the lead plane and released their bombs with a toggle switch when they saw the bombs fall from the lead plane. For this reason it was ordinarily possible to evaluate only the lead bombardiers. This greatly reduced the numbers of individuals in the validation studies.

TABLE 4.10.—*The relative value of the Air-Crew Classification Test for predicting combat average circular error in samples of lead bombardiers*

| Tests | Test No. | Number of individuals in combination | Combined[1] weighted r |
|---|---|---|---|
| Discrimination Reaction Time | CP611D | 1,272 | 0.099 |
| Spatial Orientation I | CP503B | 1,271 | .090 |
| Technical Vocabulary (P) | CE505C | 1,046 | .085 |
| Mechanical Principles | CI903A (AC10D6)[2] | 920 | .073 |
| Reading Comprehension | CI614G (AC10D2) | 925 | .072 |
| Dial and Table Readings | CP622-21A | 1,149 | .061 |
| Complex Coordination | CM701A | 1,149 | .049 |
| Mathematics A | CI702A | 1,046 | .041 |
| Spatial Orientation II | CP503B | 1,259 | .029 |
| Speed of Identification | CP610A | 1,120 | .026 |
| Mathematics B | CI710A, (CI706A, CI206B) | 1,051 | .018 |
| Technical Vocabulary (N) | CE505C | 1,046 | .016 |
| Two-Hand Coordination | CM101A | 1,006 | .006 |
| Technical Vocabulary (B) | CE505C | 1,044 | .002 |
| Rotary Pursuit | CM803A | 757 | —.009 |
| Mechanical Information | CI905A | 432 | —.014 |
| Aiming Stress | CM211A | 1,075 | —.030 |
| Finger Dexterity | CM116A | 1,148 | —.035 |
| Numerical Operations | CI702B | 1,119 | —.047 |

[1]The following studies were included:
AERD No. 1—Eighth Air Force, April and May 1944.
Eighth Air Force, June 1944.
AERD No. 2—Fifteenth Air Force, November 1944.
Ninth Air Force, March 1945.
Eighth Air Force, April 1945 (four studies on B-17's and B-24's above and below 21,000 feet).
AERD No. 3—POA, March 1945.

[2]The combined correlations for these tests include a composite of all of the forms indicated. The test number or numbers in parentheses constitute the smaller proportion of cases.

The combined results of all studies of bombardiers' circular errors in combat operations are given in table 4.10. The tests at the top of the list include Discrimination Reaction Time, Spatial Orientation I (Target Identification), Technical Vocabulary (P) (Interest in aviation and related fields), Mechanical Principles, Reading Comprehension, and Dial and Table Reading. The first three correlations are significant at the 1-percent level and the other three at the 5-percent level. Small negative values insignificantly different from zero were obtained for five tests including Rotary Pursuit, Mechanical Information, Aiming Stress, Finger Dexterity, and Numerical Operations. The negative value for Numerical Operations is consistent with the indication from other studies that scores on this test are negatively associated with the interests and personality characteristics which are typical of the outstanding participants in combat aerial operations.

The differences between the correlations obtained for these tests with combat circular errors as compared with training circular errors are striking but not unreasonable. In the training situation where a great deal of uniformity exists for the conditions of the bomb drops, about the only errors

TABLE 4.11.—*The relative value of Air-Crew Classification Tests in samples of bombardiers for predicting combat ratings, bombing accuracy, reclassification, and importance of positions held*

| Tests | Test No. | Data combined | Approx. number of cases in combination | Combined[1] weighted $r$ |
|---|---|---|---|---|
| Discrimination Reaction Time | CP611D | 1–31 | 2,690 | 0.074 |
| Spatial Orientation I | CP501B | 1–31 | 2,690 | .068 |
| Mechanical Principles | CI903A, (AC10D6)[2] | 1–6,9–11,14–30 | 2,430 | .066 |
| Spatial Orientation II | CP503B | 1–30 | 2,680 | .051 |
| Mathematics A | CI702E | 1–6,9–30 | 2,510 | .045 |
| Technical Vocabulary (P) | CE505C | 1–6,9–30 | 2,510 | .042 |
| Technical Vocabulary (N) | CE505C | 1–6,9–30 | 2,510 | .042 |
| Complex Coordination | CM701A | 1–6,9–31 | 2,520 | .035 |
| Reading Comprehension | CI614G (AC10D2) | 1–6,9–11,14–30 | 2,430 | .032 |
| Dial and Table Reading | CP622–21A, (CP621A, CP622A) | 1–6,9–31 | 2,520 | .028 |
| Technical Vocabulary (B) | CE505C | 1–6,9–30 | 2,510 | .025 |
| Mathematics B | CI710A, (CI206B, CI706A) | 1–6,9–30 | 2,510 | .023 |
| Mechanical Information | CI905A | 6,10,11,14–30 | 1,670 | .022 |
| Speed of Identification | CP610A | 1–6,9–30 | 2,510 | .020 |
| Two-Hand Coordination | CM101A | 1–6,8–31 | 2,560 | .014 |
| Rotary Pursuit | CM803A | 1–6,9–11,14–30 | 2,430 | .006 |
| Numerical Operations | CI702B | 1–6,9–30 | 2,510 | —.002 |
| Finger Dexterity | CM116A | 1–6,9–31 | 2,520 | —.011 |
| Aiming Stress | CM211A | 1–6,9–30 | 2,510 | —.012 |

[1]Studies included in combination:

| No. | Study | Number of cases |
|---|---|---|
| | ***AERD No. 1*** | |
| 1. | Ratings provided by staff personnel, 8AF | 220 |
| 2, 3. | Radial error and percent hits, 8AF April and May 1944 | 250 |
| 4, 5. | Radial error and percent hits, 8AF June 1944 | 158–162 |
| | ***AERD No. 2*** | |
| 6. | Range error, 15AF | 187–295 |
| 7. | Range error on bombardiers with 2 or more missions, 15AF | 7– 95 |
| 8. | Radial error for unit lead bombardiers who sighted for both range and deflection, 15AF | 115–123 |
| 9. | Percent "wins"—evaluation by wing and group staffs of lead bombardiers after each mission, a "win" being a satisfactory bombing performance considering all circumstances. Criterion measure was percent of "wins" in total missions, 15AF. | 38– 51 |
| 10. | Range error, sample drawn from 6, consisting of those whose inaccuracy was judged as due to bombardier error, 15AF. | 73–117 |
| 11. | Analysis from strike photographs of percent of bombs falling within 300 feet of target, for lead bombardiers with 4 or more usable photos, 12AF. | 45– 85 |
| 12. | Hit-miss ratio for lead bombardiers = hits + ½ "near misses" — total number missions, 12AF. | 38– 46 |
| 13. | Ratings by group bombardiers on all bombardiers who had completed 30 or more missions on a five point scale, 12AF. | 35– 39 |
| 14. | Average accuracy score — a converted score based on the radial error for lead bombardiers with scores for 3 or more missions, 9AF. | 71–172 |
| 15, 16. | B-17 radial error and percent hits below 21,000 feet, 8AF | 59–109 |
| 17, 18. | B-24 radial error and percent hits below 21,000 feet, 8AF | 59– 89 |
| 19, 20. | B-17 radial error and percent hits above 21,000 feet, 8AF | 168–317 |
| 21, 22. | B-24 radial error and percent hits above 21,000 feet, 8AF | 88–122 |
| 23, 24. | Specially evaluated missions, each bombardier classified as good or poor | 27– 58 |
| 25, 26. | Ratings by supervisory officers on a five-point scale for bombardiers with 10 or more missions. | 145–331 |
| | $r$ biserials calculated from critical ratio data (*See* footnote 14). | |
| 27. | Medical Disposition Board cases versus control ..... control 75-114 MDB.. | 38– 59 |
| 28. | Central Medical Establishment cases versus control ..... CME.. | 59– 98 |
| | ..... control . | 59– 92 |
| 29. | Reclassified cases versus control ..... recl... | 13– 21 |
| | ..... control.. | 25– 35 |
| 30. | Lead versus nonlead ..... lead.. | 351–446 |
| | ..... nonlead.. | 159–223 |
| | ***AERD No. 3*** | |
| | Raw radial error corrected for standard altitude of 12,000 feet | 14 |

[2]The combined correlations for these tests include a composite of all of the forms indicated. The test number or numbers in parentheses constitute the smaller proportion of cases.

that a bombardier can make are mistakes in arithmetic calculations or failure to synchronize the bombsight in the relatively long period of time allowed on the bomb run. In contrast, combat bombing puts an emphasis on speed of reactions and adjustments, on rapid and accurate perception of the specific situation, and adaptation to it with a minimum of interference from nervousness and anxiety.

Although combat circular error was considered to be by far the most relevant criterion of the success of the bombardier, data were also collected and analyzed for a number of other criteria of success in combat operations. The combined values of these correlations are presented in table 4.11. These combined results include correlations based on ratings of general effectiveness, ratings of effectiveness on specific combat missions, selection for lead positions, and reclassification due to psychological breakdown, in addition to the combat circular error data reported in table 4.10. The tests with the highest correlations are not very different from those in the previous table — Reading Comprehension and Dial and Table Reading drop somewhat, and Spatial Orientation II (Pilotage) and Mathematics A move up on the list. Three tests, Numerical Operations, Finger Dexterity, and Aiming Stress continue to show small negative values which are not significantly different from zero.

In addition to these studies regarding the relative values of specific tests, studies were made of the value of the bombardier stanine as obtained at the time of original testing and classification. These correlations are shown in table 4.12 along with the correlation between the average circular errors made by a group of bombardiers when in training in the AAF Training Command and the average circular errors for combat bombs dropped by these same officers in heavy bombardment operations of the Eighth Air Force.

TABLE 4.12.—*Correlation between various average circular errors and the bombardier stanine in samples of bombardiers*

| | Number of individuals in combination | Combined[1] weighted r |
|---|---|---|
| Bombardier Stanine vs. ACE (AAF Training Command) | 10,432 | 0.054 |
| Bombardier Stanine vs. ACE (operational training) | 473 | .03 |
| Bombardier Stanine vs. ACE (combat) | 1,332 | .067 |
| ACE (AAF Training Command) vs. ACE (combat) | 495 | .064 |

[1]The composites include all of the data available in each of the categories.

Although the numbers of individuals involved are not sufficiently large to make a precise comparison possible, it appears that the bombardier stanine and the average circular error obtained in Bombardier Training Schools are about equally correlated with average circular error in combat operations.

The relationships between combat-bombing average circular error and the bombardier stanine and the three tests in the Air-Crew Classification Battery having the highest correlations with this criterion are illustrated in figure 4.18.

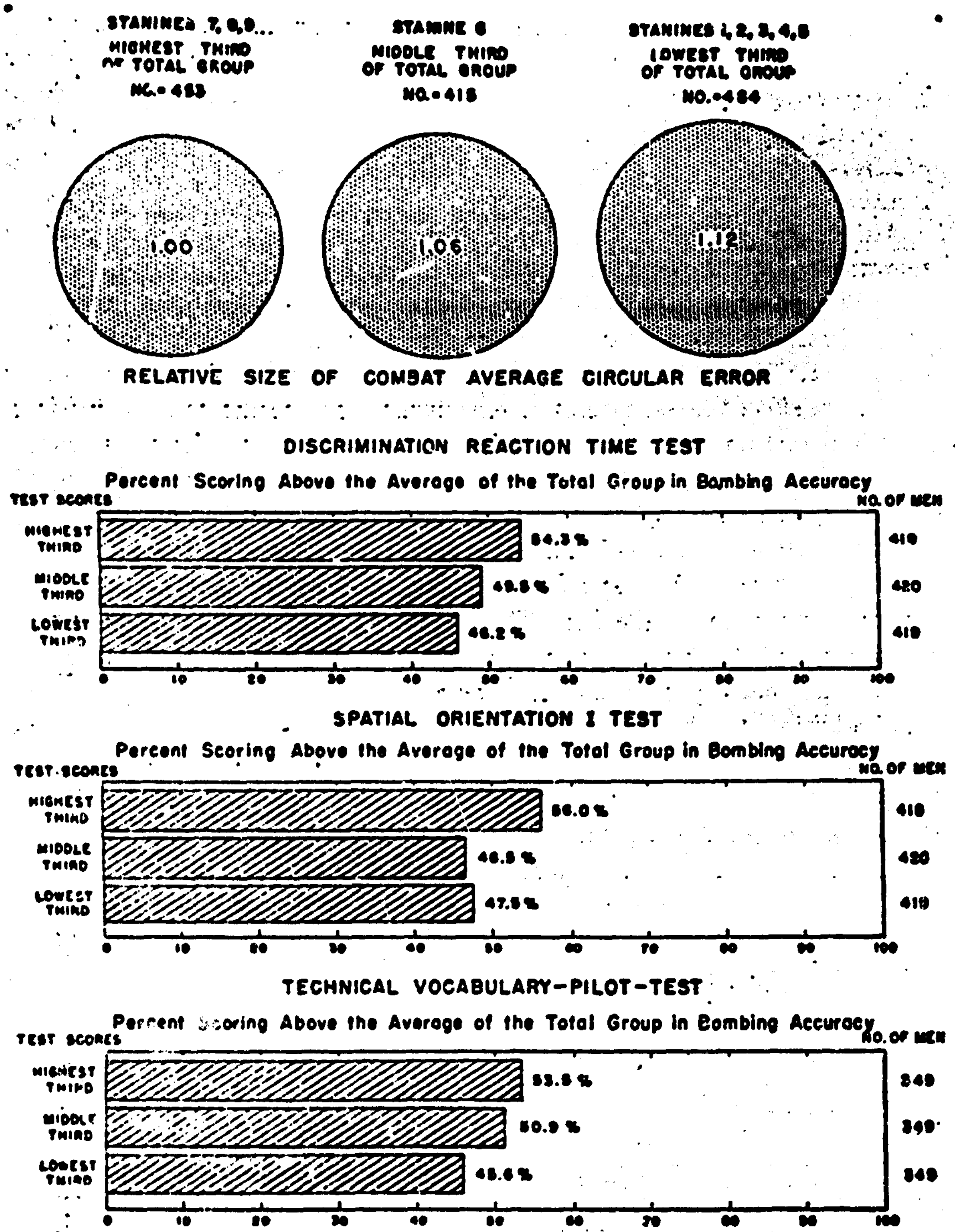



FIGURE 4.18.—Value of the bombardier stanine and the three air-crew classification tests that were best for predicting the relative accuracy of combat average circular error (lead bombardiers).

## *Bomber Pilots*

It has already been mentioned that the pilot had an important influence on the size of the average circular error in the AAF Training Command Schools. Since in this situation the bombardier students rather than the pilots of the training planes were the ones being evaluated, no studies were made of the relation between the test scores of the pilots and the average circular error of the bombs dropped by the bombardier students flown by him. One small study was carried out in the Fourth Air Force comparing scores on the air-crew classification tests with three criteria of success as heavy bombardment pilots in operational training. The

TABLE 4.13.—*The relative value of Air-Crew Classification Tests for predicting instructor's ratings, link-trainer grades, and crew average circular error of B-24 airplane commanders in operational training*

| Tests | Test No. | Number of individuals in combination | Combined weighted r |
|---|---|---|---|
| Mechanical Principles | CI903A | 131 | 0.18 |
| Mathematics A | CI702F | 104 | .15 |
| Biographical Data (N) | CE602D | 104 | .10 |
| Two-Hand Coordination | CM101A | 135 | .10 |
| Numerical Operations I | CI702B | 134 | .09 |
| Biographical Data (P) | CE602D | 104 | .09 |
| Mathematics B | CI206C | 104 | .09 |
| Aiming Stress | CE211A | 131 | .08 |
| Reading Comprehension | CI614G | 131 | .06 |
| Spatial Orientation II | CP503B | 137 | .06 |
| Spatial Orientation I | CP501B | 137 | .06 |
| Dial and Table Reading | CP622–21A | 137 | .06 |
| Numerical Operations II | CI702B | 134 | .05 |
| Rotary Pursuit with Div. Att. | CP410B | 106 | .05 |
| General Information (P) | CE505D | 104 | .04 |
| Finger Dexterity | CM116A | 137 | .02 |
| Speed of Identification | CP610A | 134 | .01 |
| Discrimination Reaction Time | CP611D | 137 | .00 |
| General Information (N) | CE505D | 104 | —.02 |
| Complex Coordination | CM710A | 137 | —.09 |

TABLE 4.14.—*The relative value of the Air-Crew Classification Tests in samples of bomber pilots for predicting combat ratings, bombing accuracy, reclassification, accidents and importance of positions held*

| Tests | Test No. | Data[1] combined | Approx. number of cases in combination | Combined weighted r |
|---|---|---|---|---|
| Mathematics B | CI710A, (CI206B, CI706A)[2] | 1–12 | 1,390 | 0.166 |
| Complex Coordination | CM701A | 1–12 | 1,390 | .151 |
| Mechanical Information | CI905A | 10–12 | 590 | .099 |
| Discrimination Reaction Time | CP611D | 1–12 | 1,270 | .098 |
| Spatial Orientation II | CP503B | 1–12 | 1,390 | .098 |
| Reading Comprehension | CI614G, (AC10D2) | 1–7,10–12 | 1,300 | .091 |
| Mechanical Principles | CI903A, (AC10D6) | 1–7,10–12 | 1,300 | .083 |
| Spatial Orientation I | CP501B | 1–12 | 1,390 | .078 |
| Technical Vocabulary (N) | CE505C | 1–12 | 1,390 | .074 |
| Two-Hand Coordination | CM101A | 1–4,8–12 | 940 | .064 |
| Speed of Identification | CP610A | 1–12 | 1,420 | .062 |
| Dial and Table Reading | CP622–21A, (CP621A, CP622A) | 1–12 | 1,420 | .053 |
| Technical Vocabulary (P) | CE505C | 1–12 | 1,390 | .053 |
| Finger Dexterity | CM116A | 1–12 | 1,390 | .050 |
| Rotary Pursuit | CM803A | 10–12 | 590 | .043 |
| Numerical Operations | CI702B | 1–12 | 1,420 | .034 |
| Mathematics A | CI702E | 1–12 | 1,420 | .033 |
| Technical Vocabulary (B) | CE505C | 1–12 | 1,390 | .014 |
| Aiming Stress | CM211A | 5–12 | 1,010 | — 005 |

[1]Studies included in combination:

| | | Number of cases |
|---|---|---|
| | *Initial Survey* | |
| 1. | Ratings on pilots by squadron officers on general over-all effectiveness, ETO | 18– 48 |
| 2. | Ratings on pilots, MTO | 38–106 |
| 3. | Ratings on co-pilots, ETO | 26–180 |
| 4. | Ratings on co-pilots, MTO | 25– 72 |
| | *AERD No. 1* | |
| 5. | Ratings by staff personnel, 8AF | 130 |
| 6. | Radial error and percent hits, 8AF, May and June 1944 | 100 |
| 7. | Radial error and percent hits, 8AF, July 1944 | 106–110 |
| | *AERD No. 2* | |
| 8. | Analysis from strike photographs of percent of bombs falling within 300 feet of target for lead pilots with 4 or more usable strike photos. | 47– 64 |
| 9. | A converted score based on the radial error for lead pilots with three or more missions r biserials calculated from critical ratio data (*See* footnote 14). | 31– 46 |
| 10. | Pilots removed from flying status by Flying Evaluation and Reclassification Boards versus matched control, ETO. | 18– 35 |
| 11. | Bomber pilots involved in pilot error accidents versus lead bomber pilots, 15AF | 78–118 |
| 12. | Lead versus nonlead pilots | 172–289 |

[2]The combined correlations for these tests include a composite of all of the forms indicated. The test number or numbers in parentheses constitute the smaller proportion of cases.

correlations of the scores with ratings of proficiency, Link-Trainer grades, and average bombing circular error, and also the average correlation of each test with all three criteria are reported in Table 4.13.

The numbers are too small to provide more than a suggestion regarding these relationships. All of the values but two are positive. The two largest correlations are for Mechanical Principles and Mathematics A. For average bombing circular error the largest coefficients are for Spatial Orientation I (Target Identification), Mathematics A, and General Information (Pilot). The correlation of pilot stanine wit' this combination of criteria was found to be 0.15 for this sample.

Several studies of the correlations between criteria of success of bomber pilots in combat and the air-crew classification tests were made. The criteria used included ratings of effectiveness, reclassification due to psychological breakdown, accidents, selection for lead positions, and average circular error on combat bombing missions. In table 4.14 are shown the average correlations for all of these studies combined. The two tests with

TABLE 4.15.—*The relative value of the Air-Crew Classification Tests in samples of navigators for predicting combat ratings, bombing accuracy, reclassification, and importance of positions held*

| Tests | Test No. | Data[1] combined | Approx. number of cases in combination | Combined weighted r |
|---|---|---|---|---|
| Discrimination Reaction Time | CP611D | 1-14 | 1,640 | 0.099 |
| Technical Vocabulary (P) | CE505C | 1-14 | 1,640 | .080 |
| Spatial Orientation II | CP503B | 1-14 | 1,640 | .078 |
| Speed of Identification | CP610A | 1-14 | 1,640 | .066 |
| Mathematics B | CI710A, (CI206B, CI706A)[a] | 1-14 | 1,640 | .064 |
| Reading Comprehension | CI614G, (AC10D2)[a] | 1-10,12-14 | 1,520 | .060 |
| Technical Vocabulary (N) | CE505C | 1-14 | 1,640 | .052 |
| Dial and Table Reading | CP622-21A, (CP621A, CP622A)[a] | 1-14 | 1,640 | .047 |
| Mechanical Information | CI905A | 9-10,12-14 | 730 | .035 |
| Rotary Pursuit | CM803A | 3-14 | 1,530 | .028 |
| Two-Hand Coordination | CM101A | 1-14 | 1,640 | .025 |
| Mechanical Principles | CI903A, (AC10D6)[a] | 1-10,12-14 | 1,520 | .018 |
| Technical Vocabulary (B) | CE505C | 1-14 | 1,640 | .018 |
| Numerical Operations | CI702B | 1-14 | 1,640 | .016 |
| Spatial Orientation I | CP501B | 1-14 | 1,640 | .015 |
| Complex Coordination | CM701A | 1-14 | 1,640 | .009 |
| Mathematics A | CI702E | 1-14 | 1,640 | —.001 |
| Aiming Stress | CM211A | 3-14 | 1,530 | —.042 |
| Finger Dexterity | CM116A | 1-14 | 1,640 | —.057 |

[1] Studies included in combination:

*Initial Survey* — *Number of cases*

1. Ratings provided by squadron officers on general over-all effectiveness, ETO .... 26-115
2. Ratings provided by squadron officers on general over-all effectiveness, MTO ..... 11-13

*AERD No. 1*

3. Ratings by staff personnel, 8AF ..... 262
4.5. Radial error and percent hits, 8AF, May and June 1944 ..... 236
6.7. Radial error and percent hits, 8AF, July 44 ..... 180

*AERD No. 2*

8. Average mission rating by Wing and Group staffs, 8AF ..... 42-69
9. Same as 8 with more cases and more mission rating ..... 51-86
10. Ratings by supervisory officers on a five point scale on over-all effectiveness (of navigators with ten or more missions, 15AF.

r-biserials converted from critical ratios (*See* footnote 14)

11. Medical Disposition Board cases versus control, 15AF ..... MDB . 31-48; control . 56-84
12. Central Medical Establishment cases versus control, 8AF ..... CME . 35-63; control 36-62
13. Reclassified cases versus control ..... recl. . 12-20; control . 23-38
14. Lead versus nonlead ..... lead . 260-408; nonlead 125-201

[a] The combined correlations for these tests include a composite of all of the forms indicated. The test number or numbers in parentheses constitute the smaller proportion of cases.

the largest correlations are Mathematics B (Arithmetic Reasoning) and Complex Coordination. Only one test, Aiming Stress, shows a negative correlation with the combined criteria. This value is negligibly small. Many of the coefficients in this table are sufficiently large as to make it unreasonable to assume that no relation exists.

*Navigators*

It was believed that the academic and flying grades received by the navigators in training schools were more relevant to their subsequent combat assignments than those obtained for the other specialties. On the other hand, it was found to be very difficult to obtain a satisfactory criterion measure of the success of the navigator in the combat situation. The results of the studies made comparing air-crew classification tests with ratings of general effectiveness, ratings of effectiveness on specific combat missions, selection for lead positions, reclassification for psychological breakdown, and average bombing circular error in combat are given in table 4.15.

The tests showing the highest correlations with the combined criteria are Discrimination Reaction Time, Technical Vocabulary (P) (interest in aviation and related activities), Spatial Orientation II (pilotage), Speed of Identification, Mathematics B (arithmetic reasoning), and Reading Comprehension. Three tests show small negative correlations with this combination of criteria. These are Mathematics A (algebra), Aiming Stress, and Finger Dexterity.

## CONCLUSIONS

To solve the problems confronting the Army Air Forces in 1941 regarding the selection and classification of men for air-crew training, two examining procedures were developed. Initial selection was accomplished by means of an initial 150-item screening examination called the AAF Qualifying Examination. From a quarter to a half of the more than a million applicants were rejected because of low scores on this examination. This test provided a measure of proficiency with regard to comprehension of the types of materials typical of the training courses in these fields. In its later forms it also emphasized aptitude and motivation for air-crew and especially pilot training. The test was designed as a practical measure of proficiency and ample time, 3 hours, was allowed for its completion. An individual could try the test again after waiting 30 to 90 days if he failed to qualify.

The results reported in this chapter indicate that it provided an effective initial screen for sifting applicants at several hundred boards throughout the country. It should not be used in situations where it is practical to examine all applicants with the more comprehensive Air-Crew Classification Test Battery.

The Air-Crew Classification Test Battery consisted of approximately 20 tests. Of these, 6 were apparatus tests of coordination and speed of decision, and 14 were printed tests measuring intellectual aptitude and abilities, perception and visualization, and temperament and motivation. Stanines based on weighted combinations of the scores from these various tests were computed for each individual separately for the various air-crew specialties. These stanines were found to have high predictive value for pilot and navigator training success and were also found to be definitely correlated with criterion measures of success in operational training and in combat. The qualifying standards were raised successively during the war to improve the efficiency and quality of training and to raise the standards of proficiency and potentiality for development in the personnel sent into combat units. After October 1944 only men with stanines of 7, 8, and 9 were qualified for pilot, navigator, bombardier, or radar-observer training.

More than 600,000 men were given this comprehensive battery of tests and the entire group were followed up in the training schools through use of a machine-records system. The outstanding success of the procedures developed was evidenced by the adoption of all of the available tests in this battery by the Royal Air Forces in April 1944 and by the French Air Forces in July 1944. In response to subsequent requests the tests have also been made available for the use of a number of other foreign governments.

A number of surveys made in the combat theaters indicated that the officers supervising the conduct of combat operations were surprised and enthusiastic regarding the continued improvement observed in the quality of the air-crew replacements received. Although much information has been collected and effective procedures have been developed, changing requirements in aerial combat operations will continually provide new problems. Therefore, a strong research and development program based on the important advances made during the war period is essential to the continuation of effective selection and classification work.

# CHAPTER FIVE

# Studies on the Problem of Evaluating Proficiency

## INTRODUCTION

Any research program on problems of selection, classification, and training must become involved with general problems of evaluating proficiency. One of the first studies carried out in the Aviation Psychology Program was a preliminary survey of the procedures used in evaluating proficiency of men being trained in various types of air-crew specialties. Fortunately, this survey indicated that the procedures in use in both the pilot and navigator training schools were sufficiently reliable and valid to warrant the expenditure of considerable time on the development of tests and other procedures to predict these proficiency measures. The proficiency measure which appeared to be clearly superior to other measures available for these types of training was graduation or elimination from the training schools. Unfortunately, a similar situation did not exist in either the bombardier or flexible-gunnery schools. This made the problem of research in these areas much more difficult.

As selection and classification tests were developed which provided increasingly accurate predictions of success in training, a need was felt for additional investigation and possible refinement of these measures of success. Various studies were carried out by the personnel of the various Psychological Research Units and in the first month of 1944 Psychological Research Projects were established for the various specialties to provide a more systematic program for the development of proficiency measures and the conduct of training research.

## STUDIES OF PILOT PROFICIENCY

### Evaluations of Grading and Checking Procedures[1]

As mentioned above, the decision as to whether a student should continue on in pilot training and be graduated or be eliminated because of failure to show satisfactory progress or for other deficiencies was carefully made. In addition to the evaluation of the instructors on daily grade slips, the students in primary flying schools were given an average of four check rides by check pilots. In doubtful cases the students sometimes received

[1]The various studies reported in this section were conducted in a number of units by personnel of the Aviation Psychology Program, including M. P. Crawford, J. T. Dailey, J. T. Cowles, R. J. Keller, F. N. Schrader, H. F. Summers, R. W. Loder, R. H. Henderman, S. H. Harris, K. C. Anderson, N. E. Miller, S. E. Ericksen, C. P. Gershenson, W. W. Ismael, R. E. Showalter, and J. R. Rohrs.

as many as 12 check rides. Almost never was the student eliminated unless at least three different pilots, his instructor and two check pilots, agreed that his flying proficiency and general promise were definitely unsatisfactory.

The general adequacy of this criterion was further indicated by the consistently substantial correlation obtained between it and the pilot stanine. From the size of these values it can be estimated that the biserial correlation coefficients between these final decisions to pass or fail a student and a very accurate measure of the pilot's proficiency and general promise obtained from an indefinitely large number of similar check rides could not be much less than 0.70. This indicates that the reliability of the procedures on which this decision is based is not much less than 0.50.

That the value could not be much greater than this was indicated by consistently low correlations found between all types of measures of flying proficiency obtained at one level of training with similar measures obtained at a later training school. These correlations were usually found to be between 0.20 and 0.35. In a typical study, where grade on General Pilot Ability in primary training schools was compared with the similar ratings in basic schools for 2,905 students, the correlation coefficient was found to be 0.27.

The implications of this can be seen by assuming that primary and basic flying are so different that the correlation coefficients obtained between the average of an indefinitely large number of check ride grades in primary is related to the average of a similar measure for basic to the extent of only 0.50. The usual statistical formula when applied indicates that under these circumstances the agreement between the grades assigned in each school and the "true" values for that measure obtained as described above could not be much *greater* than 0.70. In combination with the inference in an earlier paragraph that the value could not be much *less* than 0.70, this appears to suggest that the reliability of these procedures is in the vicinity of 0.50.

In addition to this chance element the criterion of graduation-elimination also possessed the other defects common to subjective ratings of this type. There was great variation in standards from instructor to instructor and from school to school. In a study of the elimination rates in primary school classes 43-G to 43-K, including about 50 different schools, it was found that for one class in one school the elimination rate was as low as 10 percent while in another class in a different school the elimination rate ran as high as 50 percent. These differences could not be accounted for by sampling fluctuations and the quality of students received.

The subjective character of these ratings was also indicated by the high intercorrelations obtained for grades or ratings supposed to indicate the individual's proficiency on various different maneuvers when the ratings were all made by the same instructor at the same time. Similarly high correlations were also found by ratings made by the same instructor at successive stages of flying instruction. Grades were typically given on a scale from A to F with most of the students obtaining grades of C. Each grade below C was amplified by a written statement describing the student's weaknesses. It was found that these comments concerning specific deficiencies on particular maneuvers were less subject to "halo" effects than the usual grades and ratings. The use of comments on the daily grade slips as a source of data regarding individual proficiency was impractical as a general procedure. However, studies of these comments indicated that a check list derived from them would probably be a useful addition to the existing grade slip.

In a study of the possible usefulness of subjectively graded check rides, the Pilot Project working cooperatively with the two-engine group at the Central Instructors School at Randolph Field constructed a more detailed and analytical grading check sheet than the one formerly used for grading students. This check ride covered the more important maneuvers included in the two-engine instructors school curriculum. After a preliminary tryout 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 period of from 2 to 4 days. The raters were 20 of the most experienced check pilots in the two-engine group and had never flown with the students before. The ratings were made after approximately 15 hours of flying instruction at the Central Instructors School. The rating procedures were typical. The check pilot circled a number from one to five indicating the grade on such items as smoothness on controls, landing roll, and turn-on approach. For certain other items such as gear check on approach he circled either "yes" or "no."

### Objective Measures of Flying Skill[4]

The reliability or agreement between the total scores obtained from all 14 maneuvers by the two independent check riders was found to be 0.25 when expressed as a product moment correlation coefficient. The correlation of a single check ride score with the instructor's final grade in the course was 0.34. To provide an estimate of the reliability of final course grades it can be assumed that the check pilots and the instructors are all measuring the same function, but with varying proportions of chance elements introduced by variations in conditions and individual performance. The application of the usual formula to these figures indicates that the correlation of a single check ride score of this type with the average score from an indefinitely large number of similar check rides would be 0.50. Similarly, the correlation between the instructor's final grade and

[4]Work on the development of objective measures at the primary level was principally carried on by N. E. Miller, R. P. Youtz, M. F. Connery, J. F. Kamman, J. R. Rohrs, J. J. Wasterman, J. K. Hemphill, W. W. Ishmael, R. E. Showalter, W. G. Matheny, and C. P. Gershenson. Advice as a special consultant was provided by P. J. Rulon.

The work at more advanced levels was carried on primarily by N. E. Miller, S. C. Erickson, J. K. Hemphill, W. G. Matheny, J. R. Rohrs, M. F. Connery, W. V. Hagin, W. W. Ismael, B. N. Henderson, J. O. Bean, E. Glaser, H. H. Hagy, J. G. Gleason, A. C. Tucker, W. E. Galt, I. Robbins, J. W. Stratton, G. L. Heathers, and R. M. Rust.

the average from an indefinitely large number of check rides would appear to be about 0.70, indicating a reliability for the final course grades of approximately 0.50.

These studies of existing grading and rating procedures and a few exploratory efforts to improve these procedures led to the conclusion that much greater objectivity would have to be obtained if flying skills were to be adequately measured. Accordingly, the Chief of the Psychological Branch prepared a directive entitled "An Objective Scale of Flying Skill" which was sent by Headquarters Army Air Forces to the Psychological Research Project (Pilot) through Headquarters AAF Training Command. This directive was dated 25 May 1944. Two of the principal paragraphs are quoted below:

> Numerous previous efforts have been made to improve grading procedures by such organizations as the Royal Air Force, the Civil Aeronautics Authority, and the United States Navy. Examination of the methods developed by these groups reveal that the procedures developed thus far possess two serious defects: (a) they are subjective, that is, the rating received is dependent on the standards and personal estimates of the individual doing the grading, and (b) the procedures have not been standardized so that the flying skill of the individual tested can be expressed in terms that are easily understood in relation to the performance of other individuals.
>
> It is believed that a satisfactory scale for the measurement of flying skill can be developed by standardizing a large number of specific test elements. These should be carefully selected on the following basis: (a) each item should describe a specific, well-defined maneuver, giving precise directions to be followed in performing it, (b) each item must be scored as either passed or failed and the task must be so defined that all raters observing the performance would agree as to whether it should be scored as passing or failing, (c) each item must measure an element of flying skill that is sufficiently stable so that performance on the assigned task will be consistent from one test to another, (d) each item must be shown experimentally to cover an element of flying skill which is related to later success in flying training, (e) the group of items selected at any particular level of flying instruction must sample all essential phases of flying skill.

Early efforts to develop objective measures of flying skill clarified the importance of controlling all sources of variation in score which were unrelated to the student's skill in the particular task being studied. The three main sources of variation which must be controlled in this situation were found to be:

1. The judgment of the examiner observing the student's performance of the assigned task or observing a record of the student's performance. The making of this judgment must be standardized in such a way that practically all observers can agree precisely on the scoring category to which a specific observed performance should be assigned.

2. The student's interpretation of the nature of the task. It is essential that in performing a complex task in flying the student know exactly what he must do to achieve the maximum possible score. He must also know the relative losses in his score which will accompany partial failures on various aspects of his task. Unless variations in the interpretation of the task are controlled by defining the task clearly and completely, two stu-

dents with the same ability may get different scores because they believe the examiner to regard errors on two related aspects of the maneuver as of differing relative importance.

3. The nature and difficulty of the task itself must be controlled for various students. Unfortunately, planes vary widely in their flying characteristics and the calibration of their instruments. Changes in wind, turbulence, density, and visibility tend to introduce large differences in the difficulty and nature of the tasks the individual is called upon to perform.

In developing the specific objective measures the Pilot Project concentrated on two crucial areas: (1) the early part of primary training, and (2) instrument flying. The early part of primary training was critical because it is at this point that most eliminations occur. Instrument flying was selected because of its increasing importance and also because the use of instruments adapts itself well to objective measurement.

The first step in developing standard objective measures of flying skill was the determination of the types of observation which check pilots could make with a reasonable degree of objectivity. It was found that check pilots could agree as to whether the air speed increased or decreased by more than a fixed amount over a specified period of time, or during a particular maneuver. In general they were found to agree on any similar maximum or minimum reading of instruments. They could also provide objective records as to whether or not a plane "bounced more than 3 feet," "landed in the first third of the field," or whether a sufficient angle of bank was achieved so that the cabane strut was parallel with the horizon at some time during a particular turn.

On the other hand, they found it very difficult to give an objective report as to whether or not the student "looked around at least 50 percent of the time." "Satisfactory ground pattern," "maintains altitude in turns," and "excessive use of the throttle" were found *not* to be sufficiently well defined statements to enable observers to give an objective report. On the basis of considerable informal investigation and a few preliminary studies the general conditions necessary for obtaining an objective report from the check pilot were quite satisfactorily determined.

More than 500 measures of flying skill were experimentally investigated at the various levels of training. In developing these items three different methods of scoring instrument deviations were compared. These included (1) the time-sample method in which the check rider read the deviations from the correct altitude heading, etc., at specified intervals, throughout the maneuvers, (2) the range method in which he scored the difference between the highest and the lowest reading on an instrument during the maneuvers, 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 instruments.

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 the instructor's grade than the range method. Furthermore the time-sample method was harder to administer. The comparison between the range method and the limits method were inconclusive but tended to favor the range method.

The next step in the development of the measures proved to be more difficult. This was the matter of controlling the individual's interpretation of the task. It was found, for example, if one merely asked the student to make a steep turn of 90 degrees and then obtained an objective record from the check pilot as to what he did, a very unsatisfactory measure of his ability to perform a steep turn was likely to be obtained. Students of some instructors would use an angle of bank of between 30 and 45 degrees instead of the conventionally accepted 60 degrees. Some would stress coordination and avoidance of slips and skids, others would stress maintaining altitude and still others would stress making exactly a 90 degree turn. If the scoring system placed greatest weight on gain or loss of altitude many students were unfairly scored because they could easily have controlled altitude better if they had realized that that was the main element of the maneuver.

After considerable work with this particular maneuver, an item was developed which penalized the students one point for each 20 feet gained or lost from the starting altitude during the maneuver and which also provided two conditions to be fulfilled, failure in either of which meant a zero score on the whole maneuver. The first condition was that the student reach a 60 degree bank in each turn as indicated by the cabane strut being parallel with the horizon and the second condition was that the final roll-out should be within plus or minus 45 degrees of the starting heading. The important consideration was found to be that the student know precisely what the scoring system was so that variations in the task which he set for himself and in his approach to this particular maneuver would be minimized as much as possible.

Developing items in which the task would be stated so objectively that the students would all interpret the task similarly and attempt the maneuver in such a way that their true relative ability to perform this maneuver in comparison with the other students was reflected in their score was found to be very difficult. However, very substantial progress was made in constructing such items.

It was only after the first two steps had been fairly adequately carried out for a large number of items that the very great influence of the third source of variation—the variation in condition of the plane and other equipment and in weather conditions—was fully recognized. Studies of the agreement of scores on objective check rides made on the same day in the same plane by a particular student and the relative lack of agreement between the check rides made by the same student in a different plane on the following day or the following week made it very clear that for certain types of measures, equipment, weather, and probably variations in the individual's ability itself were such as largely to obscure differences in the individual ability of students.

Fortunately this was not true of all types of measures. An objective check ride of single-engine instrument flying skill at the basic level of flying training prepared and tried out by personnel of the Pilot Project in September 1945 showed a substantial amount of agreement between total scores for the same standard check ride given on successive days by different check pilots. The correlation coefficient between these scores for 55 students was 0.46. The correlation between the sum of the total scores on the 81-item scale for the 2 check rides and an over-all rating of instrument-flying proficiency made immediately before these rides by the student's instructor was 0.51 for the same group of students. These figures provide good evidence that it is feasible to develop a reliable and valid objective scale of instrument flying skill.

It is believed that very important progress has been made toward developing satisfactory objective measures of flying skill. Further research in this area should make it possible to develop practical measures for standardization and use in all pilot-training schools.

### Tests of Flying Information*

It was found very early that there were practically no eliminations in pilot training schools for reasons other than flying deficiencies. The students were sometimes held over because of deficiencies in academic ground-school courses but rarely were they eliminated for such deficiencies. This led to an emphasis on the aquisition of flying skill both in training and in prediction of success as a military pilot. Later studies, particularly those carried on in combat theaters, emphasized the importance of judgment and flying information in the pilot's combat work. Objective tests of flying information of the multiple choice types were prepared by the Psychological Research Project (Pilot) to select pilots for special assignments such as lead-crew commanders or instructor training. In developing these tests, an attempt was made to cover points of information which were important for the pilot to know in carrying out his operational assignments, regardless of whether or not they were adequately covered in the training-school curriculum.

The increasing emphasis on instrument flying led the Pilot Project to develop a test of Pilot Instrument Flying Information in cooperation with the AAF Instrument Flying Standardization Board. A 150-item multiple choice test was prepared including items relating to principles of operation and limits of the flight instruments, radio-compass operation and procedures, AAF instrument-approach system, radio range, flight rules, instrument

*Work on these tests was done at Psychological Research Project (Pilot) by I. Robbins, R. Levine, W. E. Harris, S. Foulke, and W. H. Wiesenberg and under the general supervision of N. E. Miller.

procedures, instrument-flying problems, weather fronts, thunder storms, icing, wind, fog, atmospheric stability, weather reports, flight planning, weather maps, weather flight procedures, and judgment based on these facts.

Use of this test was established by the AAF Training Command as one of the requirements for securing an instrument card which authorized the pilot to fly under instrument conditions.

## STUDIES OF NAVIGATOR PROFICIENCY[4]

### Evaluation of Grading Procedures

Unlike the pilot training school, the navigator training school gave substantial weight to ground-school grades to determine whether an individual should be eliminated or graduated. Partly for this reason it appears that final grades in navigation schools were much more reliable than grades in pilot training schools. A study of reliability of grades for 308 graduates of class 44-2 at Selman Field indicated a fair degree of reliability for measures entering into the final grade. These reliability coefficients were: for lecture grade based on daily class-room performance, 0.82; for examination grade based on weekly examination, 0.90; and for flight-mission grade based on approximately 20 navigation missions, 0.72. In all cases these coefficients were obtained by comparing grades obtained on odd weeks with grades obtained on even weeks and using the usual formula to correct for the full length of the course.

In this same group, the flight final average grade correlated with the lecture final average to the extent of 0.46 and with the examination final average to the extent of 0.49. The lecture final average and the examination final average correlated 0.78.

Two points should be noted in connection with these correlation coefficients. First, some of the measures, particularly the flight and lecture grades, tended to be subjective in nature. Therefore, since they were made in many instances on odd and even weeks by the same instructor, they tend to give an overestimate of the agreement of these grades with a really independent appraisal. The second point is that the three types of grades cannot be expected to be truly independent for the same reason. The instructors usually were informed of previous examination grades at the time the other grades were assigned and were probably influenced by this knowledge.

Since recommendations for elimination of students tended to be made on the average about the sixth week, the actual reliabilities of the grades on which the decision to eliminate or continue students were made were much lower than those previously cited. The reliabilities for grades at the time of elimination were estimated to be approximately as follows: lecture grades, 0.67; examination grades 0.80; and flight grades, 0.40.

The Psychological Research Project (Navigator) cooperated with the Research and Development Department of the AAF Instructors School in

[4]The work reported in this section was done chiefly by L. F. Carter, J. M. Christensen, L. N. Wiley, R. M. Rust, E. Erskine, R. T. Joseph, R. Glaser, S. T. Friedman, W. A. Zielonka, and H. A. Smith.

developing a new flight-mission grading system. This new system standardized procedures to a much greater extent than had been true formerly. After each flight check the instructors discussed the flight and agreed upon minimum requirements for the particular mission. The students' logs were analyzed in considerable detail and graded according to a standard grading form. Unfortunately, no studies of this new grading system have been made, since it was not installed at Selman Field until August 1945.

### Objective Aerial Measures of Navigation Skill

The preceding discussion has indicated that for navigation training just as for pilot training there was an important need for objective measures of the navigator's skill as shown on flying missions. To meet this need, Psychological Research Project (Navigator) developed procedures for obtaining objective measures of flight performance in the dead-reckoning phase of air navigation. Dead-reckoning was chosen since it is basic to all other phases of navigation.

In general, the problem of developing objective measures for measuring the flight performance of navigators involves the same problems of control of the sources of error as were encountered in developing objective measures for pilots. Because of the nature of the navigator's task it was possible in this case, however, to require that the navigator make a complete record of his work in the form of a log which could then be evaluated after the mission was completed. Thus, it was possible to obtain an accurate statement of the work done by the individual navigator and what was needed was a standard for comparison indicating precisely how his record would read had he done perfect work all the way through.

In this situation also, there was the problem of controlling the interpretation of the relative importance of the various aspects of the task on the part of the individual navigator. Finally, there was the same general problem of the variation introduced by differences in planes, instrument characteristics, pilot, weather conditions, and so forth.

In order to have a standard for comparing students' logs which would apply to a large number of individuals and also in order to control air conditions as much as possible, a procedure was developed for flying missions in formation. By flying formations in which the planes were reasonably close together, all 40 students in one flight would be flown over the same course at the same speed and under the same weather conditions. To standardize student procedure as much as possible, explicit instructions were given to the students before each flight mission and the instructions were repeated before take-off.

The students performed "follow-the-pilot" procedures throughout the mission and made 5-minute entries of instrument readings in their logs. Winds on two headings were computed after each turn. The dead-reckoning positions which were required for each turn were determined using "follow-the-pilot" procedures. Compass deviations were obtained by astro-compass on each heading. On the last leg of the mission the student precomputed a dead-reckoning position for 35 minutes after the turn. Thirty minutes after the turn he made out an Alter-Course Slip and handed it to the pilot. The Alter-Course Slip contained spaces in which the student filled in his dead-reckoning position computed for 35 minutes after the turn, the magnetic heading from his dead-reckoning position to destination, the ground speed for his new heading, and the estimated time of arrival at destination on this new heading. By accomplishing these requirements the student was in essence shifting from "follow" to "direct-the-pilot" navigation.

To obtain standards with which to compare the students' records, experienced lead navigators rode in the lead ship of each formation. These navigators directed the lead ship (and the formation) over the intended course. These lead navigators kept precise logs for use as standards for grading the students' logs. Approximately 150 students were flown on 12 formation missions during October and November of 1944.

It was found possible to standardize the course, the speed, and of course the weather was the same for the various ships in the formation which flew together over the course. Similarly, the techniques employed by the students were sufficiently uniform to constitute no great problem in obtaining an objective evaluation of their performance. The development of precise standards with which to compare the students' logs proved more difficult than was anticipated. In some instances the readings and computations of the lead navigators were considerably at variance with those obtained by the students or by the project's observers and supplemental data had to be applied to establish the standards.

By scoring each leg of the mission in such a way that errors on one leg did not affect errors on another, four independent samples of performance were obtained under the conditions of the mission. The methods used also provided a separate record of error with respect to each of the navigational tasks which made it possible to analyze the components of the task of navigation. By comparing errors made by students on legs one and four with those made by the same students on legs two and three, coefficients of reliability were obtained for the various navigation entries. Appreciable reliability coefficients were found for such measures as true air speed, drift, deviation, and distance-off for the various legs of a single mission. However, in the first study the coefficients obtained by comparing errors on one mission with errors on another mission tended to be around zero, with a large proportion of them negative.

An analysis of the results indicated that a substantial amount of variation in errors was associated with the plane used. The drift meter used also seemed to contribute to the student's accuracy score. On the other hand, position in formation did not seem to be of any importance. The procedures were redesigned to make use of larger planes, TC-47's, which made it possible for an observer to accompany each group. The alignment of the drift meter was made the responsibility of the students who were to use them. Each student was also made responsible for obtaining his own readings with the astro-compass. These changes and a generally more rigorous control of position were expected to improve the accuracy of the measure obtained by the individual.

An analysis of the results obtained by students on missions flown from May to September 1945 indicated that some improvement had probably been effected. This was suggested by the fact that the tendency for the students to make errors in obtaining drift-on-course and in obtaining true air speed tended to show a small amount of consistency from one mission to another. On the other hand, deviation showed only a very small amount of consistency and distance-off showed practically no consistency from one mission to another.

Although the very narrow range of individual differences in ability to do navigational tasks could be expected to reduce the coefficients of reliability or consistency found in this type of study, these studies indicate both the need and the difficulty of evaluating an individual's proficiency in actually carrying out the task as a navigator on a flying mission. It is clear that much additional research on the nature of the navigator's task is required before procedures currently in use for selection and training can be precisely evaluated. The studies just described defined the problem and indicated some of the most promising procedures for the solution of the many specific problems arising. Further research along these lines should make possible the development of procedures for a standardized objective evaluation of navigational skill.

## TESTS FOR USE ON THE GROUND

As previously indicated, much more attention was paid in navigation training schools to abilities which could be performed on the ground than was the case in pilot training. The Psychological Research Project (Navigator) developed a series of ground measures consisting of a special printed test, a motion-picture test, a photographic test of pilotage, and an instrument phase check.

For the purpose of selecting superior navigators to be placed with superior personnel in the other specialties as potential lead crews for operational training, a printed test of navigator proficiency was prepared. Emphasis was placed upon the aspects of navigation related to flight. Photographs and diagrams of aircraft instruments were used to make the problems as realistic as possible and diagrams, tables, and work sheets requiring the individual to perform a complete unit of a navigational procedure were used. This printed test required three hours and was designed to emphasize those aspects of navigational skill most needed in actual combat operations. The intercorrelations obtained on the several parts of this test were very low. Only 2 of the 36 intercorrelations for the nine parts were above 0.30.

The correlations of students' scores on this test with ground and flight-mission grades were found to be very low. These coefficients were below

0.20 for a sample of 74 graduating cadets. This tends to confirm the point raised earlier that the agreement between the various subjective grades was probably spuriously high.

A tentative form of a pilotage test using still photographs was prepared. In this test the navigation student was asked to view a picture for one minute and he was then given 30 seconds to locate the position of this picture on a map and mark the corresponding letter on his answer sheet. A number of pictures were shown in this way with each succeeding picture representing a position a certain number of minutes flight from the last picture. After several trials of this sort, a new series of pictures involving a different mapped area was presented. Although this test was never evaluated it appeared to the navigators who examined it and took it that it was a fairly good approximation to the task of pilotage as encountered by the navigator on a flight mission.

In cooperation with the Psychological Test Film Unit the Psychological Research Project (Navigator) developed a motion picture test known as the Navigation Proficiency Test—Map Reading and Dead-Reckoning. The continuous film for this test was taken from an altitude of 12,000 feet with the camera tipped slightly forward so that a view of the terrain directly below and slightly ahead of the plane would be included.

Under these conditions the film gave a view of the terrain about 4 miles wide. Superimposed on the lower part of this terrain view were representations of the navigation instruments so the examinee could perform dead-reckoning navigation or pilotage navigation. The instruments gave compass heading, indicated air speed, and drift correction while numbers appeared on the film giving altitude and temperature.

On the basis of certain preliminary tryouts it appeared that the reliability of this test comparing first half with second half was very low. This was due in part to the accumulative nature of errors during an entire half of the test. Like the pilotage test the motion-picture test did not present a sufficiently large portion of the terrain to permit an unrestricted view around the plane and this tended to make pilotage difficult. There were also certain technical difficulties such as the tendency for the terrain to move jerkily under the grid lines and for the markings on the air speed meter to appear hazy. It is believed, however, that the technique is a promising one and should be further exploited.

## STUDIES OF BOMBARDIER PROFICIENCY MEASURES

### Studies of Bombing Circular Error[5]

In previous sections it was pointed out that the bombardier elimination rate was much smaller than that in either pilot training or navigation training. Most studies of the relation between graduation-elimination and other criteria indicated that this measure was not one in which a great deal of con-

[5] These studies were performed chiefly by E. H. Kemp, A. P. Johnson, C. W. Brown, W. W. Grings, M. Rokeach, L. F. Carter, R. W. Faulkner, R. M. Bellows, C. E. Wilder, and M. J. Warrick.

fidence could be placed. The criterion which appeared to be of most importance for use in measuring bombardiers' proficiency either in connection with selection experiments or training experiments was the circular error of the bombs dropped by the individual bombardier students. Ordinarily the bombardier dropped at least 150 bombs during his training in the AAF Training Command schools. This measure showed very little relation to either final grades or records of graduation-elimination.

A number of studies were made of the reliability of the circular error score. It was found that there was fair agreement between the error scores for odd and even releases on the same mission. However, the relations between the circular errors made on different missions were uniformly quite low. The results of all of the studies on different samples which were available in the files at Headquarters Army Air Forces are shown in table 5.1. The weighted average for these reliability coefficients, which is 0.18, is quite low. This indicates that the bombing circular error as obtained in the Training Command has a correlation of only 0.42 with the circular error which would be obtained from an indefinitely large number of similar bombing missions. This latter value, therefore, represents the highest value a correlation coefficient between this criterion and an aptitude test or any other independent measure could reach. If one of the air-crew classification tests had a correlation with the average circular error obtained from an indefinitely large number of bombing missions equal to 0.40, the correlation of this test with the circular error scores ordinarily obtained in bombardier schools would be 0.17.

Several studies were made in an effort to determine which factors other than the bombardier's ability were influencing the circular errors in order that these extraneous factors could be controlled so that the circular error would more accurately reflect the ability of the individual bombardier. Studies of San Angelo records indicated that differences which could not be explained on the basis of chance variations were obtained for different altitudes, different rated degrees of turbulence, different bombing ranges, different bomb sights (specifically the M-7 and the M-9), between practice and record missions, and between day and night practice missions. The differences due to drift, temperature, barometric pressure, ground speed, and later

TABLE 5.1.—*A summary of data regarding the reliability coefficients obtained for the bombing circular errors of bombs dropped on separate missions in AAF training command schools*

| Type of measure | Station | Class | N | Reliability coefficient |
|---|---|---|---|---|
| Odd and even missions | Barksdale | SE41-A | 41 | −0.01 |
| do | do | SE41-B | 70 | .67 |
| do | do | SE41-C | 70 | .27 |
| do | San Angelo | 43-3 | 129 | .03 |
| Average intercorrelation for 5 altitudes | do | 43-3 | 129 | .04 |
| Odd and even missions | Victorville | 43-1,2,3,4 | 94 | .06 |
| Odd and even missions (within altitudes) | do | 43-14,15 | 208 | .18 |
| Record missions with combat missions | Deming | 43-5,6 | 138 | .21 |
| Odd and even missions | Big Spring | 645 | 94 | .03 |
| Missions in first half with second half | do | 645 | 94 | .16 |
| Odd and even missions (first half only) | do | 645 | 94 | −.03 |
| 4000 feet missions with 7000 feet missions | do | 645 | 89 | .18 |
| Weighted average for combined samples of different individuals | | | 844 | .18 |

as compared to earlier releases on each mission were sufficiently small to have been merely sampling fluctuations.

In a study of Victorville bombing records the factors found to be important in influencing average circular error were: altitude, bombardier, pilot, sight number, mission number, time of impact, time for sighting, drift angle, flight type (M-7 or D-8), and type of mission (whether combat or record). These investigators reported that the following factors appeared not to contribute significantly to the circular error scores: plane number, trail, temperature, target number, release number, change of dropping angle, compass reading, track, and approach control. In a later study, reported in September 1943, it was found that when the effect of altitude was controlled the following factors no longer appeared related to circular error: target number, temperature, true air speed, time of impact, time for sighting, ground speed, and drift angle.

Under wartime training conditions it was not possible to standardize the training situation well enough to control the effects on the bombardier's circular error of these many extraneous factors sufficiently well so that the circular error would be a reliable indicator of his bombing skill. Therefore the Psychological Research Project (Bombardier) spent a considerable portion of its time developing proficiency tests including both printed tests using objective multiple choice test items and aerial checks utilizing check lists and similar observations and ratings. These are described in the next section.

## Other Proficiency Measures*

In the summer of 1944, Psychological Research Project (Bombardier) tried out several preliminary forms of an aerial phase check. In the first form the examiner was asked to rate the performance for each of a number of items on a scale from one to five. These subjective ratings turned out to be inadequate because of difficulties usually associated with such ratings and it was decided to attempt to make the items more objective in their scoring.

On the basis of this experience a new form was developed and this final phase check was supplied to the schools and to the Bombardier Standardization Board in November 1944. After some revision this was published in February 1945 as the Standardized Bombardier Phase Check—Air. All items on this form were scored either as plus, minus, or zero. Most of the items on this particular check were points of procedure which the observer recorded as having been done or not done. These included such items as "check bomb bay doors", "set-in trail (one mil error from No. 5 allowed",) "solve for sighting angle on E-6B", and "locate target."

This final phase check was used extensively in checking instructors, supervisors, and cadets in the various schools by the Bombardier Training Standardization Board. It was generally accepted by both schools and the board

*The work described here was performed by E. H. Kemp, M. J. Warrick, J. L. Sutton, J. J. Gibson, R. M. Gagne, E. E. Hicks, N. W. Stewart, J. W. Nagewitz, A. P. Johnson, J. S. Helmick, C. H. Humphries, R. L. Larson, R. H. Schwartz, J. M. Christensen, M. Hirsh, O. H. Summers, G. Wexler, S. F. Swenson, R. E. Barker, and L. F. Smith.

as a useful examining device. In early studies the reliability of the proficiency check was found to be quite low. In later samples in which two different checking groups administered the proficiency check to the same seventy individuals with considerable time intervening, a reliability coefficient of 0.46 was obtained.

Reports from combat theaters stressed the importance of target identification for bombardiers in practical operating situations. The aerial phase check included the identification of six check points, an initial point, and a target on each 3- or 4-hour mission. However, the missions varied in difficulty, depending upon the terrain flown over, the familiarity of the examinee with that terrain, the weather, and the amount of wind tending to blow the aircraft off course.

It was therefore proposed that a motion-picture test which could include a large number of targets and check points in a relatively short period of time of known difficulties would be a valuable addition to the measures for evaluating bombardiers on this important aspect of their work. This was considered especially important for potential lead bombardiers and work was therefore begun in the spring of 1945 on the development of such a test. The Psychological Research Project (Bombardier) collaborated with the Psychological Test Film Unit in the preparation of this test.

In August 1945 after only seven target identification runs (comprising 25 items) had been filmed over strategic targets on the scale model of Japan, production was stopped because of the cessation of hostilities. These seven runs together with introductory titles and blank film in place of briefing frames were edited and spliced together to provide with an accompanying script, a sample form of this type of test. No evaluation of its effectiveness has been accomplished to date.

The Psychological Research Project (Bombardier) prepared a printed Bombardier Proficiency Test in the fall of 1944. This was used extensively to provide a single objective command-wide means of evaluating individual cadets. It was also used to compare instructors and cadets at various schools.

This examination used charts, maps, tables, diagrams, and photographs very extensively and appears to have been one of the most successful in providing realistic tasks in printed multiple-choice form to the students. The correlation between the score on the Bombardier Proficiency Test, Form B, and ground-school average for more than 500 cases was found to be approximately 0.57.

## STUDIES IN PROFICIENCY OF FLIGHT ENGINEERS[7]

The position of flight engineer for the B-29 superfortress was one of the last air-crew specialties to be established. Psychological research on this specialty was not begun until September 1944 when a detachment from Psy-

[7]The work described in this section was carried on by M. P. Crawford, J. T. Cowles, J. T. Dailey, W. A. McClelland, A. A. Canfield, J. E. French, E. T. Seaman, R. W. Unger, L. D. Brokaw, H. F. Schmonsees, H. W. Riecken, and L. Grappe.

chological Research Unit No. 2 made field trips to Second-Air-Force and Training-Command stations specializing in B-29 training.

The study of proficiency measures followed the pattern for most other air-crew specialties. Preliminary investigations indicated that the graduation-elimination criterion of the proficiency of flight engineers was not a useful one because elimination rates at all flight-engineer schools were extremely low. The large demand for trained flight engineers made it imperative that practically every man in training complete training.

Turning to ground-school grades it was found that these were generally based on both performance tests and objective examinations of the short-answer type. Grades in the separate courses in both the preflight engineer school at Amarillo and in the flight-engineer school which was located first at Lowry and later at Hondo tended to correlate with each other to the extent of 0.20 to 0.50. For a sample of 747 men the correlation between the final grade at the preflight engineer school at Amarillo and the final school grade at the flight-engineer school at Lowry was .60. This agreement suggested a fair degree of reliability in the final grades of both schools.

Early investigation of criteria of proficiency in flight indicated that these measures, which were usually ratings made at the end of the particular flight, were probably not very satisfactory. In one study at Lowry Field the correlation between flight ratings and ground cruise-control grades was found to be 0.13 for 188 students. Since both of these measures were intended to indicate cruise-control proficiency and since the ground cruise-control grades had been found to have a fair degree of reliability, doubt was cast on the flight ratings.

This and other studies led the personnel of the Psychological Research Project (Flight Engineer) to develop a daily check list to be used in the air by the instructor to evaluate student proficiency in flight. This check list was put into use in the transition school at Maxwell Field in April 1945. The marks from these check lists were converted to normalized standard scores and were found to have significant correlations with records of Amarillo preflight engineer final grades and Lowry flight-engineer school final grades. For a sample of 224 cases the correlation coefficients were 0.27 and 0.26 respectively.

For use at stations giving B-29 operational training to crews a flight-engineer descriptive rating scale was developed. The agreement between ratings made by two instructors on this descriptive rating scale for 180 students at six different stations was found to be about 0.60. These ratings were made by the regular instructors and it is not known to what extent they may have depended on subjective reputation factors. The flight-engineer instructors in operational training schools regarded this descriptive rating scale as a very promising device.

During the spring and summer of 1945 a Flight Engineer Proficiency Test was developed primarily to be used in selection of flight engineers for B-29 lead crews. The final form of the test consisted of two booklets. The first booklet contained 102 verbal items and 48 diagram items. The verbal questions sampled knowledge of information and procedures while the diagram items showed the diagram of particular pieces of equipment or system and asked the student to indicate by selecting one of five numbers where to look for trouble under certain circumstances or which control lever to use to rectify a specific situation. The second booklet included items testing judgment regarding general theoretical aspects of cruise control, cruise-control computational problems, and problems to be solved by use of the flight-progress chart and cruise-control chart. Although this test was completed too late in the war for a comprehensive evaluation, tryout on small samples of students at Hondo flight-engineer school indicated that the scores on the test were substantially correlated with final grades in both the preflight engineer school at Amarillo and to a somewhat less degree with the Hondo flight-engineer final grade.

## STUDIES OF THE MEASUREMENT OF RADAR OBSERVER PROFICIENCY*

Just as in the case of the flight-engineer psychological research, research on radar-observer training at Langley Field was initiated in September 1944. Also, as in the case of the flight-engineer training program, this was just at the time that the AAF was greatly expanding training of radar observers to meet the needs of the B-29 crew-training program. Some exploratory work had been done with respect to radar-observer selection and training prior to this time and the Director of the Psychological Research Project (Radar Observer) had just returned from several months in the European Theater of Operations where most of his time had been spent studying the problems of selecting and training personnel to use radar equipment in bombing-through-overcast in the European Theater.

In this situation also the graduation-elimination criterion for judging the success of radar observers was of no practical value because the demand by the Operational Air Forces for radar observers was so great as to prohibit the failure of all but the most inferior students. The rapid expansion of training necessitated the use of so many instructors with no previous teaching experience and only limited motivation for this assignment that it was believed that instructor grades would not represent a reliable measure of success in training. Because of the rapid expansion of the program, sufficient photographic equipment was not available at the radar training stations to make it possible to obtain suitable records of bombing accuracy.

For these reasons the Psychological Research Project (Radar Observer) concentrated a large portion of its efforts on the development of a comprehensive battery of standardized proficiency measures to be used in all of the radar-observer training schools to provide uniform measures of proficiency.

*These studies were conducted by S. W. Cook, K. R. VanSaun, A. H. Hastorf, B. C. Sullivan, R. E. Johnston, H. Sofer, S. D. Murford, A. Babcock, A. S. Arnott, G. M. Ballinger, H. I. Roth, L. I. Foster, J. D. Hennessy, G. D. Oberh, S. Lottier, L. G. Carpenter, H. H. Kelley, T. P. Kincaid, N. Graf, D. W. Bray, P. H. Kriedt, H. F. Kussman, G. S. Klein, S. H. Nerby, W. J. Mauran, I. H. Harrison, G. S. Blum, R. T. Mitchell, W. F. Long, D. Collins, J. S. Harding, H. Heller, S. Blumberg, and N. L. Gage.

This battery consisted of five printed proficiency tests and six individually administered performance checks. Four of the tests were used at intermediate points in the course to measure proficiency in specific phases of training while the fifth served as a final comprehensive examination. Four of the performance checks measured proficiency on ground trainers while two measured aerial performance. One of these latter checks was administered half-way through aerial training and the other served as a final check on aerial performance.

The printed tests utilized two types of items, both of which were in the objective multiple-choice form. One type consisted of verbal questions measuring technical information and the other type required the solution of problems duplicating parts of a student's actual job as a radar observer. The reliability coefficient was computed for only one of the printed proficiency tests. For two different forms of the final examination on samples of approximately 200 cases each, the split-half reliability coefficients corrected for length were found to be around 0.80.

One section of this final comprehensive examination was designed as a simulated radar-observer bombing mission. This section consisted of interrelated items which required solution of navigational and bombing problems at various points along a typical mission route. A full-scale navigational chart and photographs of the radar scope were used in the students' computations. It was believed that by this procedure organizational abilities, integrative aspects, possibly personal habits and characteristics important for conducting an actual mission but not included in items measuring separate skills could be tested.

In the final comprehensive examination there were among the various sections three sections each of which measured one of the skills related to the total task of this simulated mission. It was found that when combined the score on the three sections measuring separate skills correlated with the section containing the simulated mission to the extent indicated by a correlation of 0.32 for one sample of about 200 cases and approximately 0.60 for each of two additional samples of 200 and 300 cases respectively. Since the reliabilities of the various tests are not known it is not possible to make a very satisfactory estimate of the extent to which unique factors are measured by the section on the simulated mission.

Although, as indicated above, accurate measures of the reliability coefficients for these tests were lacking, the intercorrelations of the various printed tests as found for a sample of 190 cases suggest that the reliability coefficients were not high. Most of the intercorrelations between the separate tests were in the neighborhood of 0.35. For the subtests on one of the final examinations the intercorrelations were about 0.45. Correlations with performance checks were also found to be quite low, most of them being around 0.15.

It was believed that these low reliability coefficients were due in part to the use of interdependent problem-solving items. This interdependence of items has the unfortunate characteristic that if an individual misses one of the first items he automatically misses the whole series, whereas he can miss the last item without serious penalty. Experience with these items suggested that the organizational and sequential aspects could be retained to some extent by periodically breaking the sequence at the end of a typical series of items and supplying information for starting a new group of items. Another device found useful was to make separate items out of intermediate steps in the solution of a problem. As in other examinations prepared for air-crew specialties extensive use was made of maps, diagrams, pictures, and other materials which added reality to the problems.

As indicated previously, six standardized performance checks were prepared by the Psychological Research Project (Radar Observer). These included checks for evaluating student proficiency on airborne radar missions, on supersonic trainer missions, and on the classroom radar equipment. These checks contained statements of standard conditions under which the checks were to be administered. Special boards of examiners were trained at each radar school to administer these checks.

During a performance check given on the supersonic trainer or in the air, the radar-observer student simply flew a typical mission. To achieve as much standardization as possible the missions were flown over certain standard routes. Standardized directions to the student were provided and the examiner was supplied standardized instructions for evaluating performance in terms of specific items. Although the examining boards were trained to administer these performance checks, adequate control of the individual judgment and competence of the examiner was not attained according to the observations of the aviation psychologists. The correlation between the printed proficiency measures and the performance checks were found to be around 0.15 for the sample of 190 cases previously referred to.

As mentioned before, these tests were used throughout all radar observer-training schools and thus provided a standardized measurement program. In the opinion of the personnel who supervised the radar-observer training schools, the introduction of these examinations had a very beneficial effect on the quality of instruction in these training schools.

As radar-observer training became better established, additional photographic equipment became available so that during the last few months of the war it was possible to make studies of the bombing errors of students in radar-observer training. The two principal types of bombing practice were simulated or camera-bombing missions, which were the most frequent, and missions on which practice bombs were dropped. When bombs were actually dropped, the procedure for recording and scoring was similar to that used in bombardier training. This consisted of photographing the target and surrounding terrain at the moment of impact of the bomb. The circular error was then obtained by measuring the distance on the photograph between the point of impact and the center of the target.

To score a simulated bomb drop a series of vertical terrain photographs were taken during the bomb run. From these photos the point of impact of

703323—47—10

the simulated release was plotted and the amount of error was scored. Scoring was found to be facilitated by the use of a mosaic photograph including a wider area than any of the single photographs taken during the bomb run. The principal limitations of this type of scoring of simulated missions were deviations from vertical alignment of cameras caused by turbulence and inadequate visibility due to poor weather. It was possible to overcome the first difficulty by using gyro-stabilized cameras. To overcome the difficulties due to lack of visibility a method was developed of using gun-laying radar to record the exact position of the plane at various points on its bombing run and at the moment the bombs were "released." These procedures involved radio liaison between the aircraft and the ground radar installation.

Two studies were made of the reliability coefficients for bombing errors obtained on odd and even missions. For a sample of 122 students at Boca Raton a coefficient of 0.19 was obtained, which indicates a reliability for the total number of bomb releases equal to 0.31. In a similar sample of 372 students at Victorville a correlation between odd and even missions of 0.11 was found, which corrected for the full course was 0.20. Both of these results were for camera-bombing missions only.

For 178 cases at Boca Raton and Langley Field, the correlation between camera-bombing error scores and course grades was found to be 0.20. For 679 cases at Victorville, Boca Raton, and Langley Field, the correlation between camera-bombing error scores and course grades was found to be 0.16. It appears that the reliability coefficient of scores obtained from camera-bombing missions are somewhat higher than those obtained from visual-bombing missions, although both sets of coefficients are very low.

Although the development of procedures for scoring bombing runs with the aid of ground radar installations occurred so late in the war that evaluations of this procedure were not possible, it appears that this method may provide a much more satisfactory basis for the measurement of bombing proficiency than has been available.

## STUDIES OF PROFICIENCY TESTS FOR FLEXIBLE GUNNERY TRAINING*

### Printed Tests and Phase Checks

For the first few months after the training of men for duty as flexible gunners was initiated in December of 1941 training and grading procedures were rather inadequate due to the newness of the program, the lack of equipment and the lack of trained personnel as instructors and training supervisors. Assistance was first given to personnel in the flexible gunnery schools by psychologists in the fall of 1942 when research detachments were sent to three of the gunnery schools in connection with the selection program for D-8 bombardiers.

*These studies were done by R. N. Hobbs, R. W. Russell, J. V. McQuitty, L. M. Stolurow, M. Waldman, G. J. Wischner, J. D. Adams, A. L. Irion, A. C. Jensen, R. J. Ellington, K. B. Henderson, P. Freeman, W. C. Mahoney, J. G. O'Hara, L. A. Carey, H. W. Goldstrom, S. Kern, J. A. Valentine, W. S. Gregory, G. R. Pascal, R. G. Bainbridge, C. A. Babl, J. T. Huber, P. M. Brown, R. N. Hubbell, J. H. Lickleia, R. Wilmeth, E. W. Roy, T. P. Gallagher, T. R. Valance, J. F. Lawrence, M. E. Smith, M. J. Haire, R. C. Payne, and A. W. Melton.

Shortly after these detachments had completed their work a detachment was sent to Fort Myers to work on psychological research projects in connection with the flexible-gunnery training program. The initial efforts of this detachment in the spring of 1943 were directed toward improving testing procedures by training instructors in methods of developing and administering achievement tests. As a more adequate psychological staff became available, the construction of standard tests for all flexible-gunnery schools was taken over by the psychological personnel.

In attempting to develop tests for the various skills and proficiencies included in the gunnery-school curriculum, attention of the supervisors of training programs was called to a need for clarifying the objectives of instruction. The basing of tests on the principal objectives agreed upon assisted materially in improving the instruction in the flexible-gunnery schools. After considerable experience had been gained with these examinations the officers in charge of the gunnery schools directed that all gunnery schools use a standard final comprehensive examination prepared by the psychological research unit working on flexible-gunnery problems. This gunnery achievement test made considerable use of pictorial items, reducing as much as possible the demand for verbal facility. The later forms of this examination were found to have a reliability coefficient as obtained from Kuder-Richardson Formula No. 21 of about 0.87 based on several samples totaling more than 6,000 cases.

When gunnery training was established, planes, ammunition, guns, and other necessary equipment for such training were not available. In the absence of guns, planes, and training devices the training authorities were forced to resort to describing guns and naming the various parts and asking the students to memorize these descriptions and names. For several months the personnel in gunnery training listened to long lectures about guns and read descriptions of guns but there was not even adequate pictorial material concerning guns and their functioning.

In order to facilitate the shift from this emphasis upon verbal learning to a more functional learning-by-doing, gunnery performance tests known as phase checks were developed. Each phase check consisted of a series of carefully prepared items which described precisely each step in the performance of some gunnery task such as stripping and assembling the caliber .50 machine gun. In developing these phase checks an attempt was made to include all important steps in a task so that the check list could be used for training as well as for testing. Phase checks were developed for the caliber .50 machine gun, for all turrets, and for all preflight, in-flight, and postflight procedures. They measured proficiency in all of the important duties necessary for the gunner to be prepared to aim and fire his guns. In using the check list all items were marked on an all-or-none basis. Items correctly done were checked and items incorrectly done were marked with a zero. The score was recorded in terms of the number of errors indicated.

The average correlation for two samples totaling about 1,000 cases between scores on the phase check on stripping and assembling of the caliber .50 machine gun and the weapons section of the gunnery final examination (a printed multiple-choice test) was 0.01. There is practically no data available as to the reliability of the scores obtained from the phase checks. After the check lists became generally used as training devices there was a tendency for a larger and larger proportion of the students to obtain perfect scores. This large proportion of perfect scores made the computations of reliability coefficients using the usual procedures of very limited value.

These phase checks were used in all basic gunnery schools, in all gunnery-training stations of the Continental Air Forces, and in most combat air forces. The widespread use of these check lists is indicated by the fact that for one series of phase checks 1 million and a half booklets and about 9 million answer sheets were printed and distributed. This extensive use of these standard devices insured a high level of information and proficiency regarding the equipment and procedures which were necessary for carrying out the gunner's main job of aiming and firing his guns.

## Studies of Air Firing

As in all of the other air-crew specialties, the real test of the proficiency of flexible gunners was believed to be their ability to perform their assigned tasks in the air. The main task of the gunner in the air was to point and fire his guns in such a way as to hit and destroy enemy aircraft. The precise nature of his task differed with the various types of gunnery sighting equipment and the turret used. Usually the gunner had to move his controls in such a way that the guns moved with the attacking plane and at the same time to make some estimate of the range of the attacking fighter. In the later forms of the gunnery sighting equipment the gunner's task was to keep his sight pointed right at the attacking plane and at the same time vary the size of his range control so that the wing spread was precisely framed by the sight.

In training it was of course impossible to measure directly the gunner's proficiency in hitting and destroying fighter aircraft. In the earlier days of gunnery training, gunners shot at a sleeve towed by an airplane along side the training plane. Practically the only similarity between this type of sighting and firing and the actual combat situation was that the gunner was in the air and was firing a gun.

In order to simulate the combat situation more accurately various procedures were developed. Probably the most successful of these was the use of gunsights equipped with 16-millimeter moving-picture cameras. The gunner would aim the sight at a fighter aircraft which simulated an attack on the plane. Unfortunately, it was found that the conditions were very difficult to standardize because of variability in attack and in the course of the bomber, variations in weather (including both turbulence and visibility), and harmonization of the camera with the sight. Because of this, unless special efforts were made to obtain standardization, the films obtained were of only limited value. Furthermore, it was found that unless great care and close supervision of the scoring of the film was possible this processing introduced large individual errors into the proficiency measures.

To provide a criterion of gunnery ability for certain crucial experiments in connection with the selection and training of flexible gunners three studies were made in which the use of the gun cameras and the scoring of the film were carefully supervised by research personnel. These studies were conducted jointly by psychologists of the Research Division of the Central School for Flexible Gunnery and the Department of Psychology of the AAF School of Aviation Medicine.

The earlier experiments usually used 10 missions, including about 6 attacks each. In the last experiment a total of four missions was obtained for most of the students. An analysis of these results indicated that the average score for tracking and framing or the average percent hits as obtained from the attacks on one mission agreed to the extent indicated by a correlation coefficient of approximately 0.20 with similar measures obtained from another mission. The number of students involved in the first study was 32, in the second 64, and in the third 463. It is apparent that sufficient reliability for experimental use could be obtained by combining the results for a few such training missions.

Just before the end of the war a new development was introduced into the gunnery-training program which appeared very promising as a proficiency measure. This was the Frangible Bullet Trainer. This involved a special .30-caliber projectile which was fired at an armored fighter. The bullet was made of powdered lead held together with a plastic base and it fell apart on impact. The P-63 fighter planes used were specially armored and used bullet-proof glass. Hits were indicated by the flashing on of a light in the nose of the fighter plane and were recorded by counters in the cockpit of the fighter. The short range of these special .30-caliber bullets and the damage done by them to the planes, and especially the glass of the fighter aircraft, appeared to introduce serious limitations in the value of this procedure. However, it had great realism and high motivational value and it is possible that the procedure could have been perfected had the war continued.

Another promising device which was in an even earlier stage of development at the end of the war was the Firing Error Indicator. In this procedure a transmitter was mounted on the nose of a small glider which was towed about 300 yards behind and 45 degrees below a B-26 tow plane. By flying the B-26 as though it were attacking the bomber in a pursuit curve except about 200 yards higher, the glider could be made to approach the bomber in a manner closely simulating that of an attacking fighter plane. The gunner fired regular .50-caliber bullets at the target which contained two microphones and radio transmitters which informed the receiving station in the bomber whether the bullet passed to the right or left of the target. It also gave an estimate of the approximate minimum distance of the bullet from the target by means of the strength of the shock wave created by the bullet in the air.

A considerable amount of experimental work was done on this problem both on the ground and in the air but the procedures had not been brought to a point where they were of practical use by the end of the war.

It is apparent that the problem of obtaining an adequate measure of the proficiency of gunners in carrying out their primary mission was one of the most difficult encountered during the war. Significant progress was made toward the solution of this problem and the experience gained should be of great value in connection with research on the gunnery problems of the future.

## CONCLUSIONS

Studies of existing measures of proficiency in the various air-crew specialties revealed that these were inadequate for a precise evaluation of individual proficiency. The work of the aviation psychologists in this field was accordingly concentrated on the development of objective measures of the various skills involved and especially on the measurement of these in the actual flying situations. It was found that to develop satisfactory objective measure of flying skill it was necessary to control three main sources of variation in the students' skill in performing the particular task assigned. These were:

1. The judgment of the examiner observing the students' performance.
2. The student's interpretation of the nature of the task.
3. The nature and difficulty of the task itself for the various students.

Objective check rides were developed at all levels of training for pilots in the AAF Training Command schools. Objective measures of flight performance were also developed for each of the other air-crew specialties. Although these measures are all in need of further development much important progress is reported in this chapter. It is hoped that this important work can be developed to the point where objective scales of flying performance can be used routinely in training schools.

# CHAPTER SIX
# Findings Regarding Instructional Problems in the Flying Training Schools

## INTRODUCTION

The development of the large mass training program necessary for supplying individuals trained in the duties of each of the air-crew specialties presented many problems and opportunities. These individuals needed to be fully trained in a short period of time so that they could take over immediately the duties of their particular crew positions and be ready to man with a minimum of crew training the enormous numbers of bombing planes being produced in the factories. After the first rush to get the training schools established and manned and students flowing through them, the more serious defects of these hastily organized training programs began to become apparent.

To analyze the situation and assist in removing some of the major deficiencies psychological research personnel were invited to work with training groups. For some of the specialties the procedures and equipment were changing so rapidly that the training courses could not become fixed and entrenched. In nearly all of the flying-training schools the changes of supervisory and instructional personnel were so frequent that the accepted and traditional methods did not become firmly rooted.

One of the first problems in the flying-training schools which psychologists were asked to work on was the problem of selecting and training instructors. At the beginning of the war the Army Air Forces had no systematic procedure for training instructors. Officers were assigned to training schools and the supervisory personnel were responsible for giving them whatever instructions were necessary for them to assume the duties of the type of instruction to which they were assigned. The same situation held for the civilian contract schools. Each school selected its own personnel and gave them what training it thought was necessary. The amount and quality of such training in both the Army and civilian schools varied a great deal but was usually quite meager. The psychological research projects assigned to the various air-crew specialties took as one of their first problems the question of instructor personnel. The work and findings of these groups are briefly summarized in the paragraphs that follow.

# INSTRUCTIONAL PERSONNEL

## Instructors in Pilot Training Schools[1]

### *Qualities of the Good Instructor*

The procedure in use in pilot training at the beginning of the war was based on the premise that an officer who had graduated from the flying-training course had been initiated into the process of learning to fly and was capable of inducting new members into the organization. Many of the instructional methods were not written out and were merely passed on from one pilot to the next as he underwent the training process. The huge expansion of pilot training as a result of the war resulted in the accumulation in various places of many useful and valuable methods regarding pilot instruction and made the need for both training and selecting instructors more obvious. Psychologists were asked to assist in developing and teaching courses in the principles of learning and instructing in the Central Instructors School established to improve the quality of instruction.

To provide a sounder basis for the selection and training of individuals for duty as pilot instructors, several studies were carried out to determine the characteristics of a good instructor of this type. In these various studies more than 2,000 students and instructors reported in interviews or in response to check lists or rating scales their judgment as to the qualities most important for pilot instructors.

The results from these studies agreed fairly well in indicating the importance of four main areas to successful work as a pilot instructor. These general attributes are as follows:

1. *Favorable attitude and suitable temperament.*—Both the students and instructors emphasized that the student cannot learn unless the instructor wants him to learn or at least allows him to learn. In addition, the student cannot be expected to learn if he is filled with anxiety and concern regarding the instructor's reaction to his efforts. Emotional outbursts on the part of instructors are distractions which make no positive contribution to the student's learning.

2. *Insight into flying problems and the ability to express himself clearly.*—To be effective the instructor must not only understand precisely why the airplane responds to controls in a certain fashion but must also be able to express them clearly so that the student can understand them. The mechanical type of learning resulting from a mere observation of the correct technique as demonstrated by the instructor and duplication of this procedure by the student does not create the type of comprehension and judgment essential to the superior pilot. Although it is not essential that either the insight of the instructor or his expression of them to the student be verbalized, they must be clearly transmitted. This can frequently be done more

[1]The work reported in this section was done principally by N. E. Miller, M. P. Crawford, R. P. Youtz, J. T. Dailey, W. A. McClelland, J. T. Cowles, J. Weitz, C. P. Froehlich, R. J. Keller, I. E. Farber, C. H. Patterson, I. Robbins, W. Nygard, D. J. Grier, W. E. Galt, R. R. Blake, S. C. Erickson, T. N. Ewing, J. R. Rohrs, L. A. Hellmer, W. Ismael, J. W. Stratton, C. P. Gershenson, and G. L. Finch.

effectively by appropriate demonstration and supervised experimentation by the student than by abstract verbal explanation.

3. *Flying ability and confidence in himself and the equipment.*—Flying, to a greater extent than many other skills, must be taught by an expert. Both the student and the instructor must have full confidence in the ability of the instructor to control the plane under all circumstances. The instructor must allow the student to experiment and learn for himself without becoming nervous or apprehensive.

4. *Understanding the individual and learning principles and techniques.*—It is important that the instructor understand the individual and the effect on him of various types of behavior. It is also important that the instructor have an understanding of the fundamental principles of learning. Students stressed the importance of knowing immediately when they had made an error, the desirability of clear assignments for solo flights, and the necessity and value of the instructors always following the procedure of immediately verifying the students' comprehension of explanations and demonstrations.

In general, practice in the Army in selecting officers for duty as pilot instructors and in training them concentrated almost exclusively on only one of the four qualities judged to be important for successful work as an instructor. This aspect was flying ability.

### *A Study of Instructors in Contract Schools*

On the basis of these studies a number of tests were developed or selected for use in selecting individuals for duty as pilot instructors. These examinations included a number of tests of intellectual ability and special aptitude for flying such as reading comprehension, aviation vocabulary, mechanical comprehension, aviation information, and reasoning. Two tests which were designed to measure the individual's ability to analyze control movements and the types of errors made by students in performing typical maneuvers were prepared. Tests were also developed to try to measure attitude, interest, and personality factors believed to be related to the work of the instructor. These included tests of preference and biographical data.

A study was made to evaluate the tests developed. To provide criteria, over-all ratings of the teaching ability and flying ability of instructors in pilot-training schools and ratings of these instructors on a 16-item descriptive rating scale were given by flight supervisors and squadron supervisors. These were obtained for 423 civilian instructors in primary flying schools.

The ratings given the instructors on various traits were correlated with the scores obtained on the tests by these instructors. Significant correlation coefficients were obtained between ratings on the trait of pilot ability and all of the tests except one of abstract intelligence. The two traits for which the descriptive ratings showed the next largest correlations with the tests were those on ability to analyze student errors and ability to express himself. The correlations of all test scores with the over-all ratings of teaching ability were very low. The test having the highest correlation was one of mechanical comprehension.

No definite pattern was apparent from examination of the correlations between the individual test scores and the various items of the descriptive rating scale. There was a suggestion of discrimination in certain circumstances. For example, the highest correlation for the reading-comprehension test was with the ability to express himself. Similarly, there was a relatively high correlation between tests of instructional judgment and analysis of control movements and the ratings of ability to analyze student errors. It was suggested by this study that these supervisors at the primary level had at least obtained a fairly good estimate of the flying ability of the instructors under their supervision and probably had some insight into certain of their other abilities. The civilian instructors represented a very wide range of background and ages. The frequent check rides with students beginning training under these instructors, especially with students being considered for elimination, provided the supervisors a good opportunity to evaluate the various instructors under their supervision.

*Rating Scales for Evaluating Combat Experienced Instructors*

In the spring of 1944 the availability of a large number of returned pilots with combat experience aroused considerable interest in the problem of selecting from this group the ones who would be most suitable for duty as pilot instructors. For this purpose a battery of tests was administered to the combat returnees at the redistribution stations of the AAF Personnel Distribution Command.

To develop a measure of success as an instructor against which to validate these tests the Psychological Research Project (Pilot) carried out a number of studies involving the development and use of supervisors' rating scales, self-rating scales, and students' rating scales. The reliability coefficient obtained for a group of 74 supervisors' ratings was approximately zero. For a group of 44 instructors the correlation between students' ratings and self-ratings was found to be 0.12. In preliminary studies the reliability of the students' rating form appeared to be promising and it was therefore decided to concentrate work on the development of the students' rating scale.

The original 47-item student-rating scale was administered to 3 groups and on the basis of the results was successively shortened to a 34-item scale which was readministered and reduced to a 25-item scale which in turn was reduced to an 18-item scale on the basis of the third administration. Each of these administrations involved between 200 and 400 students.

The sums of the ratings given on the 18 items of the revised rating scale by two students in the same class rating the same instructor were correlated to obtain an estimate of the degree of agreement or reliability. The reliability coefficients for the average of the ratings by four students on a single instructor were estimated using the Spearman-Brown formula. These reliability coefficients are as follows and are based on the number of pairs of students indicated in parentheses: Primary ($N = 68$), 0.60; Basic ($N = 233$), 0.86; Advanced ($N = 782$), 0.69; Transition ($N = 613$), 0.83. It was believed that these results were higher than would be obtained if the students' judgments were independent. That is, it was believed that students discussed instructors among themselves and therefore, when asked to give ratings, reported what was in part the reputation of the particular instructor. For this reason it was decided to administer the Students Rating Scale to four successive classes in advanced and transition schools.

A total of 281 instructors were found who had been rated by students from two classes. An average of 3.25 students in each class rated each instructor. The reliability of the Students Rating Scale obtained from ratings by 3.25 students in each of two different classes was found to be 0.36. Using the agreement of two students in the same class, weighting the schools according to the proportions in this sample, and correcting for the fact that 3.25 students rated the same instructor, the figure comparable to 0.36 in terms of the within-class correlation is 0.75. This finding confirmed the belief that the ratings of the various students were not independent but contained a substantial reputation factor. It was found also that the class-to-class reliability coefficient of the over-all item of the Students Rating Scale obtained from this same group was 0.23 as compared with the reliability for the total score of 0.36.

The intercorrelations among the 18 items of the Students Rating Scale and the ratings by supervisors of flying ability and teaching ability were obtained for a total of 1,141 students. These intercorrelations are shown in table 6.1. A graphical analysis of the patterns of intercorrelation for the various items indicated a grouping rather similar to the four types of qualities which were derived from the study of the characteristics of successful pilot instructors. For this sample the correlations between supervisors' ratings of flying and teaching ability and the total score on the eighteen items were found to be 0.06, and 0.08 respectively. The correlations for the over-all item "effectiveness in putting flying training across" were 0.08 and 0.11 respectively.

The correlations between students' rating on "ability to demonstrate maneuvers" and the two general ratings given by supervisors were 0.07 and 0.09 respectively and for "interest in instructing" the corresponding coefficients were 0.03 and 0.09. For the item "confidence in this airplane" the coefficients were 0.07 and 0.06. Nearly all of the remaining coefficients were even smaller than these. These data suggest that the supervisor's ratings were based mainly on his opinions regarding the instructor's flying ability. This is somewhat confirmed by the correlation between the supervisor's rating of flying ability and teaching ability of 0.70. Since the students were not asked to rate the instructor's flying ability, it is not possible to make a direct comparison for this particular item.

To explain the large drop in agreement between students' ratings of instructors for successive classes as compared with students in the same class it was suggested that these ratings were based in part on a general reputation factor for the instructor growing out of discussions on the part of the students. It seemed probable that this repuation factor also carried on to

TABLE 6.1.—*Intercorrelations among items of 18-item Students Rating Scale and supervisor's over-all ratings*

[N = 1141; S.E. of a zero correlation = 0.03]

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | Tot Sc | Sup Fly | Sup Tch | M | SD | 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 | 34 | 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 | 1.59 | 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 | 59 | —03 | —02 | 6 82 | 1.59 | 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 |

[1]The scores used in computing the intercorrelations were the average of ratings by one or more students. Each of the 1,141 instructors was rated by an average of 3.1 students

some extent from one class to the next, since there are ordinarily at least two different classes, approximately 5 weeks apart, in the advanced schools at the same time. Unfortunately, it was not possible to estimate from the data available the extent to which ratings from students in different classes were independent.

### *A Validation Study of Instructor Selection Tests*

The students' ratings and supervisors ratings for approximately 200 of the instructors included in the studies just mentioned were compared with the scores achieved by these instructors on the Instructor Selection Tests administered at Redistribution Stations. For this sample the combined score based on six tests yielded correlation coefficients of —0.04, —0.18 and —0.07 with the total scores on the Students Rating Scale, supervisors' ratings of teaching ability, and supervisors' ratings of flying ability respectively. The range of the correlations for the individual tests was from —0.19 between the Verbal Comprehension Test and the total score on the Students Rating Scale to 0.04 between the Analysis of Maneuvers Test and this same score. The lack of reliability and lack of agreement among the various ratings used as criteria of success as instructors had not given any reason for expecting substantial positive coefficients. However, the tendency toward negative coefficients suggested the possibility that bias had been introduced due to uncontrolled factors. Investigations did not reveal the source of any such possible bias.

### *Development of an Instructor Check Ride*

An instrument instructor check ride was developed. This consisted of a standardized check ride with complete instructions to the instructors regarding the procedures and the method of rating. At the AAF Instructors School (Instrument Pilot) 179 of the students in the class of 44-9 were given two check rides 7 to 10 days apart by different staff instructors under conditions insuring independent ratings.

The reliability coefficients for this sample were voice, 0.25; demonstration technique, 0.40; explanation technique, 0.54; knowledge of subject, 0.52; analysis of maneuvers, 0.51; final grade, 0.50. The correlation between scores on this check ride and previous experience as an instructor was 0.30 when expressed as a biserial correlation coefficient.

For a sample of 374 students the correlation between this instructor check ride and the check ride in instrument flying proficiency was found to be 0.13.

These data suggest that such an instructor check ride as this may have promise as a measure of instructional proficiency. Its disadvantages are that the check rider served as student during the ride and was, therefore, not able to make as systematic a report of the behavior of the instructor as would an independent observer. His known skill in flying also tended to make the situation somewhat artificial. However, it is believed that further development along this line is indicated.

### *The Comparison of Ratings of Instructors with Personal Data Regarding Them*

To provide further information regarding the ratings of instructors by students and supervisors, these ratings were compared with a number of personal-data items regarding the instructors which were gathered at the same time that the ratings were obtained. This analysis is based on data from 1,284 instructors. The students gave higher ratings to the instructors who were combat returnees, who held the rank of first lieutenant or higher, were 24 years of age or older, had a year or more of college education, and had logged more than 1,100 hours of flying time. All of the differences mentioned above were sufficiently large so that it is not reasonable to suppose that they were caused by sampling fluctuations.

Smaller differences in the students' ratings favored married pilots living with their wives with no children, graduates from the Central Instructors School, officers with less than 4 months' or more than 2 years' experience as an instructor, and officers who expressed a positive liking for the job as flying instructor.

Although the controls were not adequate to permit definite conclusions there was an interesting suggestion obtained from the analysis of the items of the Students Rating Scale and length of time the officer had been an instructor at that station. The analysis 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 the ability to express themselves and reliably lower ratings on personality items such as interest in the student and patience and self-control. It is possible that long periods of teaching flying tend to improve the instructor's technical ability but to spoil his patience and personality.

The supervisors gave higher ratings to the instructors who had been instructing the longest at that station, who had not served in combat, who were 24 years of age or older, who expressed a positive liking for the job of flying instructor, who had logged more than 1,100 hours, who were married, had children and were living with their family, and who had more than 5 months' or preferably more than 15 months' total experience as a flying instructor. All of these differences exceed the size of differences which could be expected to have arisen by chance.

The supervisors favored by smaller amounts persons without Central Instructors School training and individuals with a year or more of college training. The supervisors gave their highest ratings to second lieutenants, next highest to first lieutenants; below these they rated instructors in the rank of captain or over; and the lowest average ratings were given to a very small group of flight officers. Unfortunately, records are not available of the rank and combat experience of the supervisors. It is not believed that many of these supervisors had had combat experience, since only 404 of the 1,284 instructors included in this study were combat returnees.

### *An Analysis of Stability of Preferences*

One study which bears on the general problem of the selection of instructors is an analysis of statements and ratings regarding interest in and preference for work as a flying instructor. In a sample of approximately 200 cases, a correlation of only 0.12 was found between the instructor's statement as to his desire for assignment as an instructor at the time of his return from combat to a redistribution station and his own rating regarding his liking for the instructor's job after he had been on the job for some months.

Similarly, this preference rating made at the time he was in the redistribution station had no correlation with the students' estimates of instructor's interest in his job or students' or supervisors' ratings of instructor proficiency. For a sample of 277 cases the instructor's ratings of his liking for the flying-instructor job obtained while he was working as an instructor but with an interval of at least 5 weeks between ratings agreed to the extent of a correlation coefficient of 0.82. Thus, although the instructors could not predict ahead of time whether or not they would like instructing, once they got on the job their liking for it tended to be fairly stable.

### *Implications*

Although final solutions were not obtained in this area it is believed that these studies have thrown considerable light on the very difficult problem of selecting and training instructors for pilot flying schools which also have implications for the more general problems of instruction and education. It is believed that the introduction of practice teaching in the Central Instructors School would be an important advance which would make possible much more rapid progress in this area. Further work with the Student Rating Scale and the Instructor Check Ride also appears to be desirable. As improved methods for evaluating the flying skill of students become available it may be possible to compare instructors on the basis of the level of skill which their students have attained.

## Instructors in Navigation-Training Schools[a]

### *Development of Instructor-Selection Tests*

In the early phases of the development of the navigation training program, the instructors were drawn from the few officers with ratings in navigation, a number of civilians selected to teach in the navigation schools by the Officer Procurement Boards, a number of officers who had been commissioned for this work from civilian life, and a large number of instructors who were selected from the graduating classes of advanced navigation schools. In 1943 an AAF Instructor School was established at Selman Field to train instructors for navigation schools. For a while this school trained recently graduated navigators, but in 1944 a large number of navigators with combat ex-

[a]The research reported in this section was carried out by L. F. Carter, W. A. Zielonka, R. M. Rust, E. M. Rosemark, G. L. Heathers, F. J. Dudek, A. E. Grux, G. C. Brown, and N. L. Kravetz.

perience were returned to the Training Command for assignment and all of these men were sent to the AAF Instructors School at Selman Field. Until December 1944 all navigators with combat experience reporting to the Training Command were sent to the Instructor School and assigned to advanced navigation-school teaching. By the end of 1944, however, there were such a large number of navigators available for assignment as instructors that a system of instructor selection was established.

The introduction early in 1944 of the policy of rotating instructors in all training schools to combat duty and the increasing number of combat returnees emphasized the need of procedures for selecting from among the combat returnees those most suited to instructor duty. To meet this need various groups within the Aviation Psychology Program undertook to develop suitable tests. This test-development program was discussed at a conference at Headquarters AAF Training Command early in March of 1944 and work was begun immediately on the preparation of these forms. Experimental testing of these forms was carried out in the redistribution stations of the Personnel Distribution Command and in other stations from June to November 1944. In November 1944 a program of administering a battery of instructor selection tests was established at all redistribution stations.

To provide an estimate of success as a navigator instructor, scores were combined for tests of attitude, individual adjustment, reading comprehension and reasoning, and navigator proficiency. A stanine for navigator instructor was obtained from these combined scores.

Studies carried out at the Psychological Research Project (Navigator) indicated that the correlation between the scores obtained on a group of attitude items administered at the redistribution stations correlated only 0.66 with the scores on these same items administered after these combat-experienced navigators had spent a few weeks at the Instructor School for navigators. The authorities at this school, therefore, requested that the test be given after the returnees had arrived at the navigation school. Accordingly, an attitude test similar to the one given at the Personnel Distribution Command stations and an information test of scientific background were given at the Instructor School and combined to provide a Screening Test Score.

Unfortunately, these tests were developed before there had been an opportunity for a systematic and thorough analysis of the duties of the navigator instructor. This would not have been so serious if a suitable criterion had been available to provide an empirical check on the value of these tests. However, as was frequently the case in similar situations, an investigation of the available measures of success as a navigator instructor indicated that no available measure was satisfactory. It was found that success in the navigator instructor school was largely dependent on proficiency as a navigator and this could of course be predicted by intellectual tests. Both the Navigator Instructor Stanine and the Screening Test Scores yielded significant, though not very high, correlations with the grades and reports used at these schools.

### *The Report of Cadets' Opinions Regarding Instructors*

It was apparent that these grades reflected to only a very small extent the specific qualities necessary for success as a navigator instructor. To improve the validity of the data available at the schools, Psychological Research Project (Navigator) was requested to develop a procedure for rating practice lectures given by students in the Instructor School. This led to the development of a fourteen-item lecture-rating scale. For a sample of 53 cases the total score on this rating scale was found to have a reliability of 0.60. The extent to which this apparent reliability is due to prior informal discussion regarding the abilities of the instructor-school students by the raters is not known.

This scale provided the basis for the cadet rating sheet of advanced-school instructors. This modified rating sheet was adopted for official use in all advanced navigator training schools as a part of the cadet opinion survey. It was found that for each of the 14 items the average rating given to a particular instructor by the odd- as compared with the even-numbered students in a particular class of approximately 40 students agreed to the extent of a correlation coefficient of about 0.87. This coefficient was based on a group of about 52 instructors. The average ratings given by the students in one class were correlated with the average ratings given the same instructor by a different class for a group of 20 instructors. The correlation coefficient obtained was 0.47.

In completing these scales each student selected one of five choices as being the best description of the instructor being rated. The student did not sign his name to his rating. The ratings for each class for each instructor were averaged and compared with those of all other instructors. A graphic profile giving the instructor's relative standing on the 14 items was prepared and sent to the supervisor of the particular instructor for the information of both the instructor and the supervisor. This report form together with the form "Explanation of Report of Cadet Opinions Regarding Instructors" which accompanied it are reproduced in figures 6.1 and 6.2.

In table 6.2 are shown the results of the follow-up study of the correlations of the tests given at the Navigator Instructor School with ratings obtained for these instructors in Advanced Navigation School with respect to their actual teaching performance. It will be noted that the correlations are all extremely low. There were not a sufficient number of cases in the follow-up groups who had taken the tests in the redistribution stations to make a useful analysis in terms of these other tests.

TABLE 6.2.—*Correlations of instructor-selection tests with criteria of advanced-navigation-school teaching performance*

| Teaching criteria | Survey of attitude | Scientific Background | Screening test | N | M | SD |
|---|---|---|---|---|---|---|
| Student opinion | 0.01 | 0.01 | 0.03 | 242 | 4.04 | 0.46 |
| Fellow instructors | —.03 | .10 | .05 | 178 | 4.08 | .32 |
| Flight commanders | —.04 | .03 | .01 | 127 | 4.09 | .47 |
| Supervisors | —.05 | .11 | .03 | 42 | 3.93 | .40 |
| Self-evaluation | .10 | —.18 | .02 | 143 | 4.01 | .34 |

The findings from studies of navigator instructors tend to agree with those previously discussed based on pilot instructors as to the value and limitations of various types of ratings of success in instructional work. The systematic collection and tabulation of student opinion, regarding both instructors and training procedures in general, as developed by the navigator project is believed to represent a useful technique which deserves wider investigation and use.

## REPORT OF CADETS' OPINIONS REGARDING INSTRUCTORS

| NAME | Roe | Richard | W. | 1st Lt. | I | 155N |
|---|---|---|---|---|---|---|
| | Last | First | Initial | Rank | Flight | Class Taught |

I. In estimating this instructor's teaching abilities, of those taking the survey ($N = 43$), 56 percent stated they were very certain of their opinions; 36 percent stated they were fairly certain of their opinions; 8 percent stated they gave their opinions with some reservations; 0 percent stated they gave their opinions, but with many reservations because of the little time they had been taught by this instructor; 0 percent stated they had this instructor for such a little time they felt they could not form any opinions about him (no ratings were made by these men).

*Items Surveyed*

II. Knowledge of subject matter.
III. Organization of lectures
IV. Clarity of presentation of subject.
V. Effectiveness in use of training aids.
VI. Ability to arouse and maintain interest
VII. Completeness of answers to student questions.
VIII. Understanding students' ability to absorb and use material.
IX. Voice quality
X. Ability to state thoughts fluently.
XI. Adequacy of explaining new vocabulary and technical words.
XII. Control of Class
XIII. Developing class interest in navigation.
XIV. Attitude of helpfulness toward cadets.
XV. General effectiveness of instructor in putting subject across.

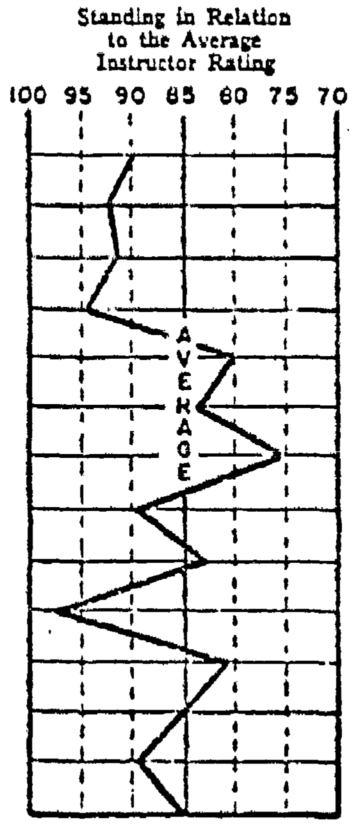



FIGURE 6.1.—Report of cadets' opinions regarding instructors.

## EXPLANATION OF REPORT OF CADETS' OPINIONS REGARDING INSTRUCTORS

These expressions of opinion by cadets should in no way be interpreted as "efficiency ratings." They are only an indication of the learner's evaluation of the instructor and as such, reflect only the learner's point of view. These expressions of opinion should be tempered by the judgment of experienced navigation supervisory personnel as to the qualities desired in the instructor and the extent to which such opinions reflect the attainment of these qualities.

It should be remembered that the expressions of opinion indicated on the reverse side of this report are based on the opinions of the cadets in one particular flight. Had other cadets, receiving instruction from this instructor, indicated their opinion, some change in the average expressed opinion might take place. However, it is believed that ratings obtained from class to class will be roughly comparable.

The value of these expressions of opinion lies in the fact that the learner is in an excellent position to evaluate the effectiveness of an instructor's presentation. The instructor should inspect his relative standings on each of the items from II through XV and note his strong and weak points. By being aware of his weak points and attempting to improve his work on these points, he will be able to increase the effectiveness of his instruction. Likewise, the instructor should note his standing relative to the average instructor. If the instructor is generally below average, he should attempt to evaluate the effectiveness of his instruction and to improve those particular points which seem to cause the learner the greatest difficulty. On the other hand, if an instructor is generally above the average he can feel that his instruction is quite effective.

Item I gives the percentage of the flight stating differing degrees of certainty of opinion about this instructor.

The graph represents the position of this instructor on each item from II through XV in relation to the average rating given by all student raters to all instructors in the school. The midpoint of 85 marks the school average for that item. The point representing this instructor on any particular item represents his standing in relation to the average rating for all instructors for that item.

FIGURE 6.2.—Explanation of report of cadets' opinions regarding instructors.

## Instruc[illegible] in Bombardier Training Schools[a]

### *Development of Tests and Rating Scales*

The situation with respect to the selection and training of bombardier instructors was quite similar to that for navigators. The first three cadet bombardier classes at Lowry Field provided the instructors for the establishment in 1941 of bombardier schools at Barksdale and Ellington Fields. For a number of classes at these fields it was also necessary to retain a number of the graduates as instructors. The bombardier schools were instructed to provide an additional 3 weeks' [illegible]aining for graduate bombardiers who were to be retained as instructors.

In January 1943 the Cen[illegible]al Instructors School for Bombardiers was established at Carlsbad Army Air Field. Initially the classes were composed of instructors and supervisors from the various bombardier schools but by the middle of 1943 combat-experienced bombardiers were being sent to Carlsbad for instructor training. As the result of the policy of replacing instructors who had not served in combat theaters, all returned bombardiers ordered to the Training Command were sent to the Instructor School. In October 1943 the Central Instructors School (Bombardier) was moved to Midland Army Air Field.

One of the first projects of the Psychological Research Project (Bombardier) when it was established in January 1944 at Midland Army Air Field was the study of bombardier instructors. The project was directed to make an analysis of the duties of the bombardier instructor. On the basis of interviews with supervisors, instructors, and aviation cadets and detailed observation of instruction on regular training flights a tentative list was developed of seven important aspects of the task of the bombardier instructor which it was believed could be measured by psychological tests. These seven traits included (1) attitudes and opinions, (2) verbal intellectual ability, (3) mechanical comprehension, (4) knowledge of basic subject matter, (5) numerical ability, (6) instructional judgment, and (7) biographical information.

On the basis of this analysis of the duties of the bombardier instructor, a battery of tests was assembled and tried out experimentally by the Psychological Research Project (Bombardier). These tests included:

1. *Arithmetic Reasoning Test, DG501A.*—A 40-item printed test designed to measure nonverbal reasoning ability. The problems were multiple-choice arithmetic reasoning problems.

2. *Bombardier Basic Knowledge, DS441A.*—A 50-item printed test designed to measure practical knowledge of bombing subjects. The multiple-choice items include such problems as setting up the auto-pilot and the use and operation of the sight.

3. *Explanation Blank, DS301A.*—This test included a number of para-

[a]This work was done by E. H. Kemp, W. W. Griggs, K. B. Levan, R. J. Larson, J. W. Birch, and L. B. Coad.

graphs followed by questions regarding the quality of the explanation provided in the paragraph.

4. *Mechanical Comprehension, DS201A.*—A 40-item multiple-choice test designed to measure understanding of fundamental mechanical principles. The items were accompanied by diagrams and drawings designed to illustrate everyday mechanical problems.

5. *Opinion Questionnaire, CE519A.*—A 90-item questionnaire of the multiple-choice type prepared by Psychological Research Project (Bombardier). The items covered attitudes toward instructing, promotion, paper work, discipline, and similar topics.

6. *Personal Inventory, DE201B.*—A 45-item inventory revised from the Shipley Inventory. It was scored by an empirical key developed by comparing combat returnees diagnosed as having operational fatigue with cases diagnosed as normal.

7. *Scientific Background, CI812A.*—A 60-item test prepared by Psychological Research Project (Navigator) and designed to measure knowledge of physical principles and background of general science.

8. *Verbal Comprehension, DG101B.*—A 40-item test of reading comprehension similar to those used in the AAF Qualifying Examination and the Air-crew Classification Test Battery. Several paragraphs were presented followed by multiple-choice items designed to test understanding of the material.

9. *Bombardier Proficiency Test, DS442A.*—A 90-item test based on a proficiency test developed for selecting lead personnel in the Eighth Air Force and the Bombardier Basic Knowledge Test.

Certain of these tests were selected for administration in the instructor battery at the redistribution stations of the AAF Personnel Distribution Command. These included all of the nine tests mentioned above except the Explanation Blank and the Scientific Background and Bombardier Basic Knowledge Tests. The Opinion Questionnaire was slightly revised and renamed the Air-crew Officer Blank for this Purpose. These six scores were combined to obtain a Bombardier Instructor Stanine.

To evaluate the success in the Central Instructors School for bombardiers, the Over-all Instructor Rating was used. This rating was a composite including the following grades with the weights indicated: ground technical-training grade, 5; teaching techniques for aerial instruction grade, 3; aerial-training grade, 1; technical aerial-training grade, 1. These combined scores were recorded in terms of ratings of superior, excellent, very satisfactory, satisfactory, and unsatisfactory. For purposes of the analysis of the Instructor Selection Test the two top categories were compared with the three bottom categories. The bombardier instructor students were fairly evenly divided between these two groupings.

To obtain measures of the actual success on the job as a bombardier instructor, the personnel of the Psychological Research Project (Bombardier) visited several bombardier schools where standardized interviews were conducted with supervisory and cadet personnel. The purpose of these studies was explained to these individuals and they were assured that their ratings would be kept strictly confidential.

Ratings on each instructor were obtained from between three to six qualified training supervisors, including both combat veterans and men who had seen no combat duty at each of the cadet bombardier schools. These ratings were made in terms of a simple 5-point scale on which they rated the instructor as being among the highest 20 percent of instructors they had known in all-around work and ability as an instructor, the next highest 20 percent, and so on down to the lowest 20 percent. Each instructor who was rated had been assigned to the school as an instructor for at least 2 months. Care was taken to obtain ratings under as nearly independent conditions as possible.

Ratings were obtained from an average of about three cadets who had known the instructor for a period of at least 1 month in these same schools. The cadet rating forms requested ratings on (1) ability to explain clearly, (2) ability as a bombardier, (3) personality as an instructor, and (4) over-all effectiveness as an instructor. Each of these was presented to the cadets in the form of a question followed by five descriptive statements. The cadet was asked to select the statement which most accurately described the bombardier instructor being rated.

Ratings of these types were obtained for 332 instructors by supervisors and 272 instructors by cadets. These instructors were distributed among four bombardier schools. In addition to these ratings, information was obtained at the field concerning the officer's efficiency rating, his score on the comprehensive Bombardier Proficiency Test, and his score on the Standard Aerial Check if available.

### *Analysis of the Validation Data*

These various ratings were then correlated with Bombardier Instructor Stanines and the scores on the individual tests that were available. A summary of these coefficients is shown in table 6.3.

Correlations between the various tests and the ratings made by supervisors and cadets were surprisingly high in view of the findings just reported for pilot and navigator instructors. The correlation for the Bombardier Instructor Stanine with the Over-all Instructor Rating was 0.54. With supervisors' ratings it was 0.38 and for the cadets' ratings 0.61. Although studies were not made of the reliability of the cadets' and supervisors' ratings, the correlation between these two ratings of 0.46 reported in the following table suggests a fair degree of reliability for each of these.

There are a number of factors which may explain the better predictive value of the Instructor Selection Tests in this instance. In general, these are conditions which suggest that the ratings of bombardier instructors in this situation are likely to have been more valid than in the case of the other specialties. One of the most important of these is the fact that these studies

TABLE 63.—*A summary of validation data on predictors of bombardier instructor criteria*

| Predictors | Criteria | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Instructors school | | On-the-job | | | | | | | | | |
| | Overall instructor rating | | Supervisors' rating | | Cadets' rating | | Officers efficiency rating | | prof. test (Form C) PRP(B) B. | | Standard phase check | |
| | $N$ | $r_{bis}$ | $N$ | $r$ | $N$ | $r$ | $N$ | $r_{bis}$ | $N$ | $r$ | $N$ | $r$ |
| Instructor stanine | 441 | 0.54 | 101 | 0.38 | 90 | 0.61 | 101 | 0.38 | 97 | 0.36 | 64 | 0.47 |
| Bombardier stanine | 326 | .11 | 140 | .22 | 126 | .16 | 353 | —.03 | 214 | .17 | 110 | .14 |
| Navigator stanine | 337 | .26 | 148 | .21 | 126 | .23 | 354 | .02 | 212 | .23 | 110 | .23 |
| Pilot stanine | 107 | .09 | 143 | .21 | 122 | .25 | 316 | .00 | 192 | .22 | 97 | .28 |
| Instructor preference | 441 | .34 | 101 | .20 | 90 | .20 | 101 | .27 | (b) | (b) | (b) | (b) |
| Arithmetic reasoning | 93 | .39 | 101 | .21 | 90 | .34 | 101 | —.03 | (b) | (b) | (b) | (b) |
| Bombardier basic knowledge | 121 | .29 | 101 | .34 | 90 | .36 | 101 | 36 | (b) | (b) | (b) | (b) |
| Explanation blank | 90 | .26 | 80 | .12 | 71 | .14 | 80 | .07 | (b) | (b) | (b) | (b) |
| Mechanical comprehension | 122 | .29 | 101 | .18 | 90 | .34 | 101 | .30 | (b) | (b) | (b) | (b) |
| Opinion questionnaire | 124 | .30 | 126 | .56 | 90 | .46 | 101 | .43 | (b) | (b) | (b) | (b) |
| Personal inventory | 119 | .29 | 101 | .15 | 90 | .16 | 101 | .12 | (b) | (b) | (b) | (b) |
| Scientific background | 180 | .27 | 64 | .13 | 37 | .43 | 41 | .20 | (b) | (b) | (b) | (b) |
| Verbal comprehension | 122 | .31 | 101 | .11 | 90 | .37 | 101 | .19 | (b) | (b) | (b) | (b) |
| Over-all instructor rating | ... | ... | 230 | .09 | 90 | .40 | 112 | a.34 | 93 | .38 | 58 | .26 |
| Supervisors' rating | 230 | .09 | ... | ... | 90 | .46 | 44 | .34 | 103 | .15 | 36 | .04 |
| Cadets' rating | 90 | .40 | 90 | .46 | ... | ... | 44 | .34 | 46 | .08 | 30 | .24 |
| Officer Efficiency rating | 112 | a.34 | 44 | .34 | 44 | .34 | ... | ... | 96 | .19 | 79 | .20 |
| Bombardier professional test (C) | 93 | .38 | 103 | .15 | 46 | 08 | 98 | .19 | ... | ... | 65 | 21 |
| Standard phase check | 58 | .26 | 36 | 04 | 30 | .24 | 79 | 20 | 65 | 21 | ... | ... |

a $r_{tet}$
b No data.

were carried on in the midst of a vigorous campaign throughout the bombardier training program to evaluate and improve instructor personnel. All instructors as well as all cadets were required to take the Bombardier Proficiency Test and the fact that the average scores for instructors and cadets were almost identical in most of the schools was given wide publicity.

The Bombardier Training Standardization Board was also active during this time, visiting the schools and giving various instructors the Standard Aerial Check Ride. The supervisory personnel were encouraged to check ride their instructors with the same standard check. Supervisory personnel were being urged to weed out the inferior instructors. It is believed that all of this contributed to encouraging supervisors to learn something about the knowledge, ability, and personality characteristics of their instructors.

It is possible that the bombardier-school cadets had an opportunity to observe a larger number of bombardier instructors and therefore had a sounder basis for giving a comparative rating. Clearly, if the cadet had never had but one instructor, he had little means for judging the relative competence of his instructor. Another factor which may have been important in the evaluation of bombardier instructors was that all ratings were obtained during an interview by a representative of psychological Research Project (Bombardier). This procedure should have resulted in a clearer understanding on the part of the raters and probably a more careful job of rating.

Another contributing factor may have been the smaller restriction in the range of abilities present in the returned bombardier group. For the bombardiers represented in the groups of 326 and 327 entering into the analyses of Bombardier Stanine and Navigator Stanine with Over-all Instructor Rating the mean Bombardier Stanine is 5.75 and the mean Navigator Stanine is 5.36, and standard deviations are 1.66 and 1.65 respectively. The level of intellectual ability was definitely lower for this group of bombardier instructors than for similar groups of navigator instructors and the range was correspondingly less restricted than would usually be found in a similar group of navigator instructors.

Although the numbers of cases involved in these studies are too small to permit definite conclusions, they offer considerable encouragement for further research on problems regarding the selecting and training of instructors for bombardier-training schools.

## INSTRUCTORS IN FLEXIBLE GUNNERY TRAINING SCHOOLS[4]

### *Early Studies and Procedures*

The general problems of selecting and training of instructors for flexible-gunnery schools were quite similar to those for the other air-crew specialties. Gunners were typically enlisted men rather than officers and their instructors were similarly in enlisted status as a general rule. There were three types of gunnery instructors: range instructors, classroom instructors, and air-firing instructors.

The range instructors supervised the firing of the students using various types of weapons, sights, and turrets on the ground ranges. The classroom instructor had a more typical instructional position in a fairly conventional classroom situation with special emphasis on training films and other visual aids. The air-firing instructor had what was generally considered to be the most important of the instructor positions. It was also a very difficult one because of the number of students to be supervised, the cramped interiors of the planes, and the lack of familiarity of the students with the general problems and conditions of flying. In addition to a heavy schedule of aerial flights, instructors held preflight briefings and postflight critiques. Because of its greater importance for the primary job of the flexible gunner, plans and

[4]The work described in this section was done by R. N. Hobbs, L. M. Stolurow, A. L. Irion, G. R. Pascal, T. P. Gallagher, J. D Adams, A. Lubin, M. A. Haire, W. S. Gregory, J. M. Wepman, A. C. Jensen, W. B. Schrader, K. G. Bainbridge, M. Waldman, P. Freeman, H. W. Goldstrom, J. V. McQuitty, W. H. Angoff, H. B. Karta, K. B. Henderson, E. E. Ketcheroid, M. G. Lieberman, G. J. Wischner, J. H. Neely, and J. L. Joynes.

studies regarding the gunnery instructor problem were usually oriented about the problems of the air-firing instructor.

As in the case of the other specialties, sufficiently valid measures of student performance were not available to permit the evaluation of gunnery instructors in terms of the proficiency of their students. Therefore, a careful analysis was made of the duties of instructors and four qualities were selected as being of prime importance for this assignment. These qualities were: (1) the intelligence of the instructor, (2) the amount of gunnery knowledge possessed by the instructor, (3) the degree of personal adjustment of the instructor and (4) the ability of the instructors in putting to use the accepted principles of teaching.

The Instructor School for Flexible Gunnery at Buckingham Army Air Field was one of the first instructor schools established in the Army Air Forces. When a psychological research detachment was assigned to this school early in 1943 one of their first projects was a survey of the personnel being sent to the Instructor School by each of the six gunnery schools. It was found that the mean score on the Army General Classification Test for approximately 200 students who had graduated from flexible-gunnery schools and were assigned to the Instructor School in February 1943 was 119. More than 20 percent of the group had scores on the Army General Classification Test of 108 or below. It was also found that the knowledge of flexible gunnery of the men sent from various schools varied over a considerable range.

This study led to a directive regarding the assignment of students to the Instructor School which specified that they should have an Army General Classification Test score of 120 or above except for those students who had received special training in combat, had teaching experience, or had received four years of college training. It was also specified that they be high-school graduates and make a score on the Comprehensive Test in Gunnery above the seventy-fifth percentile of regular gunnery students.

This procedure appears to have operated quite satisfactorily in procuring for instructor assignments a high quality of graduates from the flexible-gunnery schools. As more combat returnees became available the policy was established for this as for other specialties that these instructors were to be released for combat duty and replaced with men who had combat experience.

### *Development of the Instructors Qualifying Examination*

To select the most suitable personnel from among the combat returnees with experience in flexible gunnery, the psychological research group with the Central School for Flexible Gunnery prepared the Instructors Qualifying Examination. This examination was prepared in two main parts. Booklet I covered aptitudes considered important for the gunnery instructor and included subtests of vocabulary, mechanical aptitude, mathematics, reading comprehension, spatial visualization, and teaching aptitude and personal adjustment. Booklet II was devoted to knowledge of gunnery and included subtests on aircraft recognition, sighting, weapons, and turrets.

In addition to these two booklets the Mechanical Comprehension Test, DS201A, used for selecting other instructors at the redistribution stations was included in obtaining the final Gunnery Instructor Stanine in order to supplement the mechanical section of the aptitude booklet. The tests of gunnery knowledge were mostly selected from the Gunnery Final Examinations used in the flexible-gunnery schools. The tests for the aptitude booklet were similar to air-crew classification tests of the types mentioned with the exception of the subtest on teaching aptitude and personal adjustment. This consisted of 40 items which were the discriminating items remaining after several empirical screenings of a wide variety of items on attitudes, opinions, preferences, and personal traits. These items were retained because of their ability to discriminate between individuals who obtained high and low ratings in practice teaching as measured by a rating scale or between individuals who obtained high ratings in the courses on student teaching as compared with those with low ratings in this course.

The ratings for practice teaching were obtained from a 14-item scale. It was found that this scale which included such items as fluency, poise, clarity of expression, use of teaching aids, ability to arouse interest, and evidence of preparation had a reliability coefficient when based on the judgments of two independent raters equal to 0.90. This value was based on a comparison of the ratings of 155 students by each of two raters. The coefficient was increased by use of the Spearman-Brown formula to represent the reliability of their combined ratings.

*Use of the Instructors Qualifying Examination*

To check on the validity of the Instructors Qualifying Examination, scores on Booklet I and Booklet II of this test were compared with scores achieved on the final comprehensive gunnery examination at the Instructor School and the ratings on basic gunnery teaching and on practice teaching at this school. Data were obtained for 800 cases enrolled in classes at the Instructor School in February 1945.

For the Final Comprehensive Examination the correlations were 0.57 and 0.62 respectively for Booklets I and II. For Basic Gunnery Teaching Ratings the corresponding correlations were 0.24 and 0.12 and for Practice Teaching these values were 0.19 and 0.08. These correlations are the values obtained after correcting the original values for the restriction in the sample at the Instructor School. These corrections were accomplished using the standard deviation of the sample of people tested in the redistribution stations of AAF Personnel Distribution Command on the total score for the Instructors Qualifying Examination. This value was 26.25 and the corresponding value for this restricted sample of 800 cases was 17.22.

The complete table of intercorrelations for all of the subtests of the Instructors Qualifying Examination and the various parts of the criterion measures is given in table 6.4.

By means of the various parts of the Instructors Qualifying Examination it was found possible to obtain a multiple correlation of 0.72 with the Final

TABLE 6.4.—*Validity coefficients, intercorrelations, and coefficients of reliability of the Instructor Qualifying Examination Battery and three criteria including part and total scores (N = 675 for intercorrelations; 200 for validities)*

VARIABLES[1]

| Variable | Var. No. | M | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *75 Final Comp.:* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Weapons | 1 | 7.31 | 3.90 | (64) | 56 | 45 | 30 | 29 | 82 | 17 | 01 | 34 | 47 | 39 | 34 | 30 | 03 | 45 | 26 | 42 | 41 | 42 | 39 | 44 | 62 |
| Sighting | 2 | 7.28 | 3.02 | 56 | (41) | 42 | 45 | 38 | 79 | 20 | 13 | 41 | 45 | 54 | 39 | 31 | —02 | 53 | 30 | 54 | 47 | 37 | 56 | 46 | 62 |
| Turrets | 3 | 9.45 | 4.78 | 45 | 42 | (72) | 14 | 11 | 73 | 14 | —04 | 33 | 44 | 33 | 30 | 25 | 16 | 40 | 16 | 30 | 36 | 34 | 40 | 19 | 46 |
| Teach Tech. | 4 | 6.87 | 2.60 | 30 | 45 | 14 | (28) | 40 | 57 | 14 | 13 | 25 | 24 | 34 | 32 | 21 | 10 | 37 | 20 | 32 | 27 | 22 | 33 | 28 | 39 |
| T. Aids & Proc. | 5 | 4.38 | 2.14 | 29 | 38 | 11 | 40 | (23) | 50 | 13 | 05 | 26 | 26 | 33 | 31 | 20 | —01 | 33 | 18 | 28 | 29 | 23 | 33 | 31 | 39 |
| Total Score | 6 | 35.29 | 10.66 | 80 | 79 | 73 | 57 | 50 | (77) | 22 | 08 | 43 | 54 | 55 | 45 | 35 | 01 | 57 | 30 | 53 | 57 | 44 | 62 | 52 | 63 |
| *Teaching:* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Basic Gun Teach. | 7 | 2.87 | .56 | 17 | 20 | 14 | 14 | 13 | 22 | (a) | 28 | 20 | 15 | 18 | 15 | 10 | 13 | 24 | 06 | 13 | 08 | 06 | 12 | 15 | 21 |
| Practice Teach. | 8 | 5.31 | 1.72 | 01 | 13 | —04 | 13 | 05 | 08 | 28 | (a) | 11 | —02 | 20 | 10 | 08 | 22 | 19 | 01 | 13 | 07 | 00 | 08 | 01 | 13 |
| *Inst. Qual. Bk. I.:* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Vocabulary | 9 | 9.64 | 3.64 | 34 | 41 | 33 | 25 | 26 | 43 | 20 | 11 | (66) | 43 | 47 | 45 | 32 | 11 | 77 | 25 | 38 | 30 | 18 | 42 | 48 | 67 |
| Mech. Princ. | 10 | 6.75 | 2.19 | 47 | 45 | 44 | 24 | 26 | 54 | 15 | —02 | 45 | (55) | 40 | 41 | 37 | 03 | 65 | 27 | 43 | 45 | 39 | 53 | 59 | 69 |
| Math. | 11 | 11.18 | 2.97 | 39 | 54 | 33 | 34 | 33 | 53 | 18 | 20 | 47 | 40 | (65) | 44 | 34 | 09 | 70 | 19 | 47 | 38 | 30 | 45 | 44 | 65 |
| Read. Comp. | 12 | 4.96 | 2.10 | 34 | 39 | 30 | 32 | 31 | 45 | 15 | 10 | 45 | 41 | 44 | (42) | 30 | 16 | 67 | 16 | 35 | 32 | 35 | 41 | 45 | 61 |
| Surface Dev. | 13 | 12.81 | 2.28 | 30 | 31 | 25 | 21 | 20 | 35 | 10 | 01 | 32 | 37 | 34 | 30 | (82) | 03 | 55 | 21 | 37 | 31 | 30 | 41 | 42 | 55 |
| Teaching Apt. | 14 | 24.09 | 4.35 | 01 | —02 | 16 | 10 | —01 | 01 | 13 | 22 | 11 | 03 | 09 | 16 | 03 | (61) | 47 | 01 | 01 | 06 | 04 | 05 | 07 | 21 |
| Total Apt. Bk. I. | 15 | 70.33 | 9.75 | 46 | 53 | 40 | 37 | 33 | 57 | 24 | 19 | 77 | 65 | 71 | 67 | 55 | 47 | (82) | 27 | 51 | 45 | 42 | 55 | 61 | 88 |
| *Inst. Qual. Bk. II.:* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Aircraft Rec. | 16 | 8.90 | 2.66 | 26 | 30 | 16 | 20 | 18 | 30 | 06 | 01 | 23 | 27 | 19 | 16 | 21 | 01 | 27 | (50) | 40 | 33 | 29 | 60 | 21 | 42 |
| Sighting | 17 | 11.11 | 3.35 | 42 | 54 | 30 | 30 | 28 | 50 | 13 | 13 | 39 | 43 | 47 | 36 | 39 | 04 | 52 | 40 | (54) | 50 | 39 | 76 | 44 | 70 |
| Weapons | 18 | 16.14 | 4.21 | 61 | 47 | 36 | 27 | 29 | 57 | 08 | 07 | 30 | 45 | 36 | 32 | 31 | 06 | 45 | 33 | 50 | (65) | 54 | 84 | 41 | 70 |
| Turrets | 19 | 9.60 | 3.76 | 42 | 37 | 34 | 22 | 23 | 44 | 06 | 00 | 18 | 39 | 30 | 35 | 30 | 04 | 42 | 29 | 31 | 54 | (54) | 77 | 45 | 67 |
| *Total Achiev. Bk. II.:* | 20 | 45.76 | 9.31 | 59 | 56 | 40 | 33 | 33 | 62 | 12 | 08 | 42 | 53 | 45 | 41 | 41 | 05 | 56 | 60 | 75 | 84 | 77 | (79) | 54 | 86 |
| *Inst. Qual. Bk. III: Mech. Comp.* | 21 | 17.69 | 5.12 | 44 | 46 | 39 | 28 | 31 | 52 | 15 | 01 | 48 | 59 | 44 | 45 | 40 | 09 | 61 | 27 | 41 | 42 | 46 | 54 | (64) | 78 |
| Total | 22 | 133.79 | 17.22 | 63 | 62 | 46 | 39 | 39 | 63 | 21 | 13 | 67 | 69 | 63 | 61 | 55 | 21 | 88 | 47 | 73 | 70 | 67 | 86 | 78 | (80) |

[1] Decimal points omitted for correlations. Coefficients corrected for restricted range. See text.

[a] Reliability coefficient could not be computed.

Comprehensive Examination in the instructor course, a multiple correlation of 0.25 with Basic Gunnery Teaching Ratings, and a multiple correlation of 0.32 with the Practice Teaching Ratings.

In addition to the use of the Instructors Qualifying Examination to select gunnery instructors from among combat returnees the examination was used as a part of the procedure for giving authority to certain individuals to be classified as Specification Serial No. 938, a classification ordinarily given only to graduates of the Instructor School. This examination was used as an initial screening device. Candidates were also required to obtain a minimum qualifying score on the phase checks and to be judged as satisfactory on the basis of an interview with a board established for that purpose. Those qualifying for an interview on the basis of test results and passing this board were classified as flexible-gunnery instructors. The examinations held by these boards indicated that many of the enlisted personnel who were functioning as gunnery instructors possessed rather inferior qualification for this assignment. These men were gradually replaced by combat-experienced men selected as described above.

It is clear that this program produced a substantial improvement in the general level of aptitude and information possessed by instructors in gunnery schools. Although evidence based on actual improvement in the quality of students produced by these instructors or based on actual observation and evaluation of the gunnery instructors in their instructional assignment in the schools is not available, the ratings of practice teaching in the Instructor School tend to indicate that the selection procedures possess validity for the practical side of instructions as well as the theoretical.

It is important to note that if these selection procedures posssessed only a moderate amount of validity they would still have been of great practical value because of the large numbers of instructors involved and of the extensive use of these selection procedures and the large number of men rejected by them. For example among the combat returnees tested in the redistribution stations no one was sent to the AAF Instructors School for Flexible Gunnery who obtained a Gunnery Instructor Stanine below the average for the gunners tested.

### Summary of Instructor Research

The work of psychological research personnel associated with instructor schools for the various specialties resulted in the accumulation of a substantial amount of important research experience in the field of instructor selection and training. In connection with the Instructors School for Flexible Gunnery the selection program was started early and was an important practical influence on the training program in flexible gunnery.

The research on instructors for pilot, bombardier, and navigator schools was largely concentrated on problems of selecting combat returnees for instructor duty. The findings of these groups were utilized to some extent in developing policies and in rejecting combat returnees for instructor assignments. However, their main contribution was in the research findings regarding the analysis of the instructor problem and especially on the evaluation of the effectiveness of instructors.

Since there was not a significant number of combat returnees with experience as flight engineers or radar observers, the instructor problem was not given as high priority in research in these specialties and only a preliminary survey of this area had been made by the close of the war. These surveys made it apparent that here as elsewhere procedures for the selection and training of instructors were in need of improvements. For example, in the radar-observer projects it was found necessary to conduct extensive training programs to obtain adequate standardization of the administration of examinations and proficiency checks.

The instructor is a key figure in any training program and these studies make it apparent that much additional research is necessary if satisfactory procedures for selecting and training instructors for the various Air Force schools are to be developed.

## RESEARCH ON THE CONTENT OF TRAINING COURSES

### Introduction

The primary problem of any training program is the determination and specification of the outcomes of instruction. It is believed that most of the defects in training programs are related not to inefficient learning situations but to having individuals learning the wrong things. The fundamental basis for determining what shall be learned is a thorough analysis of the operating and combat requirement of the assignments for which the individuals are being trained. Unfortunately, many training programs have to be developed without having the benefit of combat experience with the equipment and procedures which the individuals are being trained to utilize. Under these circumstances simulated combat situations must be substituted, or in some instances the programs must be originally based on a rational analysis of the anticipated requirements.

### Studies of Training Requirements Based on Combat Surveys

One of the first studies conducted in the Aviation Psychology Program regarding combat requirements was the survey of air-crew personnel in the Eighth, Ninth, Twelfth, and Fifteenth Air Forces conducted from November 1943 to March 1944 and reported in April 1944. One of the studies made in connection with this survey was an analysis giving the percent of the unsatisfactory bombing missions of squadrons and groups due to various causes. The results of this analysis are shown in table 6.5.

The bombing accuracy in these medium bombardment groups (B-26) in the Ninth and Twelfth Air Forces was higher than that of the heavy bombardment groups in the Eighth Air Force. The accuracy of the groups of the 42nd Wing was especially high. As noted above, the type of bombing and amount of opposition is quite different for the groups in the various Air Forces. This analysis of the reasons why various groups and squadrons failed to achieve

TABLE 6.5.—*Analysis showing the percent of the unsatisfactory bombing missions of squadrons and groups which can be attributed to the various reasons indicated*

| Reason | Percent | | |
|---|---|---|---|
| | 8th[1] AF | 9th[2] AF | 12th[3] AF |
| 1. *Weather:* Bombs not dropped at aiming point because of cloud cover.... | 23 | 3 | 7 |
| 2. *Smoke Screen:* Bombs not dropped at aiming point because of smoke screen | 5 | 0 | 2 |
| 3. *Enemy Action:* Bombing inaccurate because lead ship hit on bomb-run, etc. | 4 | 6 | 6 |
| 4. *Command Decision:* Bombing inaccurate because of change of plan by command pilot | 4 | 0 | 1 |
| 5. *Formation:* Bombing inaccurate because of loose formation and stragglers | 3 | 3 | 0 |
| 6. *Collision Courses:* Bombing inaccurate because of alteration in course necessary to avoid other groups on bombing run | 3 | 3 | 5 |
| 7. *Navigation:* Bombs not dropped at aiming point because of failure to locate initial point, or bombing inaccurate because of different heading caused by overrunning initial point | 16 | 0 | 2 |
| 8. *Target Identification:* Bombing inaccurate because aiming point not located or identified too late | 20 | 34 | 21 |
| 9. *Bombsight Technique:* Bombing inaccurate because of inferior technical skill or gross errors in using the bombsight | 11 | 48 | 33 |
| 10. *Emotional Control:* Bombing inaccurate because of nervousness and anxiety interfering with effective action on the bombing run | 7 | 3 | 5 |
| 11. *Bombing Equipment:* Bombing inaccurate because of failure on bombsight or bomb-release mechanism on bombing run | 4 | 0 | 18 |
| 12. *All reasons for failure* | 100 | 100 | 100 |

satisfactory results on particular bombing missions includes only missions on which bombs were actually dropped by the groups on enemy territory.

Since only unsatisfactory missions were included, the table does not necessarily indicate that a particular difficulty was more frequently encountered in one Air Force than in another, but merely that the unsatisfactory results in one Air Force could be attributed in a larger proportion of the missions to this reason than they could in another Air Force. For example, weather, smoke screens, command decisions, and navigation difficulties accounted for a larger proportion of the failures in the Eighth Air Force than in the Ninth or Twelfth Air Forces.

The findings summarized in this table contain a number of implications for the training program. For example it was stated that lack of adequate discipline was chiefly responsible for straggling formations and collision courses. Ordinarily these may be taken as failures on the part of pilots in subordinate command positions. They appear to reflect lack of adequate training in leadership.

There was a substantial number of mission failures in the Eighth Air Force because of navigation errors. Many of there errors were attributed to pilotage, though other aspects of navigation also contributed. Bombardiers in all three Air Forces had considerable trouble with target identification. This and similar reports emphasizing these failures on the part of trained bombardiers resulted in greater emphasis being placed on target identification in the training schools.

One of the techniques developed for providing training in target identification was camera bombing. For these missions bombardiers were briefed to bomb a particular building or factory or road intersection. The briefing was

[1]Data from all groups in combat operations at the beginning of the period for missions from 1 September 1943 to 31 December 1943.
[2]Data from the 357th Bombardment Group (Medium) for missions from 15 August 1943 to 15 January 1944.
[3]Data from medium bombardment groups of the 42nd Wing for missions from 1 December 1943 to 31 January 1944.

accomplished on the ground in the usual fashion with photographs and maps and the missions were flown simulating to a greater degree combat missions. The bombardier was required to identify a target and make a bomb run on it. A series of pictures were taken at intervals during the bomb run and at the moment of release of the bomb. These photographs made it possible to determine where bombs would have dropped had they been dropped at the release point.

The analysis also indicated that there were many errors in bombsight technique. Further studies of the types of errors made suggested changes in the emphasis given to various aspects of the bombardier's procedures in the training schools.

In addition to this analysis of reasons for mission failures, this survey included interviews with group and squadron commanders and operations officers regarding the training of the replacement personnel they were receiving. Their principal complaints were in regard to the failure of training missions, particularly in operational training, to simulate closely combat missions. It was stated that while the crews were training as units, many of the individual members of the crew had so little opportunity to practice their specialties that they were unable to maintain proficiency. These officers reported that pilots did not have sufficient training in using the automatic pilot in cooperation with the bombardier. It was also reported that pilots did not have adequate knowledge of the engines and other mechanical equipment in the airplanes.

A general point of great practical importance for the Army Air Forces training program was the extent to which the men in air-crew training should be taught theory and principles as well as procedures. It is much simpler in a mass training program to teach everything by rule and as a series of routine mechanical procedures to be done in a certain sequence rather than to attempt to provide a basic understanding of the reasons for various elements of procedures to produce certain results.

The officers in the combat theaters regarded it as very important that air-crew personnel understand the basic principles for the operation of their equipment rather than learn procedures and rules. It was reported that in combat operations the individual frequently found himself in a situation in which rules had to be violated if the mission were to be successfully completed. Under these circumstances it was very important for the individual to know the basis for the rules so as to have a full knowledge of the consequences of violating various rules in a situation where equipment, lives, and the success of the mission were at stake. There is some evidence that the type of flexibility in operations which comes from a complete understanding of the theory and principles of the combat operations of their equipment was the principal feature which distinguished American pilots from those of certain enemy countries during the past war.

In addition to this initial survey, studies of implications for training were made by each of the Air-Crew Evaluation and Research Detachments which visited combat theaters at later dates. The findings of these groups were communicated to training authorities for use along with other data in revising training curricula.

Surveys were made by the psychological research projects in the AAF Training Command of training requirements for combat groups by using questionnaire and interview procedures. The navigator project administered a 150-question "Survey of Conditions in Combat" to 618 navigators who had participated in combat operations. The results of this and similar studies by other groups were reported to Headquarters AAF Training Command and other supervisory organizations and personnel.

In conjunction with a survey of operational activity and requirements conducted by the psychological branches in the redistribution stations, which is discussed in detail in a later chapter, several hundred air-crew officers reported on specific errors in combat operations which resulted in mission failures. The men also suggested training procedures which might have prevented these specific errors. The bombardiers' suggestions were chiefly concerned with improved training in target identification and in techniques for handling malfunctions in equipment, especially at high altitudes and very low temperatures. The navigators suggested that their major errors might have been prevented by more closely simulated combat missions, improved training in procedures for crew coordination, better training in pilotage, and more adequate information in the prevention and correction of mechanical failures in navigation equipment. The bomber pilots suggested that more formation flying under simulated combat conditions would have been helpful. They also believed that better training in commanding and coordinating a crew, more knowledge of equipment and engines, and experience in leading a formation would have been valuable. The suggestions of fighter pilots were mainly concerned with more information and experience on combat-type planes, more training in navigation and instrument flying, and more practice on strafing and dive bombing.

## Studies of Amounts of Training and Learning Curves

Another fundamental problem in developing the training program was the relative emphasis to be given to the development of skill in using equipment as contrasted with the providing of information about the equipment. In the early phases of the war it was believed that there was too much emphasis on learning about machine guns and turrets in flexible gunnery and not enough emphasis on skill in using the type of equipment which his combat assignment would require.

A number of research studies were carried on by personnel of the Aviation Psychology Program to determine the effect of spending varying periods of time in training on final level of proficiency reached.

### *Studies in Pilot Training*

The staff of the Psychological Research Project (Pilot) took advantage of a 5-week training "freeze" to measure the effect of 5 weeks of training on the pr ciency of students at various levels of training. More than 8,000 pilots re given standard objective check rides which had been developed partic ly for this purpose. The introduction of various controls cut down the n ber of pilots included in the final analysis to a somewhat smaller num

It s found that the group with 15 weeks' training at the primary level was superior to the group with 10 weeks' training at this level on all 13 of the objective measures used. Similarly, the group with 15 weeks' training at the basic level was superior to the group who had had 10 weeks, at this level on 15 out of the 17 objective measures. For the group with additional training in advanced schools, superior scores were made on 10 of the 16 objective measures. The final comparisons at the primary, basic, and advanced levels were each based on more than 1,000 pilots.

In primary schools most improvement was shown in the following measures: (1) maintaining constant altitude while making a 360-degree steep turn to the right immediately followed by a 360-degree steep turn to the left; (2) performing an Immelmann without stalling, starting with an air speed not greater than 150 miles per hour and coming out lined up within 45 degrees of the correct heading; (3) performing a loop without stalling, starting with an air speed no greater than 110 miles per hour and coming out within 45 degrees of the correct heading. For these three measures, the superiority of the 15-week group was of a size which would be expected to appear by chance less than one time in a hundred.

For students in basic training schools the group with additional training was superior at this level of significance on the items relating to: (1) ability to hold heading, air speed and altitude during straight-and-level instrument flight; (2) making a 360-degree steep instrument turn maintaining a bank of at least 30 degrees, deviating no more than 100 feet in altitude and coming out on a heading not more than 5 degrees off that assigned; (3) making an instrument constant-speed climb and descent, holding air-speed deviations to no more than 5 miles per hour and heading deviations to no more than 10 degrees and leveling off within 50 feet of the correct altitude at the top of the climb and the end of the descent.

For the students with additional training in advanced schools 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 also of a size which could be expected by chance less than one time in a hundred.

The conclusions which can be drawn from these findings are rather limited but they indicate that it is possible to obtain an appreciable amount of improvement by additional training for these particular maneuvers.

An analysis made of students who had soloed in light planes before starting military flying training and of groups with different amounts of training at basic and advanced levels in various types of planes tended to confirm the

703323—47—12

general conclusion that additional training at an early stage of the training program provides an apparent immediate gain in proficiency at the next level of training. However, this gain does not persist but decreases at each succeeding stage till by the time the training course has been completed and the students are tested in their ability at fixed gunnery, no perceptible difference persists in favor of groups who had soloed and had considerable light-plane experience before entering primary flying school. From these results it may be inferred that a general principle in developing such training programs is to abbreviate the elementary phases as much as possible and give as large a proportion of the training as can be arranged for on planes which closely resemble those to be used in combat operations.

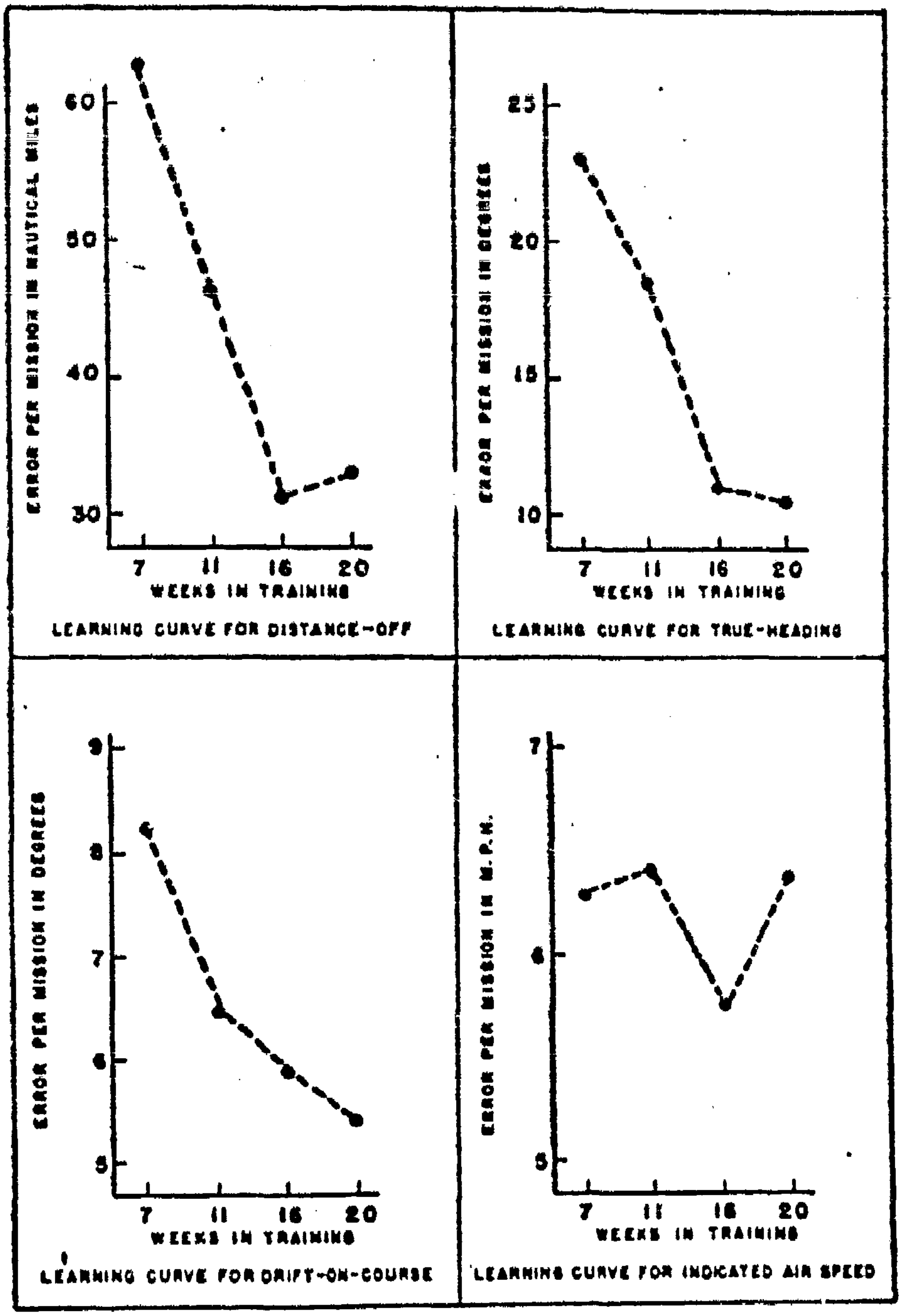
### *Studies in Navigator Training*

Psychological Research Project (Navigator) carried out a study involving 80 cadets undergoing navigation training at Ellington Field. This study consisted of comprehensive and detailed measures obtained from formation flights of the proficiency of navigation cadets in performing precision dead-reckoning ground plots. The standardized missions were flown in the seventh, eleventh, sixteenth, and twentieth weeks of training. It was found that for most of the variables the average error score was decreasing up to and including the fourth standardized mission, which took place during the twentieth week. It appeared that a longer period of training would result in further increase in proficiency. The slopes of certain typical learning curves are indicated in the charts shown in figure 6.3.

In the navigation schools, dead-reckoning navigation was taught for the first 10 weeks. This was followed by instruction in celestial and radio phases of navigation. The slope of the learning curves suggested that precision dead-reckoning ground-plot missions spaced at intervals throughout the latter part of the course would have materially increased the level of proficiency in this basic form of navigation. The curves also suggested that little improvement was being obtained with respect to certain measures such as the estimation of the air speed. On the other hand, it appeared that much further improvement might have been made in such variables as heading and drift-on-course. Greater emphasis in the training program to these latter two variables might enable a higher general level of proficiency at the completion of the course.

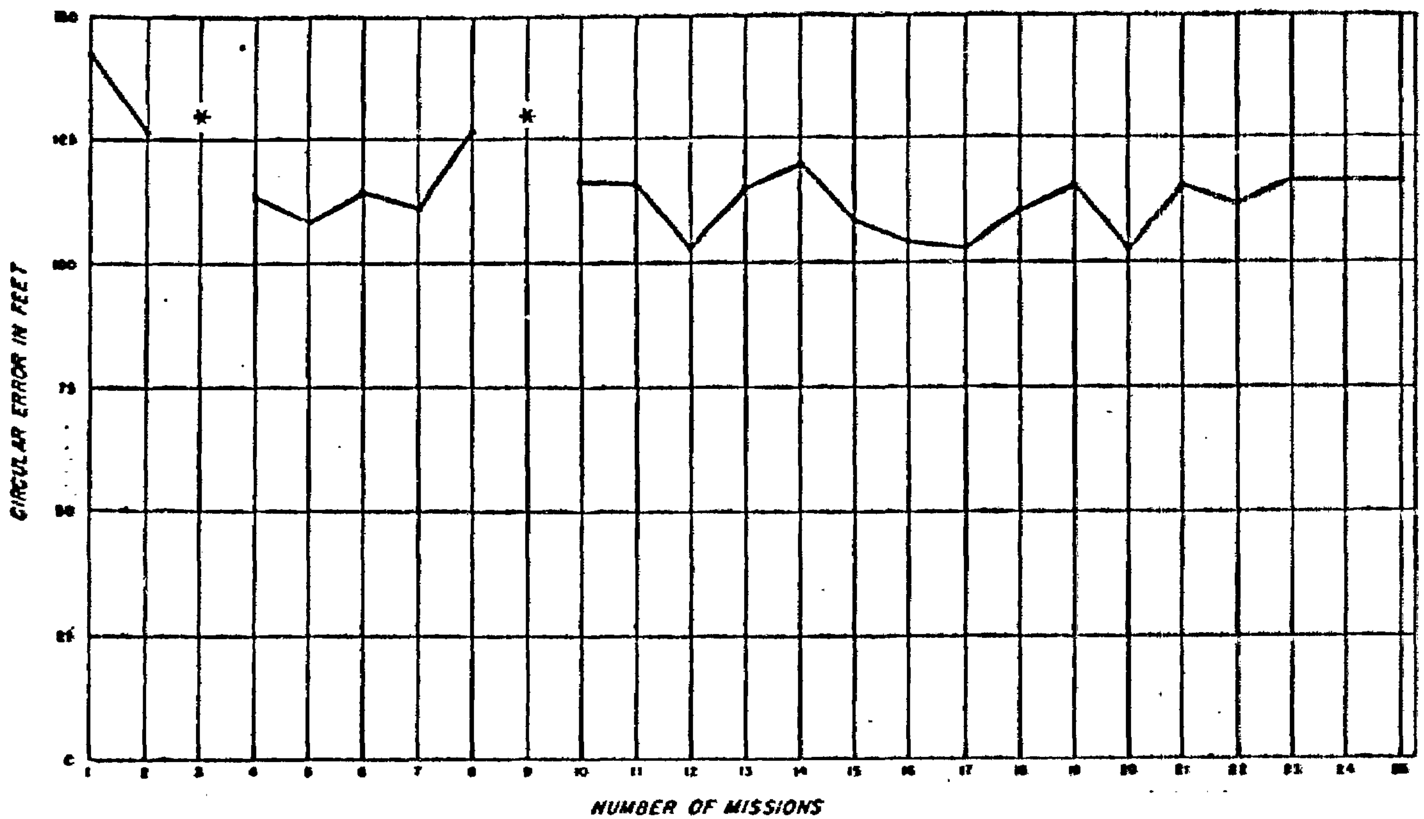
### *Studies in Bombardier Training*

The Psychological Research Project (Bombardier) carried out a study of improvements in the circular error of bombs dropped in the bombardier-training course. The 100 students for this special class were selected from a pool of preflight-school graduates awaiting entry into bombardier-training schools. Approximately equal numbers of average and superior men were included in the sample. With only two or three exceptions they had received no previous aerial training and none of them were eliminees from pilot training. Their bombardier stanines ranged from 5 to 9 with a mean of 7.15 and a standard deviation of 1.14. Their navigator stanines ranged from 4 to 9



ERROR PER MISSION IN NAUTICAL MILES
60
50
40
30
7
11
16
20
WEEKS IN TRAINING
LEARNING CURVE FOR DISTANCE-OFF
ERROR PER MISSION IN DEGREES
25
20
15
10
7
11
16
20
WEEKS IN TRAINING
LEARNING CURVE FOR TRUE-HEADING
ERROR PER MISSION IN DEGREES
9
8
7
6
5
7
11
16
20
WEEKS IN TRAINING
LEARNING CURVE FOR DRIFT-ON-COURSE
ERROR PER MISSION IN M.P.H.
7
6
5
7
11
16
20
WEEKS IN TRAINING
LEARNING CURVE FOR INDICATED AIR SPEED


FIGURE 6.3.




*Majority of bombs dropped at 4,000 feet.

FIGURE 6.4.—Class learning curve by missions, circular errors are unconverted values bombing altitude—7,000 feet.

with a mean of 6.28 and a standard deviation of 1.41. Their pilot stanines ranged from 1 to 9 with a mean of 5.80 and a standard deviation of 2.01.

This class began training at Big Spring Army Air Field late in June 1945. To standardize conditions for this experiment a group of about 30 instructors, 35 pilots, 30 airplanes of one year and model (with C-1 auto-pilots installed), 50 newly reconditioned Norden bomb sights and 6 targets on the bombing range were set aside for the exclusive use of the experiment.

The first phase of the experiment was concerned with a study of learning on ground trainers. Six days were spent on the A-2 trainer followed by 14 days on the A-6 trainer. On this latter trainer, an accuracy check was obtained on each of the first 10 days based on the circular errors under standard conditions on six circular targets. The average circular errors for the group on days 7, 8, 9, and 10 are all at least a fifth of a standard deviation below any of those obtained on the 4 preceding days when accuracy checks were given. During the last 3 days on this trainer, complex conditions were used which caused a slight increase in the circular error. This error was reduced perceptibly on the second day but did not drop on the third day using complex conditions. It can probably be concluded that this is approximately the right amount of training on the ground trainer before aerial training is introduced.

For the study of the learning of aerial bombing, missions were run during the first week for the purpose of auto-pilot familiarization and simulated bombing runs but no bombs were dropped during this period. With certain exceptions due to weather conditions all bombs during the first phase of the studies were dropped from an altitude of 7,000 feet with all missions scheduled in the morning. Auto-pilot was used except when malfunctions were found after take-off. The average circular errors for the bombs dropped on each of the first 25 missions are reproduced in figure 6.4.

It is clear from this figure that very little learning occurred under these conditions after the first three or four missions. The results for the bombs dropped at 4,000 feet also tend to confirm this finding. It was concluded from this study that where more time for training is available it could probably more profitably be employed in target identification, camera bombing, offset bombing, bombing through overcast, and operational bombing. The 150 bombs ordinarily dropped during the course of training appeared to be quite adequate for the development of a satisfactory level of bombing accuracy. Considerable variation in altitudes and other conditions during the course of this training would probably improve the quality of the bombardier in terms of ability to do combat-type bombing.

### *Studies in Radar-Observer Training*

Learning studies were also conducted by the Psychological Research Project (Radar Observer). These studies were based on data which had already been collected and it was found difficult to control all of the factors in the learning situation. These studies suggested that the radar-observer training program then in effect was of insufficient length to develop a satisfactory degree of skill in radar bombing.

A carefully controlled experiment in extended training for radar observers was conducted at Victorville by Project SC-70, NS-146 of the National Defense Research Committee. In this experiment 20 radar observers who had received the full course of radar-observer training, including approximately 30 to 35 hours' training on the scope, received approximately 140 hours, additional training. A substantial amount of learning occurred during the first 40 hours of extended training and definite though less marked improvement was obtained during the next 40 hours. The average circular errors decreased from over 3,000 feet at the beginning of the period of extended training to approximately 1,600 feet at the end. It should be noted that the individuals making up the crews with which the various radar observers bombed were maintained relatively constant for each particular radar observer throughout this experiment. The results tend to confirm the findings of the Psychological Research Project (Radar Observer) to the effect that the course could be considerably improved by increasing the amount of bombing practice given to the radar observers.

***Studies in Flexible-Gunnery Training***

A number of studies in learning were conducted by the psychological research group in the flexible-gunnery program. These included a substantial amount of research on learning on ground trainers. One of these, the Reflectone Trainer, was used to give instruction in range estimation. It was found that, under the conditions of the classroom, gunners learned little after about 150 settings of the aparatus and that most learning in estimating range on this trainer occurred during the first 30 settings. On the basis of these experimental results it appeared that approximately 8 hours of training time could be saved.

For most of the other ground trainers studied it was found that a significant amount of improvement was obtained in the scores during the training period. The periods were usually too short to establish the point at which further practice would not result in improvement. These data were probably not crucial, however, because of the high degree of specificity found for most of these ground trainers.

Studies of improvement in aerial-gunnery performance were made using gun-camera missions. In most of these studies the personnel of the Department of Psychology of the AAF School of Aviation Medicine cooperated with the psychological research group at the Central School for Flexible Gunnery. The first study of learning of this kind was a study of learning in a group of graduate basic gunners during 10 successive gun-camera missions. Sixteen Martin turret gunners and 16 Sperry upper-turret gunners were used in this study. All missions were carried out under carefully controlled conditions. On the basis of the results it was concluded that graduate basic gunners showed no consistent improvement during 10 carefully controlled gun-camera missions.

In a later study a group of 54 men, 18 Sperry gunners, 18 Martin gunners, and 18 waist gunners were given a series of 10 gun-camera missions. The purpose of this experiment was to determine the nature of learning for an untrained group of gunners. These untrained gunners showed a significant degree of improvement during the 10 gun-camera missions. Most of the improvement occurred during the first three missions. Comparing the results of the trained and untrained gunners, it was found that in most respects the proficiency of untrained gunners was lower during the first few missions than that of the trained gunners. However, during the last five missions the untrained gunners, both Sperry and Martin, are on the average slightly, though not in all comparisons significantly, better than the graduate gunners. It appears that insofar as ability in gun-camera missions is concerned, the training received in the basic gunnery course was balanced by three or four gun-camera missions, or about 3- or 4-minutes of actual turret operations in the air.

Another study carried out in flexible-gunnery schools was the comparison of the 6-week training course with an experimental 8-week training course. In terms of proficiency measures applied at the end of the training course, the 8-week group appeared to be slightly superior. However, when both groups were examined later at a training-air-force station there was no evidence of superiority for the 8-week group. Surveys indicated that the morale of the 8-week group was better than that for the 6-week course group.

## Analytical Studies of Training Results

An example of an important type of research study providing information regarding the desirable content of training courses is supplied by the Psychological Research Project (Navigator). This study was a by-product of the previously mentioned research on the measurement of proficiency and learning on navigation flight missions. The end result of most practical navigation problems is a position report. This results in an error score called distance-off.

This research study undertook to determine the relation of errors in the various steps involved in a dead-reckoning navigation mission and the final errors in terms of distance-off. For this study the intercorrelations of errors in the following navigation variables were computed: track-made-good (TMG), drift-on-course (DOC), true heading (TH), magnetic heading

TABLE 6.6.—*Intercorrelations between 6 navigation variables for 165 students who were in their seventh week of training at Selman Field*

| | TMG | DOC | DEV | TAS | WF | WD | GS | DO | *M* | *SD* |
|---|---|---|---|---|---|---|---|---|---|---|
| TMG | | 0.11 | 0.76 | −0.03 | 0.12 | 0.17 | 0.06 | 0.78 | 15.40 | 8.04 |
| DOC | 0.11 | | .01 | .00 | .72 | .27 | .53 | .12 | 5.10 | 3.35 |
| DEV | .76 | .01 | | −.02 | .07 | .17 | .11 | .67 | 13.53 | 8.10 |
| TAS | −.03 | .00 | −.02 | | .04 | −.13 | .28 | .12 | 8.20 | 6.65 |
| WF | .12 | .72 | .07 | .04 | | .12 | .57 | .18 | 10.30 | 4.10 |
| WD | .17 | .27 | .17 | −.13 | .12 | | .27 | .18 | 135.00 | 47.00 |
| GS | .06 | .52 | .11 | .28 | .57 | .27 | | .30 | 10.23 | 9.76 |
| DO | .78 | .12 | .67 | .12 | .18 | .15 | .30 | | 31.40 | 14.80 |

(MH), compass deviation (DEV), indicated air speed (IAS), calibrated air speed (CAS), true air speed (TAS), wind force (WF), wind direction (WD), ground speed (GS), distance traveled (DIST), time (TIME), and distance-off (DO). These correlations are shown in table 6.6.

On two smaller samples of 40 cases each the correlations between true heading and magnetic heading with compass deviation were found to vary between 0.90 and 1.00. Similarly, the correlations between indicated air speed and calibrated air speed were found to be over 0.90 in each case and the correlations of these two variables with true air speed were found to be between 0.63 and 0.83. Distance traveled had correlations of 0.48 and 0.82 with ground speed in the two smaller samples and time had correlations of 0.50 and 0.58 in the two samples with distance traveled.

It is clear from an inspection of the coefficients in the accompanying table that errors in distance-off are highly correlated with errors in track-made-good and deviation which are also substantially correlated with each other. Errors in drift-on-course, true-air-speed, wind-force, and wind direction have a very small relation with distance-off, and errors in ground-speed are only moderately related to distance-off. These general findings were confirmed by the correlations obtained on 4 different missions by a group of 80 students in navigation.

Using track-made-good and ground speed, very good predictions were obtained for distance-off. Using the relatively independent basic navigation variables, including deviation, drift-on-course, and true air speed, it was also possible to obtain substantial multiple correlations with distance-off. These multiple correlations for the 5 groups including nearly 500 cases were about 0.70. Substantially larger correlations were obtained between track-made-good and ground speed in predicting distance-off. However, these two variables are used directly in computing the position report, and therefore this large correlation indicates only that very few errors were made in measuring and plotting the distance and in the final calculations necessary for obtaining distance-off.

It will be recognized that the most important outcome of this study was the finding of the predominate importance of errors in compass deviation in determining errors in the final solutions of dead-reckoning navigation problems. On the basis of this study Psychological Research Project (Navigator) recommended that more emphasis should be placed on obtaining accurate estimates of compass deviation in the training courses. It was also recommended that modifications be made in the procedures taught regarding navigational practices. The students were taught to make readings every 5 minutes on most of the navigation instruments in the plane such as the compass, air speed meter, thermometer, altimeter, and drift meter. The usual procedure was to take only one of two shots with the astro-compass. The estimate of compass deviation was based on these shots. It was therefore recommended that more reliable estimates of compass deviation be secured by making more frequent observations with the astro-compass.

It was found that some of the errors made in obtaining estimates of compass deviation from the astro-compass were due to the apparatus itself. It was therefore recommended that the design of the instrument be modified in certain specific ways to avoid these types of errors.

This analysis of the relation between errors in obtaining measures of specific variables used in dead-reckoning navigation and errors in the final position report is seen to have provided valuable basic data for the modification of the navigation training course. It is believed that the wider application of this research method to training problems would be of great value to the Air Forces.

A somewhat similar study of one specific aspect of the bombardier's job was carried out by the Psychological Research Project (Bombardier). A preliminary survey had shown that very few bombardiers were able to compute altitudes correctly. Therefore, in developing the section on altitude computation of Form B of the Bombardier Proficiency Test, items were prepared covering each of the various steps in the procedure. An item analysis based on 1,494 bombardier students revealed the main sources of difficulty. It was found that 58 percent of students made errors in computing the true bombing altitude when they were given the target pressure altitude, the target temperature, and a view of the bombardier's instruments including free air temperature gauge, air-speed meter and altimeter. The survey pointed to two serious lacks, first, an adequate understanding of the fundamental principles of altitude computation, and second, clerical facility in the simple arithmetical steps involved.

## STUDIES ON TRAINING DEVICES AND EQUIPMENT

### Introduction

A fundamental problem in the planning of all air-crew training courses was the relative emphasis to be given to ground and air training, particularly with reference to the use of equipment. It was, of course, much cheaper to provide training on the ground and it was also feasible to provide very much more training than was usually possible on an air mission. Ground trainers could be used regardless of weather and certain types of changes in conditions could be introduced very readily. There was also no dependency on a large crew of other individuals and theoretically at least conditions could be more easily standardized and the proficiency of the individual students more accurately measured.

In view of these advantages many synthetic training devices were produced and used. However, in many cases doubts arose in the minds of instructors and training authorities regarding the value of these devices. Personnel of the Aviation Psychology Program were therefore frequently asked to evaluate the effectiveness of specific training devices in producing desired training results. This was especially true during the latter part of the war period.

## Study of a Navigation Training Device

A typical study of this type was the experimental evaluation of the G-2 (Dead-Reckoning) Trainer by the Psychological Research Project (Navigator). This trainer consisted of a master control unit and 48 individual booths. Each booth accommodated one student and contained the instruments necessary for performing dead-reckoning navigation: an air-speed meter, compass, altimeter, temperature gauge, clock, and drift meter. The instruments on the panels of the individual booths were controlled from a master control panel. Because of the differences of opinion regarding the value of this apparatus as a training device, a preliminary study was made comparing the flight-mission grades of students having much practice on the trainer with average grades of students having little or no experience with the trainer. No differences were found between the groups but since a number of factors were not controlled in this comparison authority was requested to perform a more thorough experiment.

Accordingly, an experiment was designed to test the hypothesis that groups trained only in the classroom would not differ significantly from groups receiving extensive training on the G-2 Trainer. Four flights of approximately 40 students each were divided into an experimental and control group, half of each flight being in each group. In the classroom groups the navigational ground problems were presented by either providing the data on the blackboard or on a large mock-up in the front of the room. In the G-2 Trainer the ground problem consisted of "navigating" according to data presented on the instrument panels.

The two groups were evaluated by three especially graded flight missions. The differences between average errors made by the two groups on these missions were compared with the differences which could have been expected to arise by chance if no real difference in the two groups existed. This analysis indicated no consistent differences favoring either one or the other of the experimental groups. An analysis of variance was also used to test the likelihood that the differences had arisen because of sampling fluctuations. The results of this analysis failed to cast doubt on the hypothesis that no difference existed between the trainer and classroom groups.

The results of this study indicating that simple blackboard presentations or classroom mock-ups are not significantly inferior to the very expensive ground trainers of the type here evaluated is of considerable value both in this specific situation and also because of its broader implications for training situations in general.

## Evaluation of a Novel Skeet Sight for Fighter-Pilot Training

A somewhat similar study of one small aspect of training in fixed gunnery was carried out by the Training Aids Section of Aloe Army Air Field in cooperation with the Psychological Research Project (Pilot) which carried out the analyses of the data. This research study was designed to evaluate the effectiveness of a self-reflecting optical sight as an attachment to the conventional shotgun for use in teaching the sight picture for air-to-air fixed gunnery. The sight gave the students when firing the same sight picture that is obtained in fixed-gunnery training. It was believed that increasing the similarity of the skeet firing to fixed-gunnery firing in this way should increase the transfer of training.

Two groups in fixed-gunnery training were compared. The 145 students using the optical sights in skeet training obtained an average fixed-gunnery score of 32.2 in terms of percent hits in air-to-air fixed-gunnery firing. For the 144 students using the conventional shotgun sight in skeet training the average percent hits was 29.0. This difference of a little more than 3 percent hits was sufficiently large as to make it unreasonable to suppose that no difference existed between the two groups. These findings encouraged the introduction of this optical sight into skeet training as a valuable and useful contribution to the improvement of fixed-gunnery scores.

## Evaluations of Ground Trainers in Flexible-Gunnery Training

The Department of Psychology of the School of Aviation Medicine and the personnel of the psychological research group in the Central School for Flexible Gunnery cooperated in an evaluation of the usefulness of two ground trainers, the Sperry Spotlight Trainer and the DeVry Panoramic Gunnery Trainer. Sixteen Sperry gunners were given a series of three carefully controlled gun-camera missions to establish an initial proficiency level for the group. This was followed by 2 weeks of intensive ground training on the Sperry Spotlight Trainer. After this ground training the 16 men were given three more gun-camera missions. Similarly 16 Martin turret gunners were given 3-gun-camera missions and 2 weeks of intensive ground training on the DeVry Panoramic Trainer followed by 3-gun-camera missions.

It was found that practice on the Sperry Spotlight Trainer decreased the tracking error of a Sperry gunner on succeeding gun-camera missions but his framing error remained the same. On the other hand, the Panoramic Gunnery Trainer did not significantly improve the performance of Martin gunners in the gun-camera missions which followed their training period. There was a slight indication that the effect of ground training was greatest on the first gun-camera mission following training. Difficulties in the mechanical functioning of the trainers may have limited their effectivenss in this experiment. If these difficulties are general as is suggested by other experience with these types of trainers, this would also have a similar limiting effect on their usefulness under ordinary training conditions.

An even more comprehensive experimental evaluation of ground-training devices was provided by the comparison of the gun-camera scores obtained by students who had completed the basic gunnery-training course and individuals who had had no training in basic gunnery. This study indicated that after five gun-camera missions there was no discernible difference in the performance of the trained and untrained groups on gun-camera missions. The error scores for the untrained gunners are slightly better for both the Sperry and Martin turrets in most respects than the scores for the gunners as to preference and strength of preference for a specific type of training were found to have a small correlation with subsequent records of success in a particular activity. However, much more substantial correlations with training success were obtained between items on the Biographical Data Blank as mentioned above and also items on the General Information Test. The General Information Test measured the individual's knowledge of special types of activities. The items on this test were found to provide more dependable measures of the individual's real functioning interests than the reports of biographical history and preferences on the Biographical Data Blank.

It was found that these measures of basic interests and values were only moderately correlated with specific aptitude-test scores and could therefore be used to obtain a substantial improvement in the prediction of success in the various types of activities. It is believed that further exploration of this field would be profitable.

## THE NATURE AND SIGNIFICANCE OF FACTORS OF PERSONALITY, TEMPERAMENT, AND INDIVIDUAL ADJUSTMENT

### Introduction

Of the various determiners of success in military positions, probably the most has been written and the least is definitely known about the factors of personality, temperament, and individual adjustment. The military situation develops and accentuates emotional problems. The uniformity of general environment and activities which exists for large numbers of men makes it possible to study these problems in a much more controlled situation than is typical of ordinary life.

The military situation is therefore especially well adapted to the study of anxiety reactions which constitute an important problem in the services in time of war. Problems of character and temperament also occupy an important role in military activities and valuable opportunities to study traits in these areas are provided.

### The Nature of Anxiety Reactions

Repeated strong emotional reactions arising from apprehension of a threat to something valued tend to induce an anxiety reaction in which the individual shows the usual symptoms of the psychoneurotic.

Early analyses of reports of faculty-board proceedings stating causes for elimination of aviation cadets from pilot training indicated that a very large proportion of these cadets showed some signs of tension, nervousness, and apprehension during the period prior to their elimination from training. This was thought to indicate that personality and temperament factors were of great importance in determining success or failure in flying training. However, repeated studies failed to indicate as substantial a relationship as was anticipated from the analysis of the board proceedings. Further study of the situation suggested that the lack of progress in learning to fly was not so much the result of the tension and apprehension but that to a greater extent the tenseness and emotional symptoms were produced by the student's knowledge of his lack of progress.

Studies of individuals showing anxiety reactions to combat stress indicated that this anxiety was typically a learned phenomenon following all the principles of learning usual for the acquisition, reinforcement, and extinction of associations. One of the most striking of the findings regarded the specificity to the original inducing situation of anxiety symptoms. As mentioned in the previous chapter, a pilot or bombardier who showed severe anxiety symptoms when flying in the usual place in the airplane was frequently able to ride through difficult combat missions as a tail gunner with only very mild symptoms.

Available data suggest, however, that there do exist individual differences in the ability to undergo combat stress without exhibiting emotional reactions beyond the normally accepted limit. This factor appears to be definitely secondary in importance in causing breakdowns in comparison with other factors in the situation including leadership, amount of stress, and administrative procedures used in the combat situation. This predisposition is minor but not negligible and can probably be predicted in terms of what might be called the tough and tender temperament types. The sensitive, sedentary, and highly cultured individual appears to be much more susceptible to combat stress than the rough, active, and more primitive type of person. The findings also suggest that men with greater intellectual ability are better able to control their emotional reactions. Unfortunately, it was not possible to obtain a more thorough confirmation and definition of these findings during the war period. It is hoped that further light may be shed upon this topic, however, by certain data which are still to be analyzed and are currently in the files of the School of Aviation Medicine.

The combat returnees reported that the symptoms of anxiety reactions which provided their most severe problems were related to sleep and dreams, memory, depression, worries and fears, and concentration, decisions and plans, nervousness, and smoking and drinking, in the order named. These reports tend to supplement and confirm the information regarding the nature of the anxiety

Fortunately, the findings during the war period suggest that if adequate aptitude and motivation are assured the problems of personality, temperament, and individual adjustment are comparatively minor ones. Evidence for this is provided by the very small elimination rates for fear and own request in flying training and also the very low breakdown rate in combat operations. However, the effects of anxiety are serious in individual cases and have far more important later consequences than a little awkwardness in a landing or a mistake in arithmetic caused by slight deficiencies in specific aptitudes.

## The Control of Anxiety Reactions

Anxiety can best be controlled by reducing the individual's subjective estimate of the likelihood of the undesired event's occurring and by decreasing the valuation placed by the individual on the thing which is threatened.

In the preceding chapter there is a discussion of the specific administrative regulations found effective in combat units for controlling combat reactions. There is also an analysis of reports by individuals returned from combat regarding the methods which they found most effective in controlling fear while on combat missions and also during the intervals between missions. The administrative procedures found effective were of two types: first, those including rest and recreation and the opportunity to deemphasize a vivid emotional experience; and second, those concerned with factors of motivation, leadership, and a strong group spirit.

The individuals reported that they found an example of confidence and calmness and also personal activity of great value in controlling fear. The men reported little immediate value from such matters as having a clear idea of what we were fighting the war for. Between missions the men reported that letters from home, getting away from the combat group, sleep, music, and the movies were most helpful in relaxing them and relieving anxiety. They indicated that athletics, work, hobbies, and radio programs other than music were not very helpful to them in providing the needed relaxation.

The above findings have immediate application in dealing with stress-producing situations in the civilian situation. As previously mentioned, the military situation provides a desirable laboratory situation for the study of these individual reactions to stress situations. The principles governing the emotional reactions and the development of symptoms in this special situation appear to be general and can be put to immediate use.

# CHAPTER TEN

# General Contribution to the Theory and Knowledge of Education and Training and the Evaluation of Effectiveness

## THE NATURE OF LEARNING

### Introduction

For the reasons stated in the first chapter, it was believed at the beginning of the war that greater improvement could be made immediately in the effectiveness of AAF operations through improved selection and classification procedures than by improvements in training. However, as selection and classification procedures became well established and this field had been fairly well worked over, attention was directed to a greater extent to procedures related to training.

The first of the psychological research groups to be activated for work with a particular training specialty was the unit assigned to the field of flexible gunnery. Partly because of the special difficulties related to problems of selection and evaluation of these men who were being trained as gunners for bombardment planes the psychological personnel of this unit spent a larger proportion of their time on problems of training than did the personnel assigned to psychological research on other training specialties.

Another group which was very active in research on problems of training and the nature of learning was the Psychological Test Film Unit at Santa Ana Army Air Base. This group conducted extensive experiments especially regarding problems of learning to recognize aircraft. They also conducted a study to evaluate the motion-picture film as a training device.

A number of other units, chiefly the psychological research projects in the AAF Training Command, and the psychological groups in the Continental Air Forces, conducted studies of training and of the evaluation of effectiveness in performing air-crew tasks. These were reported in chapter 5 and chapter 6. Some of the general implications from these studies are reported in the paragraphs that follow.

### Course Content

The content of training courses is likely to be much less effective than could reasonably be expected, unless it is based on a systematic analysis of the requirements of the activity for which the training is intended to prepare the individual. Studies of the training courses in a wide variety of situations tended to confirm the generalization that most of the defects of training programs were related not to inefficient learning situations, but to having individuals learning the wrong things.

In all training and educational programs the fundamental basis for determining what shall be learned must be a thorough analysis of the requirements of the activity for which the educational training and courses are designed to fit the individual. It should also be emphasized that under modern conditions of rapid change, not only must a systematic analysis of requirements be made, but this must be continuously revised in the light of new conditions.

Too often in our educational institutions, especially those concerned with elementary and secondary training, the course content is determined solely by an analysis of the content of the next training course for which the individual is being prepared. Improved education must depend on a comprehensive, scientific analysis of the critical requirements of working and living in our modern civilization.

It is believed that the most important benefit which education can derive from the recent war experience is an increased awareness on the part of leaders in education of the need for empirical evaluations of educational outcomes in terms of the requirements of life itself. A new approach to the development of objectives for education is needed. The job-analysis procedures of Dr. W. W. Charters provided an enumeration and tabulation of the various activities comprising a particular job. This was a valuable initial step in approaching the study of desirable educational objectives. The defect was that the hypotheses and theories which grew out of these studies were never experimentally evaluated.

The procedures for defining educational objectives utilized by Prof. Ralph W. Tyler stimulated teachers to think about the goals of education and to gather in conferences and attempt to clarify their thinking regarding these objectives. This procedure has been valuable in that it provides and defines the objectives which educational experiences hope to achieve so that others may criticize these objectives or attempt to evaluate the success attained in achieving them. Unfortunately, the typical teacher cannot hope, from the limited observations of his personal experience in a profession somewhat insulated from the commercial and industrial work of the world, to establish satisfactory objectives without the aid of extensive scientific studies of the content of typical life activities. Therefore, this procedure cannot be expected to provide an adequate foundation for establishing educational outcomes.

The first step suggested is a careful analysis of adult activities including commercial, civic, cultural, vocational, and avocational functions. The next step is the development of a preliminary set of *critical requirements* for these activities. An aptitude, proficiency, habit, or other requirement for a particular activity which is crucial in the sense that it frequently is a factor determining successful or unsuccessful participation, provides the definition for a *critical requirement* for that activity.

Having established definite hypotheses concerning the nature of these critical requirements, it is essential that tests and other research procedures be developed with which to evaluate these hypotheses. This is a long-term project. Progress cannot be expected if continued reliance is placed on the uncoordinated efforts of isolated individuals, organizations, and institutions. It is strongly recommended that a vigorous attack on this fundamental problem of education be made utilizing the techniques and procedures which were found successful in the Aviation Psychology Program of the Army Air Forces.

## The Principles of Learning

In developing training plans and in conducting courses certain general principles which govern learning must be carefully followed if training is to be effective.

One of the most widely accepted of the general principles regarding learning is that interest and motivation are essential for efficient learning. Closely related to this is the principle of purposive student activity. This latter point was illustrated by the study of learning in aircraft recognition under conditions of frequent repetition of the stimuli as contrasted with the same number of repetitions, but in each case having the student respond. This is also closely associated with the general importance of knowledge of results. It is important not only that the student make an active response, but that this response be reinforced or corrected immediately.

Another principle frequently neglected in developing training plans is the decrease in proficiency which can be expected if practice in a particular skill is discontinued. Follow-up studies of gunners indicated that, with respect to a number of measures of proficiency, they were at their peak at the end of the individual training course in the AAF Training Command. Their proficiency was lower by the time they completed the course in the Continental Air Forces and still lower after they had been in the combat theater for a substantial period of time. In planning training courses, provision must always be made for maintaining proficiency as long as the particular skill is needed.

A related, but less frequent, defect in training courses is the provision of more practice on a particular skill than is necessary to develop and maintain proficiency. A similar defect of training courses is the tendency to overlearn a skill in a specific situation before introducing variations. A given level of proficiency can be attained more rapidly on a complex or variable task if variations are introduced before the slope of the learning curve tends to flatten out. Thus, in bombing training the typical program included a somewhat longer period on one set of conditions before changing to a new set than was optimal for most efficient learning.

Certain studies were made regarding learning on apparatus tests used in the Air-Crew Classification Test Battery. One of the findings of most general theoretical interest was the significantly superior performance of individuals on typical pursuit coordination tasks on the first trial following a rest period. This same superiority was observed on the first trial following an interval of a day or a week after an initial series of learning trials. The score on this initial trial was uniformly better than for the last preceding trial by an amount greater than would correspond to a smooth learning curve extrapolated from the preceding trial. A similar finding was the relatively greater extent of the learning which occurred if the interval between trials was made of more appreciable duration so that more adequate time was allowed for rest, reflection, and plan of approach. Additional studies providing an explanation of the controlling principles of this phenomenon would be very desirable.

Experience in the Aviation Psychology Program in the Army Air Forces indicated that not only was it desirable to have knowledge regarding the principles of learning at the time training plans were being developed but that it was also of great importance to demonstrate the actual extent of these factors in training situations. It is very easy for the expert, especially if he is not thoroughly familiar with the particular skill for which training is being given, to overlook a detail which has an important bearing on the application of the principles of learning to that specific situation. The demonstration of the application of the principles is also very important from the point of view of convincing less technically trained individuals of the effectiveness of these principles in the situation involved.

### Training Equipment

One of the most important findings with respect to learning was the surprising degree of specificity of learning on various superficially similar gunnery trainers and on various apparently similar skills involved in flying and similar air-crew tasks. These findings point very strongly to the need for a careful analysis of the critical requirements for a particular activity and also emphasize the importance of actual empirical demonstration of the effectiveness of a particular training procedure. Numerous instances were found in which training failed entirely to produce the desired skill on the actual task to be performed in the air crew. In gunnery especially, much of the work on ground trainers seems to have been largely wasted effort. In a few instances it was found that training actually produced interference so that untrained individuals were superior to the "trained" ones.

It is essential that in selecting training equipment and materials a systematic check be made to insure that the training devices simulate the real equipment and situation in all important respects.

A good example of a thorough analysis of the effectiveness of a training procedure is provided by an experiment conducted by the Psychological Test Film Unit comparing the effectiveness of a motion-picture film, an illustrated lecture, and a training manual. This study not only demonstrated the superiority of the motion-picture film in this situation, but also investigated the reasons why the training film produced greater learning than the other two methods. This study revealed that the motion-picture presentation was superior in dealing with changing events or with the variation of one thing in relation to another.

A valuable incidental finding was the apparent value of utilizing the subjective "point of view" in the filming of motion pictures. By moving the camera back to a position where the situation was viewed from the position of a trainee in the learning situation, the student viewing the film could easily imagine himself in the actual activity rather than outside it as an onlooker. By showing the situation from a position in which only the hands of the character and his manipulations are visible and pacing the film slowly, it was possible to provide the student a large amount of active mental practice which could be immediately reinforced or corrected. These findings are discussed in greater detail in chapter 6.

### The Relative Importance of Aptitude and Training

Differences in individual aptitudes are much more important than differences in training for producing differences in individual effectiveness in the typical training situation. In the pilot-training situation there was a great deal of variation in the quality of instruction given by different instructors in the same training schools. Observations indicated that there was even greater difference between schools. The program expanded so rapidly that, especially in primary pilot training, individual civilian contractors were given a large amount of freedom in developing their training programs. The training manuals available were not very extensive at the beginning of the war period, and therefore a large amount of individual initiative was used in developing methods and materials for instruction. In spite of this situation, differences in individual proficiency in basic and advanced training schools were determined to a much greater extent by the individual's aptitude as indicated by his scores on the air-crew classification tests than by the particular instructor or school in which he took his primary flying training.

Similarly, follow-up studies of bombardiers in combat indicated that the combination of individual aptitudes used in computing a particular bombardier's stanines were correlated to the same extent with average circular error in combat operations as were the average circular errors made by these students in the AAF Training Command Schools.

The conclusion from these findings is that a greater improvement can be gained in effectiveness in the activity for which the individual was selected and trained by research on selection and classification than by research on training procedures. This generalization must, of course, be qualified with respect to the initial quality of the procedures of both types and the extent of improvement which can be expected in those procedures as a result of research studies. However, this generalization appeared to hold for a large variety of training situations in the Army Air Forces.

# SUCCESSFUL OR EFFECTIVE BEHAVIOR AND ITS EVALUATION

## Introduction

A necessary prerequisite to definitive research in regard to selection or training is a measure of the degree of success of individuals in the specialty for which they have been selected and trained. Usually in a practical situation there are a number of measures of success available for the various aspects of the performance of the duties of the specialty. These are referred to as criteria for evaluating success, or, more commonly, criteria.

In a few situations where the task is simple and repetitive, it is very easy to obtain a precise measure of the degree of success on the job. Some types of radio operators' jobs are of this simple type. The criterion in these cases is comprehensive, well defined, and easily obtained in quantitative form. Most types of behavior are not so easily evaluated. In the typical situation the evaluation of success is very complex.

In many military situations the objectives were well enough defined so that an evaluation of the individual's effectiveness could be made. At the same time there was sufficient complexity to the situation to include most of the problems of obtaining an adequate criterion measure of success which would be encountered in evaluating the effectiveness of an individual in a typical civilian position.

In general, this area was found to have been much less adequately explored and developed by research workers than most of the other areas involved in a research program on the effective utilization of personnel. Because of the fundamental importance of this area and the relatively large deficiency in established knowledge concerning the field, the work done in the Aviation Psychology Program during the war period is of especially great importance.

## The Definition of Effective Behavior

Successful or effective behavior is defined as that which results in the accomplishment of the assigned mission. The factors which make it difficult to state whether a particular sample of behavior in an activity is successful or effective, and to what degree, are principally related to the definition of the assigned mission. If the task or assigned mission is defined in objective and quantitative terms, observers can agree on the degree to which the assigned mission has been accomplished and these observations provide a completely valid criterion measure of success.

Unfortunately, in the typical situation the assigned mission is defined in only the vaguest of terms. Therefore, the research worker's first task is to obtain a satisfactory definition of effective behavior in the specific situation.

All that is possible for the research worker under these circumstances is to develop a clear and concise definition of what he proposes to accept as the objectives of the behavior in the specific situation. This definition must include weights regarding the relative importance of various aspects of the task if a single criterion measure of success is to be achieved, or, in the absence of quantitative variables, complete descriptions of all possible combinations of behavior and a rank ordering of them with respect to their adequacy with respect to the accomplishments of the objectives of the activity.

The research worker must then be careful to report this definition so that other workers may interpret correctly the significance of the findings. The research worker cannot define success in an activity in any absolute way, since this involves value judgments. He can merely report the definition which he selects for use in clear terms so that others may evaluate his judgments in terms of their own values and purposes.

### Considerations Regarding the Measurement of Success

Evaluations of effectiveness should be based on direct observation of the results achieved or on the examination of systematic records of these results. All but the most elementary and simple evaluations of the effectiveness of behavior depend basically upon the judgments of observers. In many instances, the judgments are made without direct observation of the actual behavior as, for example, in the case of judgments made on the basis of photographs or other records of behavior. In some cases practically all of the critical judgments are made ahead of time so that the process of appraisal of the behavior or record of behavior becomes essentially *objective* and free of judgments on which significant disagreements between independent observers would be found.

If the activity has been adequately defined the judgments regarding what should be included have already been made before the behavior is observed. This definition must include: first, the aspects of behavior to be observed; second, standards of the relative adequacy of performance on each of these aspects of behavior; and, third, an estimate of the relative importance of each of these aspects.

In procedures in which the judgments are chiefly made at the time of the observation and not carefully defined in advance, the opinions and biases of the observer become relatively important in determining the score. Therefore, the usefulness of such scores is seriously limited.

It was repeatedly found that ratings based on general impressions rather than systematic observations were of practically no value. This was especially true in cases in which the activity was inadequately defined in the manner discussed above.

A discussion of techniques for obtaining criterion measures of success in an activity is provided in a later chapter.

# CHAPTER ELEVEN

# Contributions to the Theory and Knowledge of the Design of Equipment

## THE IMPROVEMENT OF PRECISION IN PERCEPTION OF NECESSARY ACTION

Most of the effort of aviation psychologists during the war period was directed toward the selection and training of individuals to carry out the various air-crew tasks, utilizing the particular equipment which the engineers and operating personnel had designed for these purposes. The war had to be fought with the equipment then available, and individuals had to adjust to this materiel. Experience indicated that in many cases the task of the individual had been complicated unnecessarily by arbitrary factors in the design. It is now possible to include human efficiency factors with engineering factors in planning and designing new equipment.

An example of this type of study was the evaluation of aircraft attitude indicators on the basis of Link instrument ground-trainer performance carried on at the Department of Psychology in the School of Aviation Medicine. It appeared that the most suitable type of attitude indicator was one in which the pilot observed a miniature airplane and made it conform to the maneuvers which he was trying to perform. This procedure simplified the interpretation of the instrument for him since he could immediately observe the effects of movements of the controls on the attitude of the plane. The attitude of the plane in space could be immediately perceived by looking at the model without having to go through any translation or make allowances for other conditions as was required by the types of attitude indicators commonly used during the war.

The personnel of the Department of Psychology of the School of Aviation Medicine also performe[illegible] number of experiments on the comparative legibility of aircraft dials having different markings. These studies emphasized the desirability of relatively "clean" dial faces and not as large a number of division lines as compared to the number of figures marked on the dial as were ordinarily used. It was also found that greater accuracy of reading was obtained by reducing the number of figures on some dial faces in use.

The Psychological Branch at the Aero Medical Laboratory, Wright Field, began an active program of research on these problems shortly after the cessation of hostilities. Some of the early studies have developed general principles concerning the situations in which graphs are relatively more effective than tables and vice versa. In general, tables in which a value was entered and directly read were found to be most accurate. In situations where all of the values were not tabled and interpolations were necessary, graphs were usually found more effective.

A study was made to determine those flight-control pressures which provided the greatest accuracy of "feel of controls." It was found that relative performance was poorest for low pressures and improved rapidly to a stable state between 10 and 20 pounds. Low pressures tended to be overestimated and high pressures tended to be underestimated. The performance of pilots was more accurate and constant than the performance of nonpilots.

## THE IMPROVEMENT OF CONTROL MOVEMENTS TO INSURE THE INTENDED ACTION

One of the first studies carried on at the Psychological Branch, Aero Medical Laboratory, Wright Field, indicated that there were large systematic differences in the accuracy with which an individual could reach to various positions with respect to the orientation of his body as he sat in a typical cockpit seat. Movements directly to the front or directly to the side could be made more accurately than movements at other angles with respect to the horizontal or the front-back and left-right directions.

It is believed that this type of work is especially important at the time when new weapons are being designed and the work in this area has been very active in the period following victory in Japan. It is hoped that general principles can be discovered which will enable designers to provide equipment which can be much more effectively operated by the individual than equipment designed without the benefit of this information.

## CHAPTER TWELVE

# General Contributions to Techniques of Prediction and Experimentation

### TECHNIQUES OF TEST CONSTRUCTION AND ADMINISTRATION

#### Introduction

During the war period in various schools and research groups in the military services many thousands of tests were prepared and used. Because of the lack of adequate training in this field by many of the individuals in training schools responsible for the preparation of these tests, they were of rather uneven quality. In only a few instances and for specific purposes did Aviation Psychologists undertake the preparation of achievement examinations.

The test-development work carried on in the Aviation Psychology Program was not centralized but distributed among 10 or 15 field units. The techniques used and the quality of the tests produced tended to vary with the amount and character of training in test construction which the personnel of these units had received. However, by the use of technical notes and bulletins describing useful techniques and through the circulation of reports of test-development projects among the units, some supervision was accomplished and serious inadequacies were avoided.

In certain units a substantial number of individuals spent full time on test-development work. The work of these groups and the administration and scoring of a wide variety of tests to an unprecedentedly large daily flow of men provided an opportunity to study problems in these areas over a period of time. These studies resulted in a number of contributions to this general area.

#### Types of Measures

Tests which are objective in their administration and scoring were found to be of greater predictive value than appraisals based on interviews, projective methods, observational procedures, or ratings. The advantages of simple objective test procedures with which up to as many as 500 men a day were examined in a single one of the 10 centers are so obvious as to need little elaboration. It was therefore especially valuable to find not only that these simple objective procedures would do an adequate or satisfactory job of examining, but that they were, in fact, also clearly superior in terms of predictive value, for the criteria of success available, to the less practical examining procedures which might have been used.

### *Printed Tests*

One of the simplest test forms is the multiple-choice printed test. Great economy in both printing tests and scoring can be obtained by using separate answer sheets with such multiple-choice forms. This is true for hand scoring with a punched-out stencil, which was the procedure in the case of the AAF Qualifying Examination. This examination was administered in several hundred boards throughout the country by relatively untrained examiners. It was also true for the air-crew classification tests given in the AAF Classification Centers to large daily flows and scored on the IBM electrical Test Scoring Machine.

It was found in the Aviation Psychology Program that without any known exceptions, all functions which were proposed for measurement by paper-and-pencil methods could be reduced by the exercise of sufficient ingenuity to a multiple-choice form suitable for use with a separate answer sheet and the electrical test scoring machine. This possibility was not always recognized at the outset. In many cases preliminary work was aimed at the development of a test which would not be suitable for use with separate answer sheets of a machine scorable type. A large number of such tests were tried out. These, however, were subsequently changed to multiple-choice form with no perceivable loss in their validity as a measure of the traits involved.

It was found that by utilizing photographs of maps, ground terrain as seen from several thousand feet up, instrument faces and photographs of model airplanes in various attitudes of flight, it was possible to obtain a substantial degree of realism with respect to many of the tasks involved. Even certain psychomotor functions such as "path reproduction" were adapted to use with answer sheets for the test-scoring machine.

### *Motion Picture Tests*

Extensive research, discussed in one of the later research reports of this series, was conducted in developing procedures for obtaining objective responses on separate answer sheets for tests presented in motion-picture form. Certain abilities such as the judgments of the relative speeds of objects do not lend themselves to presentation in printed-test booklets. Similarly, the division of attention between a large number of dials, the readings of each of which are changing continuously, cannot be done without the use of either apparatus or motion pictures. In the development of motion-picture tests it was also found that these could be prepared in multiple-choice form suitable for use with machine-scorable separate answer sheets.

### *Apparatus Tests*

It is believed that the motion-picture method of presentation has many important advantages and further exploration of this procedure is recommended. However, the new test form developed in the Aviation Psychology Program which was of the greatest practical value during the war was the group form of apparatus testing. Many problems had to be solved to make the mass testing of men on a series of apparatus tests a practical and useful procedure. Because of the large numbers involved and the need for uniformity, procedures were developed which enabled a single examiner to administer one of the apparatus tests to a group of four or more individuals at the same time. All tests were standardized to require precisely 15 minutes' total administration time. Timing was done automatically by electrical control units and scoring was accomplished by means of electrical time clocks or counters which were conveniently arranged on the examiner's control desk.

Another important problem which had to be solved in connection with apparatus testing was the calibration and maintenance of the testing equipment so as to provide comparable scores at all times on all of the apparatuses used. This involved the development of a number of calibration, maintenance and apparatus-control procedures which are described in a later section.

***Test Content***

Test content and the types of functions being measured were the subjects of a considerable amount of research. A few general principles have emerged from this experience.

The best types of tests for measuring an individual's aptitude are those which provide measures of a single psychological function operating at the maximum level which the individual is capable of attaining. The Discrimination Reaction Time Tests and the Numerical Operations Test tend to be tests of this type. Such measures have obvious advantages of stability and permanence not found in other types of measures. Unfortunately, it is impossible in practice to bring all individuals up to the levels which are established by their individual psychological limits for improvement in the specific function measured.

This leads to the second principle to the effect that the content of tests should be based on materials which have been learned by individuals through common experiences over a long period of time. Examples of this type of test are provided by the Arithmetic Reasoning Test and the Reading Comprehension Test. The advantages of testing people after a uniform period of learning with respect to the particular functions are clear. Measures obtained from various individuals tested at different and unknown points on the learning curve for the particular function cannot be expected to provide dependable measures of aptitudes or proficiencies for preductive purposes. For example, Arithmetic Reasoning would be a very poor aptitude test for a group in which some individuals were in the second grade and others in the ninth grade if it was desirable to make predictions regarding performance several years from the time of testing.

This also leads to the principle that wherever possible, it is desirable to measure functions, the learning curves for which are very flat. This point brings up an important limitation of apparatus tests of some of the types used in the Air-Crew Classification Battery. For some of these tests the learning which occurred during the few minutes that the test was being taken was very substantial in terms of the individual differences in the scores obtained. Under these circumstances, a few minutes' practice on the specific task invalidated the score almost completely in terms of comparison with individuals lacking this practice.

Although such limitations must be carefully considered in relation to apparatus tests and other tests similarly affected, experience indicates that the only effective procedure for determining whether the experience on the assigned task and related tasks has been sufficiently common to result in individual scores which will have predictive value is an empirical test. Certain of the apparatus tests which showed very marked learning during a very brief period of time were, nevertheless, found to provide relatively stable and valid predictive measures. It was further observed that the scores on the first 2 or 3 minutes of a test of this type were generally more valid in predicting subsequent success in pilot training than the scores obtained by these individuals during the last few minutes of the test.

The types of tests found most useful were those based on performance on a clearly assigned task which provided a direct measure or a systematic series of samples of the function or type of information being measured. It is very important that the "sets," including methods of approach and attitudes, of the various individuals toward the assigned task be as uniform as possible. It is very difficult to obtain comparable measures on individuals who have set themselves different tasks, the nature of which is unknown to the observer. For these reasons, the observation of uncontrolled behavior was found to have little predictive value in the situations studied.

A related finding which has been touched on in a previous chapter is the instability and lack of a sound rational basis typical of questionnaires and inventories of preferences and opinions.

The best information of this type can be gained by using a questionnaire which includes only items regarding matters which are established facts. In the Biographical Data Blank in the Air-Crew Classification Battery an attempt was made to include only items on which the individual was required to report a known fact regarding his background or history. These were made as objective as possible so that any other individual familiar with the details of his history would answer the form for him in exactly the same way. Items were also selected which, as far as possible, could be verified from records concerning the individual which were known to exist even though relatively inaccessible.

Various studies using a series of directions for the filling out of the Biographical Data Blank indicated that there was no observable difference in the validity of the tests when administered without any particular emphasis on the truthfulness of the responses, as compared with its validity when administered with a statement that all answers must be precise and in accordance with the facts and investigations would be made and individuals would be subjected to disciplinary action who were found to be distorting the facts.

On the other hand, individuals requested to fill out the blank in the usual manner and immediately thereafter asked to fill it out in a manner which they thought would make them appear more desirable as an applicant for pilot training were found to be able to increase their scores about 15 points, which is a little more than 2½ times the standard deviation of the distribution of scores for the total group.

### *Clinical Types of Tests*

Numerous studies were made of various subjective clinical procedures for obtaining information about individuals. Very few procedures of this type were found which had a significant amount of validity for predicting success in the air-crew specialties for which the men examined in AAF Classification Centers were being classified. Furthermore, those few procedures which were found to have a small relation to criteria of success in these tasks were found not to measure any unique elements not already covered by the objective tests of the Air-Crew Classification Test Battery.

The Clinical Techniques Project was designed to provide a thorough and intensive evaluative study using a large variety of subjective clinical procedures such as the Rorschach Test, the Thematic Apperception Test, and a number of specially designed tests. These tests, projective techniques, observational procedures, and other methods were used to study several hundred men being classified for air-crew training. None of the procedures used was found to have enough validity to suggest its addition to the battery, even ignoring the impractical nature of most of the tasks for a mass-testing procedure of this type.

At a later date a systematic administration to a large number of air-crew candidates of most published personal inventories and preference forms was made. The results of the analysis of these tests were also almost entirely negative.

As would be expected from the general principles discussed above, the type of measure found most useful in providing information regarding personality and temperament traits was the general-information type of test. This test was, in fact, found to be the most valid test of the Air-Crew Classification Test Battery. It sampled information collected over a long period of time by individuals who participated in various types of activities, largely on a voluntary basis.

### *Timed and Untimed Tests*

Speed and time limits are important factors to be considered in test development. The untimed or power test has many advantages for many types of tests. For example, if we wish to measure the amount of information that a person possesses, we do not ordinarily wish to penalize him because he reads more slowly than others or happens to stop to deliberate over a few items. What is required is the best indication possible of the number of these items of information which he has in his possession. A time limit which prevents him from providing an accurate indication of his knowledge with respect to this collection of items tends to reduce the validity of the measure in terms of the definition of the task.

The AAF Qualifying Examination was the type of examination in which it was believed that the important consideration was how many of the items the individual could answer correctly in unlimited time. The first forms of the test were therefore given without any time limit at all. Individuals could continue to work on them as long as they wished. In practice this was found administratively awkward. Although the majority of the applicants finished in about two hours and only a very small percentage of people remained after three hours, an occasional individual would sit and puzzle over the questions for as long as five hours. This was a source of considerable inconvenience since it frequently interfered with the routines of the examining personnel. For administrative convenience therefore a time limit of 3 hours was set. It was found that this was ample to provide everyone an opportunity to attempt each item, but would not in all cases permit the extended deliberation over certain items that was desired by a few individuals.

In certain types of tests the items are simple and can be done correctly by all of those participating. The score is therefore a function of the number of items which the person can complete in a fixed period of time. The Complex Coordination Test, the Numerical Operations Test, and Finger Dexterity Test are examples of such measures. Such tests need to be very accurately timed and the directions are also of great importance, since it is essential that they provide the individual with uniform motivation and also with a uniform technique. In spite of efforts to insure uniformity in these regards, it was found that some individuals, because of habits of long standing, adopted systematic approaches to the assigned tasks which were not those intended. This suggests the need for additional practice and try-out of directions.

### Statistical Procedures

In developing a test which is to be used for selection of individuals for one purpose, the criterion which takes predominance over all others is the validity of the test in terms of the correlation of scores made on the test with an over-all measure of effectiveness in the activity for which the selection is being made. Under these circumstances the reliability coefficient of the test is a purely secondary consideration. Its role is to provide information concerning the extent to which the predictive value of the test could be improved by lengthening or refining it. Such lengthening or refining procedures tend to produce a score which measures the function correlated with the criterion with less dilution of the scores by chance or specific factors of a type which would not enter into the scores obtained from an alternate form of the test or from a readministration of the test at a later date.

Similarly, the purity or simplicity as contrasted with the complexity of a test is irrelevant under these circumstances. The sole determiner of the desired degree of complexity of a test to be used for predicting a single activity should be the complexity of the trait requirements for success in this activity.

On the other hand, if tests are to be used as a part of a battery on which individuals are to be classified for a number of different types of activities, it is very desirable that the tests not only be uncorrelated or independent, but also that they be simple or pure measures. The use of independent test scores increases the efficiency of the procedures, since each trait is measured only once. Simplicity or purity is also very desirable. To illustrate this, suppose that two traits are combined in one score and neither of these traits is represented in any of the other test scores in the battery. It will then be impossible to weight these two traits in any other way than proportionally to their contributions to the variance of the test score of which they are a part.

When a test is being used as part of a test battery, it is also important that the test not only measure only one trait but that it contain as little chance variance as possible. Mathematically it is obvious that the introduction of chance variance tends to reduce the possible correlations between the combined score on the test battery and the criterion measure of success because of a reduction in the proportion of the total variance in the predicting measure which can be associated with the criterion. When the various tests of the battery are substantially intercorrelated it is not so important that each measure have a high reliability since they tend to reinforce each other.

Item analysis procedures were extensively used in the test development work in the Aviation Psychology Program. In general, it may be said that practically all of the theory which applies to assembling a battery of tests can be directly applied to the problems of assembling a battery of items. If simplicity and purity are desired in the tests, it is desirable to correlate each of the items with the remaining group of items. If the test is to be used alone in predicting success in one particular activity, the correlations of the items with the criterion and with each other are the relevant considerations.

The appropriate weight for a particular item in comparison to the weight of all of the items in a test can be very quickly obtained by using the usual multiple correlation formula if the three relevant correlations are known. Giving each item the weight obtained in this way and correlating this composite with the criterion and correlating the item with this composite provides the basic data for making a second approximation to the solution of the multiple correlation problem. This procedure was first developed by the editor in 1934[1].

[1]John C. Flanagan, *Factor Analysis in the Study of Personality*. Stanford Univ., Calif.: Stanford University Press, 1935. (Submitted as a doctoral dissertation at Harvard University, June 1934.)

In a similar way, the problem of adding items to a test parallels the problem of adding tests to a battery. Whether or not the addition of a particular item will improve the correlation between the total test and the criterion can be determined by using the three relevant correlation coefficients in the typical multiple correlation formula. The multiple correlation obtained in this way should be compared with the value obtained for the correlation between the total test score not including this item and the criterion. This procedure was tried out in selecting items for Form AC14L and subsequent forms of the AAF Qualifying Examination. Although the samples were not of sufficient size and the experimental factors not adequately controlled to provide conclusive data, the obtained differences favored the procedure which utilized the item intercorrelations.

It was found to be very important to cross-validate item analyses or item-weighting procedures of the type described in the preceding paragraph by utilizing a second sample, or if a second sample were not available by splitting the sample into halves and carrying out the procedure independently for the two halves and determining the predictive value of the score based on the analysis from one-half of the sample on the second half, and vice versa. The final weighting procedure or scoring key to be used with tests such as the Biographical Data Blank, on the basis of this type of analysis, would be based on a combination of the data from both samples.

### Problems of Administering Tests

Lack of adequate standardization of conditions of testing, including directions to students and examiners reduces the validity of most tests in current use.

Examination of the results obtained on tests at different units indicated that it was of considerable importance to standardize as much as possible the conditions of test administration at various examining units. This led to the development of a detailed set of standard operating procedures governing all procedures for testing, scoring, and recording results. Verbatim instructions were supplied to examiners for use with both the printed tests and the apparatus tests. The directions for testing specified not only what the examiner was to say, but what he was to do and how much demonstration he should provide. As a further preventive measure, detailed procedures were specified for apparatus calibration and apparatus maintenance and for auditing and checking all scoring and conversion procedures.

In addition to these attempts at standardization, control procedures were introduced to control variation between testing units in the administration of printed tests. The mean scores of individuals tested in the various units were plotted each month on a scale using standard deviation units. Although the men sent to the various units could not be regarded as random samples from a total population, the plotting of monthly raw score means by units was valuable for studying uniformity of testing conditions in the units. If the relative standing of a unit as compared with other units was different over a period of months on two tests which were measuring very similar functions, the conditions of examining at the unit were suspect. A number of local divergencies from standard procedures of scoring, timing, or administering tests were discovered by an examination of the plot of the monthly means. Sample results from this form of systematic control procedure are shown in figure 12.1.

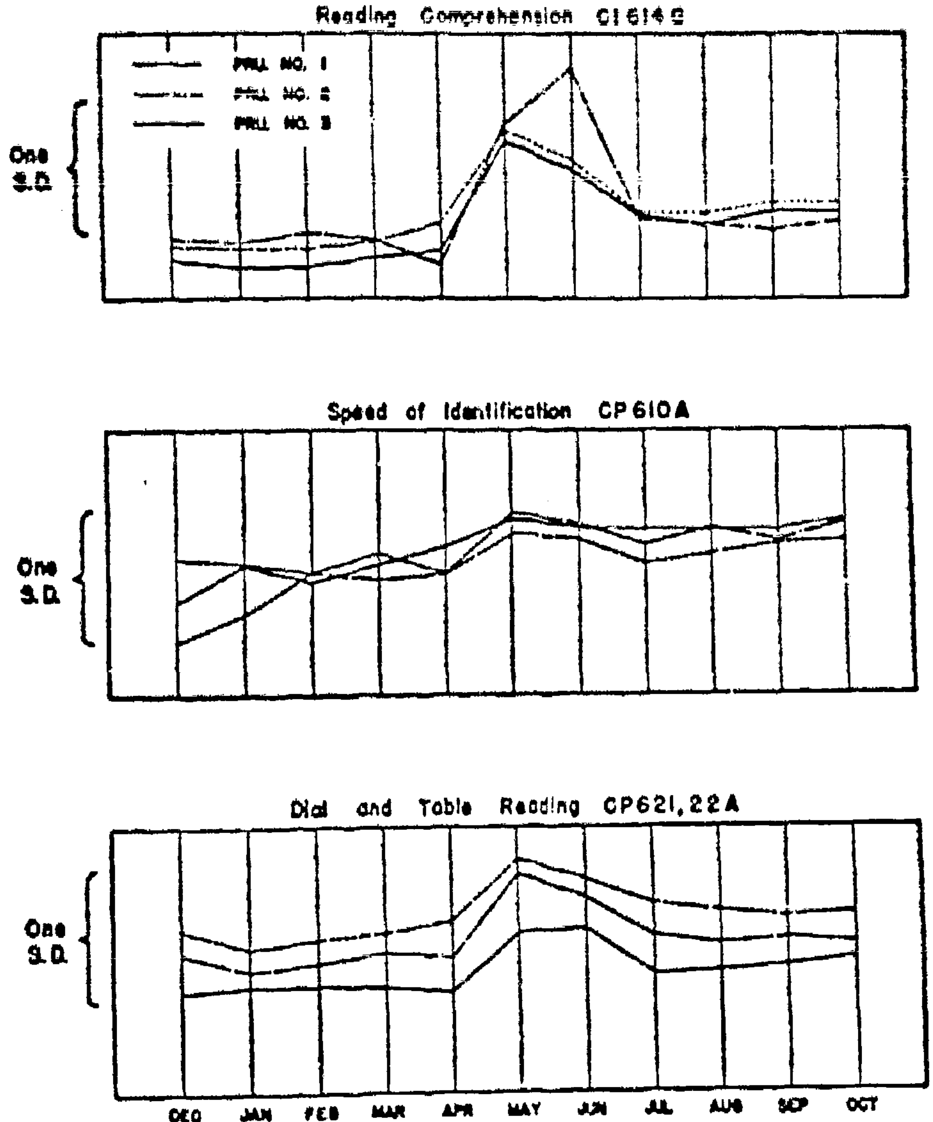



FIGURE 12.1.—Unit differences in mean test score.
(Monthly means, plotted in standard-score units.)

The monthly means for the Reading Comprehension Test, the Speed of Identification Test, and the Dial and Table Reading Test at the three psychological research units over a period of several months are shown in this figure. The Reading Comprehension Test shows a striking population shift

703323—47—19

beginning in May 1943. This was caused by the selection of those men in the AAF College Training Detachments who had the highest scores on the education examination (AC20A) for immediate assignment to classification centers and various types of air-crew training without completion of the college-training course. The variation on the Speed of Identification Test is not especially marked except for the early months at Psychological Research Unit No. 2.

The Dial and Table Reading Test shows large and consistent differences between the mean test scores at the three Units. These results, taken in conjunction with results from similar tests, made it appear almost certain that the procedures for administering the tests were not the same at all units. These differences are believed to have been the result of varied amounts of supplementation by the local units of the printed directions for administering this test. The differences appeared chiefly to have been concerned with the extent to which precautions regarding the reversal of the two scales in the table-reading part of this test were added by the local units.

These results for the Dial and Table Reading Test emphasize the importance of directions and practice exercises for homogeneous speeded tests as discussed in an earlier section. It is essential that the motivation, attitudes, and techniques used in answering the items be uniform for all individuals. It is also necessary that the task be defined in such a way that the student knows precisely the implications of various courses of action in relation to the scores.

An example of this is the type of test involving short and simple items where speed and accuracy both enter into the score but could enter in varying amounts. On such a test as the Numerical Operations Test the number of items completed using directions which emphasized speed was quite different from those completed under directions which emphasized accuracy. It must also be recognized that the directions affect individuals somewhat differently. For example, certain individuals fail to approach the test as they should under speed conditions. This situation can be improved by indicating to the students exactly what the penalty for a wrong answer will be. They could be still further improved probably by indicating through the use of short carefully supervised and paced 15-second trial periods the change in the relative scores they would earn by adopting different attitudes and techniques with respect to the tasks assigned.

A number of studies were carried out regarding the influence of time of day and order of administration on test scores. The combined results from these studies indicated that time of day was not a significant systematic factor in determining apparatus-test scores. Studies of examiner differences were carried out at various units. The results appeared to be somewhat different at different units, presumably reflecting different supervisory and training policies with respect to the examiners. Although large differences were not found, there was a definite indication that two or three examiners in certain of those units caused the group of students examined by them to get significantly higher or lower average scores than those for the other examiners.

No differences were found in the test scores of people seated in various sections of the room for the printed tests. An interesting experiment compared the performance of a group of individuals using booklets which were so tattered that they were about ready to be salvaged as waste paper with the performance of a group using new test booklets. No differences were found which approached statistical significance even though relatively large groups were compared.

As previously indicated, the administration of apparatus tests presented a number of difficult problems. One of the most important of these appeared to be the problem of standardization of the various apparatuses. From the beginning it was found that mean scores on the various apparatuses differed to a greater extent than could reasonably be attributed to sampling fluctuations. The original solution of this problem was to refer the individual's score on the test to those of a large sample of individuals who had taken the test recently, by using standard scores with a mean of 50 and a standard deviation of 10.

After further experience had been gained and procedures had been developed for calibrating and making routine checks on the apparatus tests, a single standard-score conversion table was prepared for use with all copies of a given apparatus. The standard procedures for apparatus control consisted first of obtaining an average raw score on each test each day. These means were inspected daily and any marked deviation resulted in a maintenance check of the apparatus involved. The second step involved was the preparation of a more accurate control check for each apparatus following the completion of testing of each 100 cases.

If a test-unit mean in raw-score units for 100 cases deviated by more than one-third of a standard deviation from the mean of other copies of the same test a thorough maintenance check was instigated immediately. In every case where the mean for a unit exceeded one standard deviation for 100 cases and remediable apparatus defect was found the unit was removed from the testing line. If after 200 cases had been accumulated the test copy deviated more than two-thirds of a standard deviation or more than one-third of a standard deviation for a total of 300 cases the test unit was removed from the line. When a test unit was removed from a psychomotor line because of such statistical evidence of defect, it was not returned to the line until remediable action had been effected. Any difference below one-third of a standard deviation necessitated an inspection of the deviating copy but not its removal from the testing line, since this small deviation had little effect on the final stanines used for classification.

It is believed that most tests in ordinary use are inadequately objectified in administration. Further research is needed to determine the extent to which more detailed directions and practice exercises and systematic control procedures are essential to the effective use of tests of various types.

## TECHNIQUES FOR DEFINING JOB REQUIREMENTS

### Introduction

A basic prerequisite to effective research and experimentation in the fields of selection, classification, and training is a definition of the requirements for successful participation in the job both in terms of the actual component activities and their relative importance. It is also essential that this definition be translated into abilities, aptitudes, and personality traits.

The traditional approach to this problem has been in terms of various methods known as job description, job classification, and job analysis. Use of these procedures in the Aviation Psychology Program emphasized the need for improved methods of defining job requirements and a considerable amount of effort was devoted to this problem.

Although a detailed verbal description of exactly what an individual does on a particular job supplies useful orientation and background information, it proved of little practical value in most instances. Job descriptions of this type merely provided the basis for making a long list of requirements but did not differentiate among them. Similarly, job classification procedures ordinarily provided descriptive information which gave requirements only in very general terms and provided little assistance in solving the practical problems of selection and training.

Techniques of a more analytical and less purely descriptive nature commonly included under the general heading of job analysis were found of greater use than the descriptive methods. The term job analysis has been used to include both the defining of the activity and the translation of the requirements for success from statements regarding the actual operations performed into statements of the general abilities, aptitudes, and personality traits believed to underlie this behavior.

The only satisfactory basis for studying job requirements is in terms of the actual job to be performed. Too often, studies have been made on training requirements with the assumption that these can be expected to relate closely to the requirements of the job. Experience indicates that this close similarity cannot be assumed but must be demonstrated.

### Judgments of Participants, Supervisors, and Observers

Judgments regarding job requirements are unlikely to be valuable unless based on observations of the work of a fairly substantial number of individuals on typical assignments. It has been commonly supposed that participants, supervisors, and observers of particular activities are automatically qualified as authorities on the requirements for this activity. Experience in the Aviation Psychology Program indicates that this is very definitely not the case.

In many instances *participants* know very little about the critical requirements of a particular activity. They have very little information as to whether the difficulties which they have are typical of those which others have and are apt to dismiss individual differences in proficiency with comments regarding motivation and generalities. They have never found it necessary to make a systematic analysis of requirements for the activity and their judgments are usually in vague terms and likely to represent "stereotypes" which have been suggested to them in conversation.

In general, *supervisors* are a better source of information since it is part of their job to evaluate the work of individuals participating in the activity who are subordinate to them. In the initial survey in the European and Mediterranean theaters it was found that those commanders with superior combat records had considerable insight into the requirements of the jobs performed by individuals under their supervision. As a part of this survey more than 300 of the officers in immediate supervision of air-crew personnel indicated on a check list of 20 aptitudes, abilities, and other personal characteristics, their judgments concerning the relative importance of each of these traits for an individual capable of doing superior work of a specific type in combat operations. These judgments were indicated in the form of the minimum acceptable standards which they believed should be used for each of these traits in selecting and classifying air-crew personnel for the specific types of duty assignments.

By carefully defining the traits and indicating what was meant by them in terms of actual samples of test materials and in terms of related performances, it was believed that a fairly satisfactory reporting of the judgments of these supervisors was obtained. It was also found that personnel filling out this check list under the supervision of different aviation psychologists in different theaters of combat agreed fairly well as to the relative importance, in terms of minimum standards, of the traits listed for the specific duty assignments.

The value of the judgment of an *observer* depends to a considerable extent on his training and insight into individual and trait differences, his technical skill, and the extent to which his observations are systematically made and recorded.

### Systematic Analyses of Good and Poor Performance

The principal objective of job analysis procedures should be the determination of critical requirements. These requirements include those which have been demonstrated to have made the difference between success and failure in carrying out an important part of the job assigned in a significant number of instances. Too often, statements regarding job requirements are merely lists of all the desirable traits of human beings. These are practically no help in selecting, classifying, or training individuals for specific jobs. To obtain valid information regarding the truly critical requirements for success in a specific assignment, procedures were developed in the Aviation Psychology Program for making systematic analyses of causes of good and poor performance. These procedures, together with some of the results obtained from using them, were discussed in chapters 6 and 7 of this report.

Essentially, the procedure was to obtain first-hand reports, or reports from objective records, of satisfactory and unsatisfactory execution of the task assigned. The cooperating individual described a situation in which success or failure was determined by specific reported causes.

This procedure was found very effective in obtaining information from individuals concerning their own errors, from subordinates concerning errors of their superiors, from supervisors with respect to their subordinates, and also from participants with respect to coparticipants.

Theoretically, at least, the analysis of reports and objective records of the performance of the assigned task should provide a useful source of data. However, in practice it was frequently found that such records and reports were so incomplete with respect to essential details as to make their interpretation untrustworthy unless supplemented by reports of first-hand observers.

An improvement on the procedure of collecting a report primarily on the basis of the individual's memory of the incident at some subsequent date was the procedure of using a check list during or immediately following the specific performance of the assigned task.

An example of the use of this procedure was the development of an aerial-gunnery check list by aviation psychologists from Psychological Research Unit No. 1. This check list was filled out by each student at the completion of each air-to-air fixed-gunnery mission. The check list consisted of a report of difficulties encountered in the air-to-air firing mission and included such items as whether he had trouble with slipping or skidding, (a) not at all, (b) occasionally; whether his turn into the target was (a) generally well timed, (b) often too late, (c) often too sharp, (d) often too hurried; and whether there was in regard to deflection, (a) no difficulty in estimating correct lead, (b) some difficulty in judging necessary lead, (c) some difficulty in gradually reducing lead.

There were also a few items on the check list concerning the mechanical functioning of the gun and plane and the behavior of the tow plane. For example, he was asked to check with respect to the tow plane, (a) speed and altitude maintained regularly, (b) speed irregular, (c) altitude irregular, (d) target yaw, (e) length of run varied.

The analysis of the data of this type included (1) the percent of individuals reporting each of the various types of difficulties for all missions; (2) the percent of individuals reporting difficulties on each of the items for the various missions, especially emphasizing comparison of the early and late missions and the various learning curves involved, and (3) the relation between the air-to-air gunnery scores and the responses given on the check list for a particular mission.

It is believed that analyses of skill and learning in relation to the reports of students as well as instructors and supervisors on check lists of this type should be of considerable value in the study of learning difficulties and the analysis of job requirements.

An example of the use of this procedure in a situation which was made much more difficult because of the lack of precise definition of success in the assigned task was provided by analyses of the characteristics of good and poor combat leaders. The objective of the study was to determine the principal requirements for success as a combat leader. It was believed desirable to carry on a systematic study to attempt to improve on the numerous "armchair" concepts of leadership.

It was believed that combat personnel were especially motivated to evaluate the quality of combat leadership they received. Not only the success of their operations, but also their personal survival, depended upon the quality of leadership they received. Under these circumstances it is believed that special value should be attached to a report of an incident which impressed the observer as confirming his belief in an outstanding commander or as confirming his identification of a commander in whom he lacked confidence.

Such incidents were the primary source of the individuals' opinions concerning the actions and characteristics of a successful leader. They provided directly the type of information necessary for the analysis and study of combat leadership. If the individual had been asked to list the activities and characteristics of an outstanding combat leader, it is likely that he would draw not only upon his own experience and observation, but also upon the "stereotypes" which he had been taught to accept as typical of the good combat leader. Although such "stereotypes" cannot be completely avoided or dismissed, it is believed that the procedure outlined above tended to provide a more valid basis for the structuring of the concept of leadership than that supplied by direct reports regarding the traits believed of most importance for the combat leader.

The concept presented in this section of critical requirements for success in specific training tasks is believed to deserve great emphasis in planning and conducting studies in this field in the future.

### Systematic Analyses of Superior and Inferior Participants

The most useful type of job-analysis information is a description of the traits and identifying characteristics of candidates for air-crew training which best differentiate those who are later found to be most successful. In cases where it is possible to identify individuals who have been outstandingly successful on the job and similar groups who have been notably unsuccessful, a careful analysis of the characteristics of these groups of individuals is very desirable.

Experience indicates that in practice this seemingly excellent technique of defining job requirements is of only limited practical value. This limitation is chiefly the result of the impossibility in most situations of obtaining comprehensive descriptive information which had been collected concerning the individual prior to his selection and training for the specific activity. It has been found very dangerous to assume, especially with regard to personality characteristics, that because successful and unsuccessful indivduals differ with respect to a given trait at the present time that any such difference existed at the time they were originally selected for training. The question of causality is of great importance here, and the possibility of error in assuming that the trait caused the failure rather than that the failure caused the trait must be constantly considered.

This is an extremely important and valuable technique when information collected prior to selection for the particular activity is available for analysis. In addition to the requirement of careful interpretation of the results to avoid errors of casuality when the tests or other information about the man were gathered at a time subsequent to his success or failure, great care must be used in insuring that the criterion used in selecting the successful and unsuccessful individuals is really a valid measure of success in the defined activity.

In conclusion, it is important to separate and define two different aspects of the job analysis program. First, there is the need for determining the components of the job which are critical in the sense that they have been demonstrated as making the difference between success and failure in carrying out an important part of the job assigned in a significant number of instances. The identification of these critical requirements leads to the preparation of a criterion measure of success in the activity as described in the next section. The second problem is stating the requirements of the job in terms of testable traits. These requirements, rather than referring to specific activities and types of behavior which are critical, are in terms of the aptitudes, abilities, and traits which are required of the individual himself in order to perform the assigned task in a successful manner. This type of job analysis provides the immediate basis for the development of tests of aptitude, temperament, and ability for use in selection. It also provides the basis for developing various types of intermediate measures of proficiency and knowledge for use in determining training content and methods and in classifying individuals for advanced assignments.

## PROCEDURES FOR DEVELOPING A CRITERION MEASURE OF THE EFFECTIVENESS OF BEHAVIOR

### Introduction

As indicated in the previous section, the definition of the requirements of the job is a prerequisite to effective research and experimentation. After these requirements have been defined in terms of an analysis of behavior on the job, two problems remain. The first, which was discussed in preceding sections is to translate these requirements into fundamental psychological and descriptive categories which can be used in developing procedures for predicting the relative effectiveness of various individuals prior to selection and training. The second is the problem of obtaining and combining measures of effectiveness or proficiency which have been found to be relevant to the assigned tasks to obtain a criterion of success. Techniques for accomplishing this will be the subject of this section.

There is no way in which criteria of success can be established on an empirical basis. The definition of success in any activity must always be on rational considerations. Furthermore, it is impossible to avoid making judgments regarding the relative importance of various aspects of the assigned task. In some instances, experts can quickly come to an agreement on the relative importance of various aspects such as speed and accuracy in performing a simple routine task. On the other hand, it may be very difficult to obtain agreement on the relative importance of saving lives and getting bombs on a specified target. The criterion of success must depend on the precise definition of the assigned tasks for the particular job.

The discussion which follows will be concerned with three principal topics. First, the various types of information which might be obtained for use as criteria of success in a particular activity; second, the principal factors to be considered in selecting the most suitable measures; and third, procedures for combining various measures to obtain a single criterion of success.

### Types of Measures

The types of measures most useful as criteria of success are those which are based on controlled observation of numerous typical samples of the individual's behavior in the activity. The observations should be made with as much objectivity in appraisal procedures and control of conditions affecting the nature and difficulty of the task as possible. The experience in the Aviation Psychology Program with studies of criterion measures of success in air-crew assignments has broad implications which will be discussed in the paragraphs which follow.

The final evaluation of all military personnel selection and classification procedures, training techniques, and weapons must be based on demonstrated effectiveness in actual combat. There is definite evidence that in selecting persons of the type who would excel in certain types of training activities, personnel were obtained who were actually inferior to the group rejected as judged by later success in more advanced types of training and in combat. Similarly, persons trained on certain training devices and by certain techniques were found later to be less effective than those who had had no training at all of this type.

Certain types of test materials, for example, questions on history and academic vocabulary, used to select persons for flying training in the AAF in the early days of the national emergency were found to predict preflight success but to be negatively related to success in flying training. Similarly, some gunnery trainers have been reported to teach people bad habits and interfere with their later work on combat-type equipment. Tests of ability to identify targets were found to be of little value in predicting the bombing accuracy of personnel in bombardier training schools. However, when the bombardiers transferred to targets in enemy territory which were not clearly identified by large white concentric circles this ability was found to be a most critical one. The ability to remain cool and collected in the face of enemy fire is a more obvious although much less predictable difference between the training and combat situations.

Because of the predominant importance of combat effectiveness it might be expected that combat criteria would be used exclusively in military research. That this is not the case is primarily a function of the numerous problems and difficulties encountered in attempting to make this type of study.

The chief problem in connection with such studies is finding a suitable criterion measure of combat effectiveness. For discussion these may be classified into four main categories: (1) objective measures of combat proficiency, (2) records of administrative actions in combat units, (3) ratings based on direct and systematic observation of combat effectiveness, (4) ratings based on general impressions, reports, and incidental observations.

The most readily accepted criteria of combat performance are those provided by objective measures. Among those being used by the AAF are: (1) records of bombing accuracy as obtained from strike photographs, (2) records of airplane accidents related to combat operations, (3) records of numbers of enemy planes shot down, and (4) records of combat casualties.

Another type of record which has been widely used because of its availability in spite of its substantial dependence on the subjective personal judgments of superior officers consists of the records of administrative actions in combat units. Such actions include: (1) promotions, (2) special awards and decorations, (3) reclassifications due to anxiety reactions or lack of combat proficiency, (4) special duty assignments, such as to lead crews or for important combat missions, (5) official ratings of efficiency while assigned to a combat unit.

The third type of criterion data which has been used consists of ratings based on direct and systematic observation of combat effectiveness. Such evaluations include (1) ratings of navigational proficiency based on a study of the logs prepared by the navigator on various missions, (2) ratings regarding the success or failure of bombing missions based on direct observation of the results or studies of records and reports, and (3) systematic ratings of combat performance of pilots based on directed observation of behavior on a specific fighter or bomber mission. The value of this type of data is increased by careful definition in advance of the specific aspects of behavior to be observed and rated.

The final type of criterion data studied consists of ratings based on general impressions, reports, and incidental observations. Such ratings appear to be the least valuable of the types of criteria studied in the AAF. The greatest weakness of such ratings is that they are seriously affected by extraneous biases and the well-known "halo effect." Such ratings in the AAF have been found to be spuriously related to the length of time the individual being rated has been known by the superior officer doing the rating, his liking for the individual on a personal basis, and the prestige and general reputation of the person rated.

Studies have also shown that ratings of specific abilities and traits are seriously biased because the rater, in his observation of the individual's performance, has been chiefly concerned with collecting evidence for some other purpose. The rater's evaluation of the individual is primarily in terms of some specific decision which he knows he is going to have to make. For example, the commander of a combat unit might be considering a number of the officers under him for a specific staff position which he expected to have to fill soon. If at a later time the rater is asked to make a rating on some other aspect of the individual or a decision which concerns some other ability of the individual, his estimate will be biased by the specific evaluation which he has already made. As previously suggested, one of the best procedures for avoiding this "halo" effect is to focus the rater's attention on the specific ability or trait on which it is desired that he render a report.

It is possible in some cases for certain raters to observe and evaluate simultaneously the performance of individuals in a number of abilities and traits. However, due to the limitations of the analytical abilities and memories of most raters, such a procedure frequently leads to a blurring of the various specific ratings. The value of ratings can be greatly improved as indicated above if they are based on systematic observations which are directed toward the final evaluations to be included in the final rating.

Although measures of success in anything short of the final activity for which the individuals have been selected and trained cannot be regarded as adequate or acceptable criteria, in practice these partial and intermediate criteria have to be used to a very large extent. In the ordinary situation it is necessary to establish the relevance of these criteria on the basis of a rational analysis. The most desirable types of proficiency measures of this general kind are those obtained from simulated combat missions or other simulated performances of the final assigned task. The objective flight checks and objective check lists discussed in chapter 5 are intermediate criteria which have considerable relevance in most situations.

## The Selection of Measures

In selecting measures to be used as criteria of effectiveness the principal considerations are their relevance, reliability, and freedom from bias. Although as mentioned in the previous section, the problem of relevance ordinarily must be determined on the basis of rational analyses, such analyses must be made with great care. They should also be supplemented to the maximum possible extent with empirical studies of the relation between the partial and intermediate criteria and the ultimate criterion measure. Even if the very difficult problem of a satisfactory criterion has been solved, there still remain a number of very complex technical problems which must be dealt with before an adequate interpretation of the findings can be obtained. Any such interpretation must include a careful consideration of disturbing factors of two principal types: (1) attenuating factors, and (2) biasing factors.

*Attenuating factors* tend to obscure existing relationships by reducing the relative amount of variation in combat performance associated with the predictive measures being used as compared with the variation in results produced by factors over which the individuals studied have no control. Such factors include: (1) lack of consistency or reliability of both types of measures being studied; (2) the introduction into either or both variables of heterogeneous factors which are unrelated to the variable with which the relationship is being studied. This type of heterogeneity may be due to factors associated with the test scores or predictors but the larger portion of such disturbing factors is usually found in the criterion and is composed of differences in opportunity to achieve "success" as defined by the criterion or differences in the standards used by the various superior officers in rating or judging the individuals under their supervision; (3) the restriction of the variability of the group studied with respect to the predictive measures used, commonly referred to as restriction of range. The repeated selection of those believed superior in initial selection, classification, training, and finally in the initial phases of combat itself restricts the range of individual differences in specific types of combat personnel.

*Biasing factors* introduce a spurious relation into the study and may result in a false finding of either a positive or negative relationship when in fact no relationship at all or a small relationship of opposite sign exists. For example, ratings of men in combat are known to include a factor consisting of length of time the individual has been known to the rating officer. Persons who get their training early in the war tend, therefore, to get higher ratings. On the other hand, in the AAF, men trained at later periods were more rigorously selected and better trained. Thus, a negative correlational factor is introduced by the heterogeneous factors in this case and the resulting coefficients are seriously biased unless this factor is controlled.

Such considerations indicate the importance of careful controls which will eliminate any biasing either positively or negatively of the results obtained. Because of the many extraneous factors involved in the typical combat validation situation, any correlation coefficients obtained which greatly exceed the general range of 0.05 to 0.15 would probably become greater than the values typically obtained in predicting success in training when corrected for all of the appropriate attenuating factors. The implication is that very large samples must be used if positive findings are to be expected in studies of combat criteria.

## The Combination of Measures

Since the definition of a particular job is usually not sufficiently precise to provide information concerning the relative importance of various aspects of the assigned task, the investigator must ordinarily make judgments regarding these matters. It is especially important that in reporting findings or comparisons with the final criterion used the precise definition of the activity on which the criterion of success has been based be made as explicit as possible.

In operational training a number of different criteria were available for individuals in various specialties. For example, for fighter pilots there were scores on air-to-air gunnery, air-to-ground gunnery, accidents, reclassifications, and various types of grades and ratings. Similarly, for combat such criteria as promotions, decorations, reported victories in air combat, casualties, and reclassifications were available. All of these types of criterion information appear to have some degree of relevance for judging the success of a particular individual.

If it is desired to avoid using the judgment of only a single individual in determining the relative weight of the various criteria for obtaining the composite, it is necessary to adopt some procedure for selecting appropriate judges and obtaining and combining their estimates of the relative importance of the various criteria in a systematic fashion.

## FACTORS RELATED TO MATHEMATICAL METHODS OF PREDICTION

### Introduction

Having established criteria or a criterion of success in a particular activity and having also developed a battery of tests and other predictive devices, a number of problems arise relating to the mathematical methods to be used in predicting from the information available which individuals are most likely to be successful. The information desired is the likelihood of attaining a specified degree of success in the activity given any one of the possible specific patterns of scores and other characteristics such as age, education, marital status, and so forth. In the typical situation it is not practical to compute this probability for each of the possible combinations of scores and other data. Therefore a technique is needed for placing people into a relatively small number of groups in which the chances for success are quite homogeneous within the group and vary as greatly as possible between groups.

In the process of determining the best procedure to use in a specific situation numerous decisions must be made regarding the most plausible assumptions to use in grouping people with various patterns of scores and characteristics. In any practical situation, many simplifications, abstractions, and approximations are necessary. The discussion which follows provides some evidence which may be useful in handling problems of this type in other situations.

### Obtaining the Predictive Value of Specific Measures

The most useful measures of the predictive value of specific tests were found to be product-moment correlation coefficients calculated from continuous distributions of scores obtained from a group including all candidates tested during a given period. When a sample has been restricted with respect to the range of variation of the scores because of administrative procedures the obtained values should be corrected for this restriction. When the criterion was limited to two categories such as pass and fail, biserial coefficients of correlation were generally found to be most useful. It is frequently important to correct the coefficients obtained for the attenuating effects of extraneous factors, especially in the criterion measures.

The general procedures used for the correction of correlation coefficients for the reduction in them due to the restriction of the range of scores in the population from which they were calculated followed the work of Karl Pearson originally published in 1903. Unfortunately, the formulas developed by Pearson did not apply to a number of the situations in which it was found desirable to correct coefficients for the effect of restriction of range in the Aviation Psychology Program. A substantial amount of work was carried on in the Aviation Psychology Program concerning both the theoretical problems involved in deriving appropriate formulas for these special situations and in carrying out empirical checks regarding the adequacy of the various formulas which were available. The Pearson formulas did not apply to the situation in which one of the coefficients was a biserial correlation. A procedure for this special case was developed by Gillman and Goode.[3]

Unfortunately, the G-coefficient developed by these individuals did not appear to give stable values under the circumstances which were in effect in the prediction problems of the Aviation Psychology Program. Detailed discussions of the various problems of correcting for restriction of range are presented in the reports in this series on Research Problems and Techniques, and Records, Analysis and Test Procedures.

At various times the question was raised as to whether the use of biserial correlation coefficients or point biserial correlation coefficients was preferable. The point biserial is the product-moment coefficient obtained by using the two categories such as zero and one for the dichotomized variable. The biserial correlation coefficient includes a correction factor to make it equivalent to the product-moment correlation coefficient which would be obtained if the dichotomized variable were distributed normally in continuous form. The most important advantage of the biserial correlation coefficient in the situation in the Aviation Psychology Program was that it was unaffected by changes in the proportion of individuals eliminated from training. This made it much more meaningful to compare correlations obtained from different classes. In practice it made no difference which of the two types of values were used in obtaining the multiple regression coefficients for a single specific sample, since the correction factor is a

[3]Gillman, L., and Goode, H. H. An Estimate of the Correlation Coefficient of a Bivariate Normal Population when $X$ is Truncated and $Y$ is Dichotomized. *Harvard Educ. Rev.*, 1946, 16, 52–55. (The authors developed this formula while working on a project for the Office of Research and Inventions of the Navy Department. This study was undertaken at the suggestion of Dr. Phillip J. Rulon who was acquainted with the problem through his work as a Consultant to the Aviation Psychology Program.)

constant value in any one sample of this type and therefore has no effect on the relative size of the regression equation coefficients.

It should be noted that if the relative regression-equation weights for tests or other variables are desired for use in obtaining a combined score and the data for one of the variables to be used in obtaining the combined scores consists of two categories, the correlation coefficients for this variable should be obtained using categories of zero and one for this variable rather than coefficients correcting for this grouping in the independent or predicting variable. The tetrachoric type of coefficient would only be appropriate if at the time the combined scores are calculated continuous scores for this particular variable are expected to be available instead of the two-category scores.

A frequent problem in the validation of test scores and other predictive measures is correcting for heterogeneity of rating standards, opportunity, or other similar factors which tend to attenuate the correlation with the criterion measure. In other instances it is unnecessary to make an exact correction but highly desirable to have a rough estimate of its probable magnitude. For example, discussions of the validation of procedures against combat criteria are full of statements that it is useless to use supervisory officers' ratings as combat criteria because of the large differences in the standards of the rating officers. Similarly, objective criteria such as number of planes shot down, number of accidents, and completion of a specified number of missions are also frequently stated to be valueless as criteria of combat skill because of the large variations in opportunity.

It therefore seems desirable to have formulas readily available to obtain either a rough or a precise estimate of the effect of such factors on the correlation coefficients obtained.

The general formula for the covariance between a test score composed of test ability ($T$) plus a variable factor ($A$) due to differences in methods of administering, scoring, or converting scores at different testing units, and a criterion score composed of proficiency ($P$) plus an extraneous factor ($O$) due to differences in standards or opportunity to achieve "success" as defined by the criterion is written as follows:

$$\sigma_{(T+A)}\,\sigma_{(P+O)}\,r_{(T+A)(P+O)} = \sigma_T\,\sigma_P\,r_{TP} + \sigma_T\,\sigma_O\,r_{TO} + \sigma_A\,\sigma_P\,r_{AP} + \sigma_A\,\sigma_O\,r_{AO}. \quad (1)$$

Solving for $r_{TP}$, which is the correlation coefficient desired, provides:

$$r_{TP} = \frac{\sigma_{(T+A)}\,\sigma_{(P+O)}\,r_{(T+A)(P+O)}}{\sigma_T\,\sigma_P} - \frac{\sigma_O r_{TO}}{\sigma_P} - \frac{\sigma_A r_{AP}}{\sigma_T} - \frac{\sigma_A \sigma_O r_{AO}}{\sigma_T\,\sigma_P} \quad (2)$$

In many cases this formula simplifies considerably. For example, if the two extraneous factors of test administration and scoring conditions ($A$) and opportunity or local rating standards ($O$) are uncorrelated, the last term drops out. In a similar fashion the two previous terms may vanish if the factors ($A$) disturbing the test score and the factors ($O$) interfering with the accuracy of the measures of proficiency are unrelated to true proficiency and test ability respectively.

The formula thus reduces in case the disturbing factors in each of the scores are not correlated with the combined score in the other variable to the following:

$$r_{TP} = \frac{\sigma_{(T+A)}\,\sigma_{(P+O)}\,r_{(T+A)(P+O)}}{\sigma_T\,\sigma_P} \qquad (3)$$

and if each of the disturbing factors can also be reasonably assumed to be uncorrelated with the other score with which it is combined this can be written,

$$r_{TP} = \left(1 + \frac{\sigma_A^2}{\sigma_T^2}\right)^{1/2}\left(1 + \frac{\sigma_O^2}{\sigma_P^2}\right)^{1/2} r_{(T+A)(P+O)}. \qquad (4)$$

If the testing is all done at one unit or if conditions are very similar from unit to unit so that the test scores are homogeneous the formula becomes,

$$r_{TP} = \frac{\sigma_{(P+O)}}{\sigma_P}\,r_{T(P+O)} - \frac{\sigma_O r_{TO}}{\sigma_P}, \qquad (5)$$

and if opportunity or the standards of the raters, etc., can reasonably be assumed to be uncorrelated with test score ($T$), this simplifies to,

$$r_{TP} = \frac{\sigma_{(P+O)}}{\sigma_P}\,r_{T(P+O)}, \qquad (6)$$

and if opportunity, etc. ($O$), can also be reasonably assumed to be uncorrelated with proficiency ($P$), this may be written,

$$r_{TP} = \left(1 + \frac{\sigma_O^2}{\sigma_P^2}\right)^{1/2} r_{T(P+O)}. \qquad (7)$$

These formulas have certain important implications for prediction studies. For example, it is important to note that the correlation due to the type of extraneous factors indicated by the terms preceded by minus signs in formulas (2) and (5) may either result in an increase or a decrease of the uncorrelated coefficient as it is originally obtained. Thus the obtained correlation may actually be larger than the correct value.

On the other hand, when the extraneous factors are uncorrelated with the real variables and formulas (3), (4), (6), and (7) therefore apply, the extraneous factors can only serve to attenuate the obtained coefficients.

The extent of this attenuation can readily be estimated in planning a research study if some information is available concerning the proportion of the variance attributable to the disturbing factors. For example, if it is known that opportunity ($O$) results in about as much variation as does proficiency ($P$), it can be anticipated that the obtained correlation coefficient will be about 30 percent smaller than it would be if opportunity ($O$) were constant for all. Thus, in many practical situations disturbing factors even though fairly substantial will not seriously interfere with obtaining useful results from follow-up studies.

On the other hand, if the disturbing factors for the two basic variables are correlated with each other or with the basic variables, large variations, even including change of sign of the obtained correlation, may result. Thus a difference in the method of administering tests or handling the scores may introduce a disturbing factor in the test scores obtained in one region which, even though small, may significantly affect the obtained correlation coefficients if the men with scores so affected all go into training schools or operational units where the average opportunity differs somewhat from that of the men tested in other regions.

A word on obtaining estimates of these variances is appropriate. In cases in which the disturbing factor is associated with the particular station, the variance of values within stations and the total variance for all stations or the variance of the means of the stations can be calculated for the samples available. The use of these values involves the assumption that the means in all stations would be identical were it not for the disturbing factors. In certain other cases similar analyses could be made by instructor or by flight, squadron, group, air force, or theater. In general, it is necessary to find groups in which the disturbing factor is constant or very nearly constant within each group. Although groups which satisfy these conditions precisely are frequently not available, a reasonable approximation can usually be obtained.

As has been discussed previously in this report, small correlation coefficients can be of very great importance in situations in which there is a large amount of heterogeneity in the data as discussed in the preceding paragraphs. In a large number of instances in the Aviation Psychology Program it was necessary to work with a criterion measure having an extremely low reliability coefficient. It should be emphasized that stable values can be obtained with these small coefficients provided the number of cases is increased to the point where the standard errors are small in relation to the sizes of the coefficient. It is important to remember that these small coefficients are indicative of very large relationships sometimes between "true" measures for the two variables involved. This can best be seen by correcting such coefficients by the use of such formulas as those just presented.

### Predictions Based on Several Measures

Multiple-correlation procedures were found to be best for obtaining estimates of the maximum predictive value of several measures and for determining the weights which should be applied to predict a specified criterion with maximum accuracy in the least-squares sense. A number of procedures were investigated in an effort to improve on the accuracy of the predictions which were obtained by calculating correlation coefficients and using the regression-equation coefficients obtained from these as weights for combining the various test scores to predict the given criterion.

703323—47—20

One of the first problems investigated was the question of curvilinear relations. It was thought probable that very low scores on certain tests might be predictive of poor performance in training schools, but that scores above a certain point, which represented the desirable minimum of ability in that area, would not be discriminating with respect to the criterion of success used.

Therefore, a number of empirical studies were made to check on the linearity of the relation between the test scores and the criterion measures. It was found that in certain instances samples as large as 3,000 cases would indicate a useful degree of curvilinearity in the data. However, this curvilinearity was not found to be confirmed for a second sample of 3,000 cases obtained in the same way. It therefore appears very doubtful that any attempt to utilize curvilinear relationships should be attempted unless the calculations are based on at least 10,000 cases.

The procedure which it was planned to use in cases where significant curvilinear relationships were found on large samples consisted of changing the curvilinear relationship to a linear relationship by modifying the size of units at the appropriate places on the distribution of scores. This procedure would only be satisfactory in cases in which a test score was weighted for only one criterion score, or in which the relation between the test score and all criteria had the same degree of curvilinearity.

Another problem which was extensively investigated, especially with regard to the apparatus tests, was the problem of obtaining the best weighted combination of scores on various trials included in a particular total test. To a first approximation, the multiple correlation between the trial scores and the criterion indicates the degree of improvement which could be made in the predictive value of the test score by the addition of information regarding the shape of the learning curve of the individual. Ordinarily, the total score on these tests is merely the average level of the individual's points on the learning curve. It was rather surprising to most of the aviation psychologists to find that multiple correlations of this type gave no significant improvement in the predictive value as obtained from the total score on the test.

A procedure frequently used by certain groups in prediction problems is to establish a series of cut-off scores. This procedure is usually referred to as the multiple-cut-off procedure. In order to qualify or be selected for a particular type of duty the individual had to score above the cut-off value on all of the tests which were relevant to this assignment. Analyses of data available in the Aviation Psychology Program indicated that the multiple-cut-off procedure was not only less efficient than the regression-equation technique but also quite impractical when the battery of tests contained as many as 15 or 20 scores. A detailed discussion of this problem will be found in the report in this series on Research Problems and Techniques.

In many situations a clinical method has been used in obtaining predictions from a number of test scores and related variables. The proponents of this technique argue especially that the pattern of scores as well as the average size of them is important.

A number of conditions are necessary for the use of the clinical method. In order that a sound clinical judgment may be made with respect to whether or not an individual should be trained for duty as a combat pilot, for example, it is necessary that the clinician have two types of information. First, he must have information concerning the personality and characteristics essential to success as a combat pilot. Second, he must be able to obtain accurate information concerning an applicant's personality and characteristics expressed in the same terms as is the first type of information.

Several studies were made in the Aviation Psychology Program of the effectiveness with which various types of individuals could select personnel for specific assignments on the basis of an interview and a clinical evaluation of the test scores and other information available. One of the first studies of this type carried out in the Aviation Psychology Program was an evaluation of the effectiveness with which commanding officers at AAF Flexible-Gunnery Schools selected candidates for a special course in D-8 low-altitude bombardier training from among members of the graduating classes of their schools. This was a special training project which lasted only about 3 months and there was, therefore, not time to make an empirical study of the best weights and best tests to use for the selection of these individuals before the procedures were put into effect. On the basis of a careful analysis of the requirements of the job, a battery of tests was selected and weights were assigned.

The final selection was accomplished by having the commanding officers at the gunnery schools interview the individuals in the order of their rank on the combined aptitude measure developed by aviation psychologists. The commanding officers had all the information on the tests, grades in the school, and other information about the individual at the time of their interview with him for the purpose of making the final selection of men to be sent to the D-8 School.

Although only a fairly small fraction of the graduating class were needed to fill the quota in any one school, some of the commanding officers rejected so many people on the basis of their evaluation in this clinical situation that they went down as low as the bottom quarter of the class to meet their quota. In other schools all of the individuals sent came out of the top quarter. Using the available criterion measures at the D-8 (low-altitude) Bombardier School, the effectiveness of the selections of the commanding officers were evaluated in comparison with the effectiveness of the combined aptitude test scores. The results indicated very clearly that in deviating from the order given by the combined aptitude scores, the commanding officers tended to get an inferior rather than a superior group of individuals for the D-8 Bombardier School.

Another study of this type was carried on in connection with the classification of applicants for air-crew training at the AAF Classification Centers. In each of the three centers a board was established by directive of Headquarters AAF Training Command to interview all applicants disqualified for pilot training because of low aptitude (stanines of one or two) and attempt to identify a few individuals who, in spite of low scores, appeared to be promising pilot material.

At the Nashville center the board was composed of an aviation psychologist and a flight surgeon. The board at San Antonio was composed of administrative officers and rated personnel featuring especially a young pilot, bombardier, and navigator, all three of whom had recently graduated from training. At Santa Ana the board was dominated by personnel officers, some of whom had had experience in industrial personnel work.

The results of analyzing the success in pilot training of those individuals selected by these boards as being especially promising pilot material indicate that in no case did the judgment of the board appear to have appreciable validity. Of the 420 cases selected, slightly less than one-third graduated from elementary flying training. There does not seem to be any clear-cut superiority of any one of the types of boards over another. All boards granted fewer and fewer waivers the more experience they had. In no case did the very carefully selected groups of low-stanine men do anywhere nearly as well as the average cadet, nor did they even do as well as those in the next higher stanine group, the threes.

One of the findings of the Aviation Psychology Program, of the greatest practical importance to research in the field of prediction of individual adjustment, was the sensitivity of regression coefficients to sampling errors and the instability of these coefficients from sample to sample. It was also discovered that the regression coefficients could be varied over a rather substantial range of values without appreciable loss in the multiple correlation obtained from the weighted combination. Under these circumstances it was found desirable to use logical considerations to modify the weights from those obtained in any except the very largest samples including several thousand cases obtained using criteria of unimpeachable validity.

A finding incidental to this was the spurious nature of most negative regression coefficients in practical situations dealing with ability and aptitude tests. Numerous efforts were made to develop and use valid and stable suppression tests. The function of such suppression tests would be to cancel out the effect of an irrelevant portion of a test in which valid elements were also present, thus gaining a purer measure of the valid function. For example, it was thought that the removal of the vocabulary element from various tests involving the knowledge and use of words but designed mainly to test other functions, might increase their validity since it was known that vocabulary had practically a zero correlation with criteria of success in pilot training.

Similarly, the removal of that aspect of the Instrument Comprehension

Test, Part II, which was related to Part I of the first form of this test which consisted of verbal descriptions of indications from instruments was proposed. This first part had a high correlation with the second part of the test. Moreover, on the basis of validity coefficients from more than 1,000 cases it appeared to have a much lower correlation with success in primary training than the second part. It was believed that subtracting a value proportional to the individual's score on the first part would provide a purer measure of the valid function in the second part of the Instrument Comprehension Test and therefore give an increased combined validity. This first part was actually introduced into the battery for a time as a suppression test, but analyses of subsequent samples indicated that the correlations between the two parts were somewhat lower than originally found, and the validity for the first part somewhat higher than the value obtained on the earlier sample. With these changes it was found that the value of Part I as a suppression test was entirely lost. It was therefore removed from the battery.

It appears clear from these results that any negative regression coefficients which are not based on samples including several thousand cases should be looked on with a great deal of suspicion.

## TECHNIQUES OF EXPERIMENTATION

### Introduction

Experimentation is defined as the process of confirming or disproving something which is doubtful. The first step in experimentation is the formulation of a clear and precise statement of the item or hypothesis to be tested. Experience indicates that more experimental work fails to achieve useful results because of failure to formulate the point to be tested in simple and concise terms than for any other reason. Techniques which have been found to be of assistance in insuring satisfactory completion of this step will be discussed and illustrated.

The second step is the design of the experiment. This also is a crucial matter and requires considerable training and experience. Factors found to be important in designing typical experiments in selection, training, and operations will be reported and examples given. The planning of the experiment is closely related to problems of carrying it out. The actual conduct of experiments in an operating organization involves numerous special arrangements and adjustments. Procedures which were found useful in incorporating experiments into normal selection, training, and operating routines are described in some detail.

The analysis and interpretation of results and the development of the implications of the findings for policy and procedure are discussed and examples of desirable and undesirable results cited.

### Formulation of Hypotheses

The formulation of a clear and precise statement of the hypothesis to be tested should be the first step in any planning for experimental work.

Frequently plans for studies stated in loose terms sound very significant and important. However, in many cases a clear definition of the nature of the experimental treatment planned and the method to be used in evaluating the results would make it obvious that only one result could be anticipated. In some cases it is probable that a clear statement of the problem would not have been made if the experiment had not been performed, but this seems to be a rather unsatisfactory reason for an experiment.

One of the points brought out by experience in formulating hypotheses in the Aviation Psychology Program was the desirability of formulating hypotheses in as general terms as possible. The hypothesis that one specific dial face is superior to another specific dial face or the hypothesis that there is no difference between them can be tested and information gained regarding this specific point. However, the usual situation is that information is desired which will make it possible to develop general principles about the design of dial faces. It is therefore extremely important that the hypothesis be formulated in those terms and that the dial faces selected be as suitable as possible for testing the broad hypothesis, the validity of which is in doubt.

In many situations, psychologists are inclined to set up hypotheses and then fail to put them to experimental test. Experience has indicated that it is necessary to verify even the obvious. An example of the importance of this point is provided by the experience in testing the comparative validity of Complex Coordination Tests, Models A, B, C, D, and E. In the development of this series of forms the logical analysis was so convincing that many psychologists believed that the conclusions which were in reality hypotheses needed no experimental testing. A detailed discussion of this experience will be found in Research Report No. 4, Apparatus Tests.

## Design of Experiments

The principal problem in experimental design is to plan the experiment in such a way that a crucial test of the hypothesis will be provided. This includes the control of extraneous factors, the selection of a group of suitable size and nature, and plans for the accurate measurement of results and for the collection and analysis of data.

There is no easy method of experimental design. It is essential that the design of the experiment be based on as full and complete knowledge of the conditions under which the experiment is to be performed as it is possible to obtain. The designer needs not only to have considerable technical skill in establishing the conditions of the experiment and the procedures for collecting and analyzing the data but needs to have sufficiently extensive experience in working with the problems being tested that his judgment with respect to the many matters of planning and detail will be sound.

An example of the need for substantial experience and knowledge in order to design a sound experiment is provided by a consideration of some of the factors involved in the study of the relation between aptitude test scores and the incidence of pilot reevaluation in combat. In studying the scores made by the individuals involved it was necessary to consider the operation of a very large variety of possible disturbing factors and to control these to as great an extent as possible. Among the factors which were taken into consideration were:

a. Differences in the quality and characteristics of applicants for flying training at various stations and during various periods of time.

b. Differences in testing conditions and conversion scales among the various units and for various dates.

c. Differences in policies and standards with respect to classification and qualification for training at various units and various times.

d. Differences in the character and quality of training at various flying training schools and at various times.

e. Differences in the character and level of the training standards required for graduation which were in operation at various training schools and at various times.

f. Differences in the character and quantity of flying experience after completing the training course in the AAF Training Command but before entering operational training.

g. Differences in the character and quality of training received in various units in the Continental Air Forces.

h. Differences in policies and standards for the completion of operational training in the selection of men to be sent for overseas duties.

i. Differences in the type of planes flown in combat operations.

j. Differences in the type of combat operations engaged in.

k. Differences in the quality of leadership in various combat units.

l. Differences in the quality of assistance from other members of a combat unit.

m. Differences in the character and quality of enemy opposition.

n. Differences in combat activity and in leave policies and length of combat tour.

o. Differences in the quality of airfields, living arrangements, and other operating conditions for various units.

It is obvious that in a practical situation of this type it is not possible to control all possible sources of variation. The experimenter must have sufficient familiarity with the situation to make a sound judgment concerning the relative importance of various factors and control those believed to be of most importance.

The experimenter must also have sufficient technical skill to design the experiment in such a way that efficient statistical procedures for testing the hypothesis may be used. The plan should include the use of a sample of sufficient size to provide a crucial test but not so large as to be wasteful of effort.

### Interpretation of Experimental Findings

The interpretation of experimental results requires full knowledge of the precise conditions of the experiment and of procedures used in collecting and analyzing the data. The process of scientific experimentation is not a mechanical and automatic one which consists of the performance of certain operations which lead simply and logically to an inevitable conclusion. The scientist must always use judgment based on experience in interpreting empirical data. Unless he has full knowledge of the details of the experiment he is not in a position to make an interpretation of the results. This suggests that scientific experimentation must be carefully supervised by competent personnel with scientific training if satisfactory results are to be achieved.

A common fallacy in interpreting experimental findings is to assume that failure to eliminate one hypothesis in a particular situation means that this hypothesis must be accepted as being true. This is, of course, not true and recent developments in procedures for testing hypotheses tend to emphasize the design in such a way that both the hypothesis and the alternate hypothesis can be tested simultaneously. An example of this type is provided by a series of studies in which the validities for the first half of the series of trials on the apparatus tests were compared with the validities for the whole series. Consistent small differences were found favoring the whole test over the first half of the test. However, with samples of the size used it was reported that these differences were not statistically significant. In other words, the hypothesis that the half test and the whole test were of equal validity was not rejected. On the basis of this finding it was concluded that the half test was as good as the whole test. In these experiments, if the hypothesis set up to be tested had been that the whole test was 0.05 better than the half test, this hypothesis also would not have been rejected by the experimental findings. This indicates the fallacy in the conclusions as reported.

In conclusion, it may be stated that scientific experimentation is a powerful tool, but it is also a delicate instrument which must be handled with great skill and judgment if useful results are to be obtained.

## PROCEDURES FOR COLLECTING, RECORDING, AND ANALYZING STATISTICAL DATA

### Introduction

In the process of giving the air-crew classification tests to more than 600,000 men and following them through training, the personnel of the program gained much experience in handling data. In addition to this systematic and routine processing a wide variety of special studies were made. Much of the experience gained is of a type which is difficult to record and impart, such as a desirable skeptical attitude concerning the accuracy of data reported and a passion for independent checks on all steps and particularly for verification from new samples. However, there are some general principles which have emerged which may be of value to other workers and these are reviewed briefly in the paragraphs which follow.

### Control Procedures

The use of carefully planned control procedures to provide systematic checks regarding various aspects of the collection and handling of data was found essential to adequate supervision. The use of control procedures with respect to the introduction of error variance into the individual's test scores by factors related to the procedures for administering the test or to the conditions of testing or of the test itself were discussed in a previous section. It was also found that the processes of scoring, recording, reporting, and tabulating the test results had to be carefully controlled if errors were to be kept to a relatively satisfactory minimum.

In spite of the use of thorough checking procedures, errors did creep into these procedures. The psychological examining units used comprehensive auditing procedures as described in the Standing Operating Procedures. These procedures attempted to provide a completely independent check of all steps in scoring, recording, and reporting for a small sample of randomly selected individuals.

Similarly in connection with the work of punching and the various operations of collating and tabulating the machine-records cards, a system of error control was developed. This consisted of a comprehensive series of check steps and preventive measures intended to reduce errors to a minimum. Details on both of these control procedures are contained in the Research Report entitled *Records, Analysis and Test Procedures.*[8]

### Form of Basic Data

Objectivity, simplicity, uniformity, parsimony, and ease of duplication were found to be the most important considerations in determining the form of basic data which would provide maximum utility.

The first test records obtained from the individuals were the marks made on the machine-scorable answer sheets. These answer sheets were scored and the values converted into single-digit normalized scores for use in obtaining the weighted combination of aptitude scores or stanines and for permanent record. The weighting was accomplished by means of the aggregate weighting unit of the IBM Test Scoring machine. This machine made it possible to multiply each of the test scores by a weight varying from 0 to 20 and sum the results in only a few seconds.

The use of single-digit scores for permanent records was found to produce a very important saving in reporting, punching, tabulating, and statistical analysis. The scores were entered on as few different records as possible. For example, all of the apparatus-test scores were entered on one form

[8]Deemer, W. L., ed. *Records, Analysis, and Test Procedures.* AAF Aviation Psychology Program Research Reports, No. 18. Washington: Government Printing Office, 1947.

and this form was carried by the examiner from one test room to another. Each individual was assigned a testing number on the day he reported for group testing. This number was of great importance since it enabled the positive identification of the individual and also the place and date on which he took the tests.

As previously mentioned, all test scores were punched on cards and made a part of a permanent file. After the results from the training schools had been collated with this information, rosters of scores were prepared using the IBM printing tabulator. To make these scores available to the many psychological research groups interested in using them, these rosters were printed on microfilm and a number of copies distributed to various field units.

The introduction of standard conversion tables for use with all printed and apparatus tests in obtaining the single-digit scores which had a mean of 5 and a standard deviation of 2 for the large population of cadets used for standardization purposes provided simple, uniform, easily interpreted scores which greatly facilitated subsequent analyses of such records.

### Standard Operating Procedures

The use of a detailed Standard Operating Procedure for all routines regarding the collecting, recording, and analyzing of statistical data was found to be of great value and practically indispensable for efficient operations. This Standing Operating Procedure was discussed briefly in a previous section. A detailed discussion of these procedures is given in the report on *Records, Analysis, and Test Procedures.*

### Operating Principles

All steps in the collecting, recording, and analyzing of statistical data should be carefully planned and scheduled and this planning must be done by a person who is thoroughly familiar with the operating conditions at each step. One of the most important findings regarding operating principles was the need for very close coordination between field personnel and headquarters personnel in planning, recording, and reporting procedures. Certain procedures were developed in headquarters such as the use of mark-sensing cards with apparatus tests which were not fully coordinated with field personnel. As a consequence, procedures were used in the field which resulted in this technique's being a burden rather than a help to both groups.

One of the most important operating principles was found to be that supervisors must have a thorough knowledge of the data and material. Only on the basis of such thorough knowledge can proper supervision be exercised both in the planning of processing and analysis work and in detecting errors.

## Size of Samples

In planning studies relating to the predictive value of a specific measure the following planning figures were used in regard to the appropriate size of the sample considered adequate to achieve the purposes listed:

| | | |
|---|---|---|
| 1. | For exploratory purposes | 100 |
| 2 | For testing the null hypothesis | 400 |
| 3. | For decisions with respect to use | 1,600 |
| 4 | For establishing precise regression weights | 6,400 |

The values above were those considered appropriate when the evaluations were to be made in terms of product-moment correlation coefficients from continuous series of scores. In situations where other procedures such as biserial correlation coefficients or tetrachoric correlation coefficients were to be used, the sizes of the samples should be increased to compensate for the larger sampling errors of the coefficients.

In general, data from a number of studies made on only moderately sized samples are more valuable than data from one large study including the same number of cases as the total for the several smaller studies.

It is hoped that careful study of some of the recorded experiences of personnel in the Aviation Psychology Program will make it possible to improve the quality of the techniques used in these general fields of psychological research.

# APPENDIX A

# List of Official Directives Referred to in Chapter 3

| No. | Date | Source | Title |
|---|---|---|---|
| 1 | 14 June 1941 | Office Chief of Air Corps | Selection of Flying Cadets. |
| 2 | 15 July 1941 | War Department | Special Orders 165. |
| 3 | 29 July 1941 | Psychological Research Agency | Tentative Plans for the Development of a Battery of Psychological Tests for Administration to Flying Cadets. |
| 4 | 15 Aug. 1941 | Headquarters AAF | Psychological Examinations in Aviation Cadet Replacement Centers. |
| 5 | 20 Aug. 1941 | Headquarters AAF | Study Reference Procurement of Aviation Cadets. |
| 6 | 2 Oct. 1941 | Office Chief of Air Corps | Selection of Aviation Cadets. |
| 7 | 26 Nov. 1941 | Headquarters AAF | Research Program on Selection of Aviation Cadets. |
| 8 | 25 Oct. 1941 | Headquarters AAF | Program of Testing as Outlined by the Research Section, Medical Division, Office Chief of Air Corps. |
| 9 | 19 Nov. 1941 | Headquarters AAF | Tests for the Selection of Aviation Cadets. |
| 10 | 20 Dec. 1941 | Office Chief of Air Corps | Assignment of Functions Pertaining to the Classification and Selection of Air-Crew Personnel. |
| 11 | 7 Jan. 1942 | War Department | Instructions for Aviation Cadet Examining Boards. |
| 12 | 29 Jan. 1942 | Office Chief of Air Corps | New Aviation Cadet Procedure. |
| 13 | 4 Feb. 1942 | AGO | Constitution and Activation of Psychological Research Units. |
| 14 | 2 Feb. 1942 | Headquarters AAF | Classification of Aviation Cadets. |
| 15 | 2 Feb. 1942 | Headquarters AAF | Procedure for Combining Classification Test Scores. |
| 16 | 2 Mar. 1942 | War Department | Circular No. 59, War Department Reorganization. |
| 17 | 13 Apr. 1942 | War Department | Classification and Testing Procedures Commissioned, Warrant, and Enlisted Personnel. |
| 18 | 12 May 1942 | War Department | Selection and Classification of Air-Crew Personnel. |
| 19 | 22 May 1942 | Headquarters AAF | AAF Regulation No. 35-24. |
| 20 | 1 Sept. 1943 | Headquarters AAF | Aviation Cadet Qualifying Examination. |
| 21 | 4 Sept. 1943 | Headquarters AAF | Delegation of Psychological Research Responsibilities. |
| 22 | 9 May 1942 | Headquarters AAF | Classification of Aviation Cadets. |
| 23 | 2 June 1942 | Headquarters AAF | Procedure for Combining Scores on Aviation Cadet Classification Examinations. |
| 24 | 2 June 1942 | Headquarters AAF | Directions for Administering and Scoring Aviation Cadet Classification Examinations. |
| 25 | 3 June 1942 | Headquarters AAF | Revised Weights and Directions for Administering Aviation Cadet Classification Tests. |
| 26 | 3 July 1942 | Headquarters AAF | Revised Weights, Directions and Plan for Administering Aviation Cadet Classification Tests. |
| 27 | 24 Oct. 1942 | Headquarters AAF | Procedure for Preparing Psychological Recommendations for Air-Crew Training Assignments. |
| 28 | 23 Nov. 1942 | Headquarters AAF FTC | Classification, Assignment and Reassignment of Aviation Cadets. |
| 29 | 28 May 1943 | Headquarters AAF | Procedures for Preparing Psychological Recommendations for Air-Crew Training Assignments. |
| 30 | 10 June 1943 | Headquarters AAF FTC | Procedures for Preparing Psychological Recommendations for Air-Crew Training Assignments. |
| 31 | 29 June 1943 | Headquarters AAF | Procedures for Preparing Psychological Recommendations for Air-Crew Training Assignments. |

| No. | Date | Source | Title |
|---|---|---|---|
| 32 | 7 July 1943 | Headquarters AAF | Passing Score on Aviation Cadet Qualifying Examination, Test AC12L. |
| 33 | 6 July 1943 | Headquarters AAF FTC | Procedure to be Followed by the Surgeon in Preparing Recommendations for Air-Crew Training Assignments. |
| 34 | 12 Aug. 1943 | Headquarters AAF FTC | Stanine Requirements for Air-Crew Training. |
| 35 | 24 Sept. 1943 | Headquarters AAF TC | Weights for the October 1943 Classification Test Battery. |
| 36 | 6 Oct. 1943 | Headquarters AAF | Weights for the October 1943 Classification Test Battery. |
| 37 | 15 Nov. 1943 | Headquarters AAF TC | Qualification of Trainees. |
| 38 | 17 Dec. 1943 | Headquarters AAF TC | TWX Re Stanine Requirements. |
| 39 | 27 Dec. 1943 | Headquarters AAF TC | TWX Re Stanine Requirements. |
| 40 | 7 July 1944 | Headquarters AAF | Air-Crew Classification Standards. |
| 41 | 16 Oct. 1944 | Headquarters AAF | Air-Crew Classification Standards. |
| 42 | 9 Dec. 1944 | Headquarters AAF | Air-Crew Classification Standards for B-29 Gunners. |
| 43 | 23 Dec. 1944 | Headquarters AAF | Air-Crew Classification Standards. |
| 44 | 19 Mar. 1945 | Headquarters AAF | Aptitude Qualification for B-29 Gunner Position. |
| 45 | 10 Apr. 1945 | Headquarters AAF | Air-Crew Classification Standards. |
| 46 | 23 May 1945 | Headquarters AAF | Selection of Radar Observer Trainees. |
| 47 | 25 Aug. 1942 | Headquarters AAF | D-4 School. |
| 48 | 13 Sept. 1943 | War Department | Constitution and Activation of a Psychological Research Unit. |
| 49 | 31 July 1943 | Headquarters AAF | Research on Problems of Personnel and Training. |
| 50 | 21 Aug. 1943 | Headquarters AAF | Administration of Aviation Cadet Qualifying Examination. |
| 51 | 28 Aug. 1943 | Headquarters AAF | Psychological Program. |
| 52 | 9 Aug. 1943 | Headquarters AAF | Research on Problems of Personnel and Training. |
| 53 | 19 Apr. 1944 | Headquarters AAF | Psychological Testing Program. |
| 54 | 17 Aug. 1944 | Headquarters AAF | Instructor Selection Tests. |
| 55 | 28 Oct. 1944 | Headquarters AAF PDC | AFPDC Regulation No. 20-19 AAF Convalescent Hospitals. |
| 56 | 9 Oct. 1943 | Headquarters AAF TC | Establishment of Psychological Research Project at Santa Ana Army Air Base. |
| 57 | 5 Jan. 1944 | Headquarters AAF TC | Establishment of Psychological Research Project (Bombardier) at Midland Army Air Field. |
| 57 | 5 Jan. 1944 | Headquarters AAF TC | Establishment of Psychological Research Project (Navigation) at Selman Field. |
| 57 | 5 Jan. 1944 | Headquarters AAF TC | Establishment of Psychological Research Project at Central Instructors School (Pilot). |
| 58 | 25 Nov. 1944 | Headquarters AAF TC | Establishment of Psychological Research Project (Radar). |
| 59 | 6 Apr. 1945 | Headquarters AAF TC | Establishment of Psychological Research Project (Combat Crew). |
| 60 | 23 June 1945 | Headquarters AAF TC | Establishment of Psychological Research Project (Flight Engineer). |
| 61 | 10 Nov. 1943 | Headquarters AAF | TDY Orders to Col. John C. Flanagan for European Theater. |
| 62 | 8 Mar. 1944 | Headquarters AAF | Selection of Personnel for Lead Crews in the United States Strategic Air Forces in Europe. |
| 63 | 24 Mar. 1944 | Headquarters AAF | Mission of Psychologists to U.K. |
| 64 | 27 Mar. 1944 | Headquarters AAF | Telegram to USSTAF making 6 officers and 15 EM available for 90 days TDY. |
| 65 | 3 Apr. 1944 | USSTAF | Telegram requesting research detachment. |
| 66 | 18 Apr. 1944 | Headquarters AAF | Telegram to TC directing 6 officers and 15 EM be ordered on TDY overseas. |
| 67 | 28 Apr. 1944 | Headquarters AAF | Mission of Air-Crew Evaluation and Research Detachment. |
| 68 | 28 Aug. 1944 | Headquarters AAF | Extension TDY orders AERD Personnel for 90 days. |
| 69 | 29 Sept. 1944 | Headquarters AAF | Mission of Air-Crew Evaluation and Research Detachment 15 AF. |
| 70 | 16 Nov. 1944 | Headquarters AAF | Mission of Air-Crew Evaluation and Research Detachment POA. |

| No. | Date | Source | Title |
|---|---|---|---|
| 71 | 28 Aug. 1944 | Headquarters AAF | Telegram to 15 AF requesting concurrence in sending personnel. |
| 72 | 11 Sept. 1944 | Headquarters 15 AF | Telegram approving TDY 4 officers and 8 EM. |
| 73 | 9 Jan. 1945 | Headquarters AAF | Telegram re extension of orders for 90 days TDY with 12 AF. |
| 73 | 7 Feb. 1945 | Headquarters AAF | Extension of TDY orders for 60 days 8 and 9 AF. |
| 74 | 2 Nov. 1944 | Headquarters AAF | Telegram requesting concurrence for sending 4 officers and 8 EM to POA. |
| 75 | 12 June 1944 | Headquarters AAF | Staff Study, Air-Crew Selection and Evaluation in the Air Forces. |
| 76 | 23 Aug. 1944 | Headquarters AAF | Research Officers. |
| 77 | 3 Oct. 1944 | Headquarters AAF | Authorization for Allotment of Research Officers. |
| 78 | 29 Sept. 1944 | Headquarters AAF | Air-Crew Selection and Evaluation in the Air Forces. |
| 79 | 23 Sept. 1944 | Headquarters AAF | Air-Crew Selection and Evaluation in the Air Forces. |
| 80 | 12 July 1945 | War Department | TO & E No. 8-460, Central Medical Establishment Aviation. |
| 81 | 6 Aug. 1945 | Headquarters AAF | Psychological Personnel for CME's. |
| 82 | 3 Mar. 1945 | Headquarters AAF | Establishment of Central Processing Station. |
| 83 | 7 June 1945 | Headquarters AAF | Standardized Operating Procedure for Inbound Combat Crew Processing. |
| 84 | 7 June 1945 | Headquarters AAF | AAF Letter 50-117, Screening of Combat Crew Personnel. |
| 85 | 29 May 1945 | Headquarters AAF | Psychological Research on Problems of Aviation Equipment. |
| 86 | 3 Jan. 1944 | RAF Delegation | Request for AAF Tests. |
| 87 | 26 Jan. 1944 | War Department | Authority Granted to Reprint AAF Psychological Tests. |
| 88 | 29 Aug. 1944 | RAF Delegation | Request for AAF Tests for use by Dominion Air Forces. |
| 89 | 14 Sept. 1944 | Headquarters AAF | Release of AAF Air-Crew Classification Tests to the Dominion Air Forces. |
| 90 | 7 Mar. 1944 | Joint Air Commission | Telegram Recommending an Officer be sent to MAAF Algiers. |
| 91 | 10 Mar. 1944 | Headquarters AAF | Proposal for setting up French Testing System Approved. |
| 92 | 27 July 1945 | AF Pacific Command | Request for personnel to set up system for selection of Filipinos for Pilot Training. |
| 93 | 2 July 1945 | Headquarters AAF | Selection of Filipinos for Pilot Training. |
| 94 | 18 Oct. 1945 | Headquarters AAF | Temporary Duty in Japan. |
| 95 | 29 Oct. 1945 | Headquarters AAF | Temporary Duty in Japan. |
| 96 | 5 May 1945 | Headquarters AAF | TDY Orders for Lt. Colonel Paul M. Fitts to European Theater. |
| 97 | 12 July 1945 | Headquarters AAF | AAF Regulation No. 20-59 Organization Aviation Psychology Program in the AAF. |
| 98 | 19 July 1945 | Headquarters AAF | AAF Letter 20-101, Aviation Psychology in the AAF Training Command. |
| 99 | 19 July 1945 | Headquarters AAF | AAF Letter 20-102, Aviation Psychology in the AAF Personnel Distribution Command. |
| 100 | 19 July 1945 | Headquarters AAF | AAF Letter 20-103, Aviation Psychology in the Continental Air Forces. |
| 101 | 27 Mar. 1946 | Headquarters AAF | Report of Conference in Gen. LeMay's Office. |
| 102 | 13 Aug. 1946 | Headquarters AAF | AAF Regulation No. 20-59, Organization Aviation Psychological Program in the AAF. |
| 103 | 24 May 1946 | Headquarters AAF | Psychological Research Program in the Army Air Forces. |

# APPENDIX B

# Intercorrelations of Tests and Other Variables in the Experimental Group and in Samples of United States Military Academy Cadets

702323—47—21

TABLE B.1.—*Intercorrelations of classification tests, stanines and other measures of experimental group*[1]

[N = 1,012]

| Variable | Code | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | M | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Biographical Data Navigator | CF602D | | 0.15 | 0.11 | 0.05 | 0.10 | 0.16 | −0.05 | 0.02 | 0.05 | 0.04 | 0.13 | 0.13 | 0.01 | 0.02 | 0.02 | 0.03 | 0.10 | 0.07 | 0.09 | 0.12 | 0.06 | 0.10 | 5.00 | 1.81 |
| 2. Biographical Data Pilot | CE602D | 0.15 | | .16 | .23 | .09 | .41 | .29 | .14 | .25 | .40 | .04 | .03 | .15 | .28 | .23 | .29 | .12 | .05 | .58 | .05 | .31 | .21 | 5.03 | 1.95 |
| 3. Spatial Orientation I | CP501B | .11 | .16 | | .43 | .21 | .45 | .18 | .43 | .33 | .24 | .22 | .19 | .22 | .26 | .35 | .12 | .37 | .20 | .43 | .32 | .41 | .15 | 4.83 | 1.97 |
| 4. Spatial Orientation II | CP503B | .05 | .23 | .43 | | .30 | .39 | .37 | .44 | .46 | .39 | .26 | .26 | .15 | .29 | .30 | .20 | .37 | .05 | .54 | .34 | .55 | .15 | 5.01 | 1.85 |
| 5. Reading Comprehension | CI614H | .10 | .09 | .21 | .30 | | .41 | .36 | .40 | .38 | .44 | .43 | .44 | .09 | .16 | .20 | .09 | .30 | .05 | .37 | .53 | .52 | .18 | 4.85 | 2.06 |
| 6. Dial and Table Reading | CP622–21A | .16 | .41 | .45 | .39 | .41 | | .29 | .53 | .43 | .29 | .44 | .45 | .16 | .28 | .37 | .10 | .47 | .19 | .45 | .57 | .51 | .23 | 4.96 | 1.95 |
| 7. Mechanical Principles | CI903B | −.05 | .29 | .18 | .37 | .36 | .29 | | .32 | .45 | .47 | .27 | .36 | .26 | .37 | .37 | .31 | .32 | .07 | .65 | .32 | .59 | .13 | 4.92 | 1.93 |
| 8. Instrument Comprehension I | CI615B | .02 | .14 | .43 | .44 | .40 | .53 | .32 | | .45 | .36 | .32 | .35 | .16 | .28 | .39 | .12 | .42 | .18 | .41 | .45 | .52 | .20 | 5.13 | 1.83 |
| 9. Instrument Comprehension II | CI616B | .05 | .25 | .33 | .46 | .38 | .43 | .45 | .45 | | .44 | .25 | .30 | .24 | .29 | .40 | .25 | .44 | .15 | .64 | .37 | .54 | .15 | 4.93 | 1.94 |
| 10. General Information | CE505E | .04 | .40 | .24 | .39 | .44 | .29 | .47 | .36 | .44 | | .29 | .23 | .16 | .28 | .31 | .29 | .28 | .09 | .67 | .41 | .63 | .21 | 4.95 | 1.85 |
| 11. Mathematics A | CI702F | .11 | .04 | .22 | .26 | .43 | .44 | .27 | .32 | .25 | .29 | | .50 | .04 | .12 | .15 | −.03 | .27 | .05 | .26 | .44 | .43 | .16 | 4.83 | 2.19 |
| 12. Mathematics B | CI206C | .13 | .03 | .19 | .26 | .44 | .45 | .36 | .35 | .30 | .23 | .50 | | .02 | .16 | .20 | .03 | .31 | .01 | .29 | .55 | .45 | .12 | 4.85 | 2.05 |
| 13. Rotary Pursuit | CP410B | .01 | .15 | .22 | .15 | .09 | .16 | .26 | .16 | .24 | .16 | .04 | .02 | | .36 | .36 | .31 | .23 | .29 | .41 | .07 | .20 | .07 | 4.99 | 1.80 |
| 14. Two Hand Coordination | CM101A | .02 | .28 | .26 | .29 | .16 | .28 | .37 | .28 | .29 | .28 | .12 | .16 | .36 | | .51 | .35 | .27 | .28 | .59 | .30 | .36 | .11 | 5.03 | 1.96 |
| 15. Complex Coordination | CM701A | .07 | .23 | .35 | .30 | .20 | .37 | .37 | .39 | .40 | .31 | .15 | .20 | .36 | .51 | | .32 | .43 | .29 | .73 | .26 | .41 | .13 | 4.90 | 1.88 |
| 16. Rudder Control | CM120B | .03 | .29 | .12 | .20 | .09 | .10 | .31 | .12 | .25 | .29 | −.03 | .03 | .31 | .35 | .32 | | .17 | .15 | .57 | .04 | .23 | .15 | 4.82 | 1.91 |
| 17. Discrimination Reaction Time | CP611D | .10 | .12 | .37 | .37 | .30 | .47 | .32 | .42 | .44 | .28 | .27 | .31 | .23 | .27 | .43 | .17 | | .28 | .47 | .35 | .44 | .17 | 4.94 | 1.93 |
| 18. Finger Dexterity | CM116A | .07 | .05 | .20 | .05 | .05 | .19 | .07 | .18 | .15 | .09 | .05 | .01 | .29 | .28 | .29 | .15 | .25 | | .26 | .11 | .15 | .07 | 4.97 | 1.94 |
| 19. Pilot Stanine | | .09 | .58 | .43 | .54 | .37 | .45 | .65 | .41 | .64 | .67 | .26 | .29 | .41 | .55 | .70 | .57 | .47 | .24 | | .37 | .66 | .24 | 4.01 | 2.03 |
| 20. General Classification Test | | .12 | .05 | .32 | .34 | .53 | .57 | .32 | .45 | .37 | .41 | .44 | .55 | .07 | .30 | .26 | .04 | .35 | .11 | .37 | | .55 | .21 | 4.97 | 1.99 |
| 21. AAFQE Total | | .06 | .31 | .41 | .55 | .52 | .51 | .59 | .52 | .54 | .63 | .43 | .45 | .20 | .36 | .41 | .23 | .45 | .11 | .66 | .55 | | .23 | 4.96 | 1.97 |
| 22. ARMA[2] | | .10 | .21 | .15 | .15 | .18 | .23 | .13 | .20 | .15 | .21 | .16 | .12 | .07 | .11 | .13 | .15 | .17 | .07 | .24 | .21 | .23 | | 1.34 | .48 |

[1]Group reported consisted of all cases on which data were complete which entered Preflight. Means and standard deviations for all test scores are in single-digit scores based on larger groups as follows: For classification tests, $N = 1,275$; for General Classification Test, $N = 1,090$; for AAF Qualifying Examination (AAFQE), $N = 1,270$ for total score, and $N = 1,223$ for part scores. Stanines based on conversion tables used for regular air-crew candidates.

[2]Adaptability Rating for Military Aeronautics. This variable was coded as follows: 1, predicted success; 2, predicted failure. Correlations with continuous variables are product-moment coefficients, with signs adjusted so that a positive coefficient indicates association of good performance and prediction of success. M and SD of ARMA unchanged.

TABLE B.1.—Variables entered in the following USMA intercorrelation tables (N = 815)

| Classification Tests | Test No. | Mean[a] | $SD_1$ |
|---|---|---|---|
| 1. Arithmetic Reasoning | CI206C | 5.05 | 1.96 |
| 2. Dial and Table Reading | CP621-11A | 5.06 | 1.96 |
| 3. Spatial Orientation I | CP501B | 5.04 | 1.96 |
| 4. Spatial Orientation II | CP503B | 4.98 | 1.93 |
| 5. Biographical Data Navigator | CE602D | 5.00 | 2.00 |
| 6. Biographical Data Pilot | CE601D | 5.03 | 1.96 |
| 7. Numerical Operations Front | CI702B | 5.06 | 2.00 |
| 8. Numerical Operations Back | CI702B | 5.01 | 1.98 |
| 9. Reading Comprehension | CI514H | 5.05 | 1.98 |
| 10. Judgment | CI301C | 5.14 | 2.01 |
| 11. General Information | CI305F | 5.01 | 1.98 |
| 12. Instrument Comprehension | CI416C | 5.02 | 1.98 |
| 13. Mechanical Principles | CI923B | 4.95 | 1.98 |
| 14. Mechanical Information | CI905B | 4.97 | 1.95 |
| 15. Speed of Identification | CP610A | 4.94 | 1.89 |
| 16. Rotary Pursuit | CP410B | 5.06 | 1.98 |
| 17. Two-Hand Coordination | CM101A | 4.98 | 1.94 |
| 18. Complex Coordination | CM701A | 4.99 | 1.90 |
| 19. Rudder Control | CM120B | 5.00 | 1.97 |
| 20. Discrimination Reaction Time | CP611D | 5.01 | 1.97 |
| 21. Finger Dexterity | CM116A | 5.01 | 1.93 |
| 22. Bombardier Stanine | | 7.45 | 1.46 |
| 23. Navigator Stanine | | 8.18 | 1.17 |
| 24. Bomber Pilot Stanine | | 7.09 | 1.80 |
| 25. Fighter Pilot Stanine | | 6.21 | 1.92 |
| 26. Officer Quality Score | | 5.02 | 1.96 |

| Fourth-Year USMA Measures[b] | Mean | $SD_2$ |
|---|---|---|
| 27. Language | 4.99 | 1.96 |
| 28. Mathematics | 5.04 | 1.95 |
| 29. English | 5.03 | 1.99 |
| 30. Military Topography and Graphics | 5.04 | 1.96 |
| 31. Tactics | 5.07 | 1.96 |
| 32. Military Physical Efficiency | 5.07 | 1.65 |
| 33. Total Proportional Parts | 5.05 | 1.95 |
| 34. Officer Average Rating | 4.98 | 2.00 |
| 35. Cadet Average Rating | 5.08 | 1.92 |
| 36. Final Composite Rating | 5.04 | 1.95 |
| *Third-Year USMA Measures* | | |
| 37. Language | 5.06 | 1.95 |
| 38. Mathematics | 5.08 | 1.93 |
| 39. Physics | 5.08 | 1.90 |
| 40. Military Topography and Graphics | 5.03 | 1.96 |
| 41. Tactics | 5.01 | 1.94 |
| 42. History | 5.04 | 1.93 |
| 43. Chemistry | 5.05 | 1.93 |
| 44. Tactical Officer Average Rating | 5.07 | 1.85 |
| 45. Academic Officer Average Rating | 5.06 | 1.97 |
| 46. Cadet Average Rating | 5.03 | 1.92 |
| 47. Final Composite Rating | 5.05 | 1.94 |
| *USMA Physical Measures* | | |
| 48. 300 Yard Run | 5.02 | 1.90 |
| 49. Dodge Run | 5.00 | 1.91 |
| 50. Standing Broad Jump | 5.06 | 1.99 |
| 51. Vertical Jump | 5.09 | 1.99 |
| 52. Bar Vault | 4.97 | 1.83 |
| 53. Rope Climb | 5.01 | 1.65 |
| 54. Sit Ups | 4.99 | 1.54 |
| 55. Chins | 5.08 | 1.86 |
| 56. Dips | 5.04 | 1.84 |
| 57. Softball Throw | 4.98 | 1.94 |

[a] With the exception of Stanines, all variables were converted into single-digit scores with a theoretical mean of 5.00 and theoretical standard deviation of 2.00, using the identical West Point population. For Stanines, conversion tables used in the processing of air-crew candidates in the PRU's and MPEU's were utilized. Consequently no comparisons of classification-test scores of West Point cadets and regular air-crew candidates is possible from these data.
[b] These refer to grades in the first year of study at the United States Military Academy.

TABLE B.3.—*Intercorrelations of classification tests for USMA class of 1946 (N=815)*[1]

[Decimals before correlation coefficients omitted]

| Variable | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 38 | 19 | 30 | 15 | −01 | 29 | 40 | 54 | 33 | 16 | 29 | 37 | 32 | 12 | −04 | 14 | 17 | 09 | 21 | 04 | 51 | 61 | 37 | 18 | 75 |
| 2 | 38 | | 41 | 27 | 15 | 03 | 52 | 50 | 25 | 19 | 13 | 38 | 11 | 07 | 31 | 18 | 25 | 37 | 14 | 31 | 20 | 61 | 61 | 35 | 37 | 36 |
| 3 | 19 | 41 | | 46 | 09 | 18 | 19 | 19 | 19 | 18 | 22 | 37 | 19 | 12 | 47 | 13 | 25 | 27 | 17 | 26 | 18 | 57 | 40 | 38 | 42 | 39 |
| 4 | 30 | 27 | 46 | | 15 | 19 | −01 | 08 | 38 | 31 | 29 | 38 | 42 | 34 | 39 | 10 | 29 | 30 | 17 | 23 | 08 | 54 | 51 | 53 | 48 | 51 |
| 5 | 15 | 15 | 09 | 15 | | 12 | 07 | 04 | 11 | 12 | 05 | 10 | 08 | 04 | 10 | −01 | −03 | 05 | −06 | 02 | 01 | 11 | 24 | 14 | 08 | 16 |
| 6 | −01 | 03 | 18 | 19 | 12 | | −13 | −07 | 03 | 21 | 49 | 22 | 33 | 36 | 21 | 15 | 26 | 17 | 20 | 07 | 06 | 12 | 08 | 52 | 45 | 16 |
| 7 | 29 | 52 | 19 | −01 | 07 | −13 | | 63 | 10 | 03 | −03 | 12 | −17 | −11 | 07 | 06 | 05 | 16 | −03 | 19 | 15 | 36 | 41 | 03 | 07 | 13 |
| 8 | 40 | 50 | 19 | 08 | 04 | −07 | 63 | | 13 | 04 | 01 | 22 | −04 | −06 | 13 | 07 | 10 | 11 | 00 | 24 | 11 | 39 | 44 | 11 | 15 | 22 |
| 9 | 54 | 25 | 19 | 38 | 11 | 03 | 10 | 13 | | 44 | 30 | 38 | 45 | 43 | 16 | 08 | 19 | 22 | 19 | 18 | 05 | 53 | 54 | 55 | 39 | 83 |
| 10 | 33 | 19 | 18 | 31 | 12 | 21 | 03 | 04 | 44 | | 34 | 23 | 38 | 36 | 13 | 13 | 19 | 16 | 16 | 15 | 05 | 35 | 38 | 52 | 26 | 67 |
| 11 | 16 | 13 | 22 | 29 | 05 | 49 | −03 | 01 | 30 | 34 | | 36 | 46 | 54 | 21 | 15 | 28 | 26 | 28 | 14 | 06 | 30 | 24 | 59 | 65 | 44 |
| 12 | 29 | 38 | 37 | 38 | 10 | 22 | 12 | 22 | 38 | 23 | 36 | | 39 | 29 | 33 | 17 | 34 | 36 | 31 | 30 | 10 | 49 | 43 | 58 | 67 | 44 |
| 13 | 37 | 11 | 19 | 42 | 08 | 33 | −17 | −04 | 45 | 38 | 46 | 39 | | 65 | 18 | 21 | 44 | 34 | 40 | 22 | 09 | 42 | 37 | 76 | 64 | 65 |
| 14 | 32 | 07 | 12 | 34 | 04 | 36 | −11 | −06 | 43 | 36 | 54 | 29 | 65 | | 14 | 13 | 35 | 28 | 29 | 13 | 09 | 33 | 29 | 63 | 54 | 56 |
| 15 | 12 | 31 | 47 | 39 | 10 | 21 | 07 | 13 | 16 | 13 | 21 | 33 | 18 | 14 | | 03 | 22 | 25 | 14 | 22 | 14 | 40 | 32 | 33 | 44 | 24 |
| 16 | −04 | 18 | 13 | 10 | −01 | 15 | 06 | 07 | 08 | 13 | 15 | 17 | 21 | 13 | 03 | | 34 | 35 | 35 | 22 | 33 | 29 | 14 | 39 | 37 | 13 |
| 17 | 14 | 25 | 25 | 29 | −03 | 26 | 05 | 10 | 19 | 19 | 28 | 34 | 44 | 35 | 22 | 34 | | 51 | 47 | 34 | 28 | 52 | 32 | 52 | 65 | 32 |
| 18 | 17 | 37 | 27 | 30 | 05 | 17 | 16 | 11 | 22 | 18 | 26 | 35 | 38 | 25 | 25 | 35 | 51 | | 45 | 40 | 32 | 55 | 35 | 55 | 66 | 31 |
| 19 | 09 | 14 | 17 | 17 | −06 | 20 | −03 | 00 | 19 | 16 | 28 | 31 | 40 | 29 | 14 | 35 | 47 | 45 | | 25 | 23 | 35 | 20 | 51 | 64 | 26 |
| 20 | 21 | 31 | 26 | 23 | 02 | 07 | 19 | 24 | 18 | 15 | 14 | 30 | 22 | 13 | 22 | 22 | 34 | 40 | 25 | | 28 | 57 | 37 | 34 | 47 | 32 |
| 21 | 04 | 20 | 18 | 08 | 01 | 06 | 15 | 11 | 05 | 05 | 06 | 10 | 09 | 09 | 14 | 33 | 28 | 32 | 23 | 28 | | 47 | 16 | 20 | 26 | 10 |
| 22 | 51 | 61 | 57 | 54 | 11 | 12 | 36 | 39 | 53 | 35 | 30 | 49 | 42 | 33 | 40 | 29 | 52 | 55 | 35 | 57 | 47 | | 82 | 66 | 65 | 67 |
| 23 | 61 | 61 | 40 | 51 | 24 | 08 | 41 | 44 | 54 | 38 | 24 | 43 | 37 | 29 | 32 | 14 | 32 | 35 | 20 | 37 | 18 | 82 | | 58 | 31 | 68 |
| 24 | 37 | 35 | 38 | 53 | 14 | 57 | 03 | 11 | 55 | 52 | 50 | 58 | 76 | 61 | 33 | 39 | 52 | 55 | 51 | 34 | 20 | 66 | 58 | | 87 | 72 |
| 25 | 28 | 37 | 42 | 48 | 08 | 46 | 07 | 15 | 39 | 36 | 65 | 67 | 64 | 54 | 44 | 37 | 65 | 66 | 64 | 47 | 25 | 65 | 51 | 87 | | 57 |
| 26 | 75 | 36 | 39 | 51 | 16 | 16 | 13 | 27 | 83 | 67 | 44 | 44 | 65 | 56 | 24 | 13 | 32 | 31 | 25 | 32 | 80 | 67 | 68 | 72 | 57 | |

[1] Variables identified in table B.2.

TABLE B.4.—*Intercorrelations of classification tests and USMA grades, ratings, and physical measures for USMA cadets, class of 1946 ($N = 815$)*[1]

[Decimals before correlation coefficients omitted]

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | 22 | 23 | 07 | 03 | 01 | −23 | 28 | 24 | 21 | −01 | −11 | 06 | −10 |
| 28 | 53 | 40 | 25 | 30 | 17 | 00 | 28 | 32 | 44 | 22 | 10 | 27 | 26 |
| 29 | 42 | 30 | 18 | 21 | 04 | −18 | 30 | 32 | 43 | 16 | 04 | 19 | 13 |
| 30 | 49 | 59 | 41 | 49 | 19 | 22 | 21 | 26 | 45 | 37 | 28 | 46 | 46 |
| 31 | 44 | 26 | 22 | 42 | 13 | 18 | 05 | 10 | 53 | 35 | 35 | 34 | 46 |
| 32 | −16 | 00 | 07 | 01 | −02 | 24 | −05 | −03 | −11 | −02 | 12 | 12 | 11 |
| 33 | 54 | 45 | 29 | 33 | 16 | −03 | 33 | 35 | 49 | 25 | 12 | 32 | 26 |
| 34 | 21 | 34 | 25 | 19 | 05 | 00 | 20 | 25 | 20 | 11 | 07 | 21 | 12 |
| 35 | 00 | 18 | 14 | 06 | 08 | 09 | 17 | 16 | −05 | 01 | 03 | 11 | 01 |
| 36 | 11 | 27 | 20 | 13 | 07 | 06 | 17 | 22 | 07 | 05 | 06 | 18 | 06 |
| 37 | 30 | 27 | 10 | 07 | 06 | −17 | 30 | 27 | 27 | 05 | −04 | 13 | −01 |
| 38 | 49 | 33 | 21 | 23 | 17 | 02 | 24 | 26 | 44 | 21 | 11 | 22 | 26 |
| 39 | 60 | 37 | 22 | 32 | 20 | 08 | 23 | 26 | 56 | 32 | 27 | 32 | 42 |
| 40 | 53 | 52 | 47 | 57 | 21 | 20 | 23 | 28 | 56 | 42 | 34 | 52 | 45 |
| 41 | 32 | 16 | 07 | 19 | 09 | −05 | 07 | 09 | 43 | 21 | 12 | 14 | 19 |
| 42 | 42 | 30 | 17 | 23 | 07 | −13 | 23 | 19 | 45 | 19 | −01 | 10 | 05 |
| 43 | 59 | 39 | 20 | 26 | 18 | 02 | 26 | 30 | 55 | 28 | 19 | 26 | 34 |
| 44 | 04 | 18 | 11 | 03 | 03 | 09 | 09 | 11 | 01 | 01 | 11 | 14 | 03 |
| 45 | 39 | 36 | 27 | 28 | 14 | 02 | 26 | 21 | 41 | 20 | 17 | 31 | 28 |
| 46 | −01 | 19 | 14 | 04 | 05 | 05 | 17 | 14 | −05 | −01 | 05 | 05 | −02 |
| 47 | 02 | 24 | 18 | 09 | 06 | 09 | 17 | 18 | 03 | 02 | 12 | 15 | 04 |
| 48 | 05 | 01 | 07 | −07 | 01 | 09 | 02 | 03 | −06 | −05 | 01 | −03 | −04 |
| 49 | −17 | 14 | 11 | −03 | 01 | 13 | 12 | 05 | −18 | −11 | 03 | 02 | −12 |
| 50 | −07 | 05 | 11 | −04 | 01 | 13 | 00 | 04 | −05 | −01 | 05 | 07 | 05 |
| 51 | −08 | 03 | 07 | −04 | 01 | 13 | 03 | 05 | −15 | −07 | 06 | 04 | 00 |
| 52 | −12 | 00 | 06 | −01 | 03 | 20 | −05 | −03 | −06 | −03 | 13 | 07 | 09 |
| 53 | −06 | −03 | 02 | 00 | 00 | 19 | −06 | −05 | 07 | 01 | 09 | 01 | 16 |
| 54 | −15 | 01 | 02 | −06 | 03 | 07 | 03 | −01 | −08 | −04 | 01 | −02 | −09 |
| 55 | −05 | −05 | −03 | −06 | 02 | 13 | −06 | −01 | −05 | −10 | 04 | −03 | 02 |
| 56 | −12 | −07 | −02 | −05 | 06 | 11 | −01 | −04 | −09 | −05 | −03 | −05 | −06 |
| 57 | −11 | 06 | 03 | −13 | 00 | 00 | 13 | 12 | −15 | −06 | 02 | 01 | −12 |

| Variable | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | −02 | 01 | −01 | −05 | 02 | −04 | 06 | 05 | 17 | 20 | −03 | −04 | 14 |
| 28 | 23 | 19 | 03 | 11 | 21 | 06 | 22 | 10 | 45 | 45 | 31 | 25 | 51 |
| 29 | 12 | 11 | 06 | 05 | 11 | 03 | 13 | 18 | 38 | 37 | 20 | 15 | 41 |
| 30 | 43 | 34 | 14 | 30 | 35 | 20 | 28 | 16 | 60 | 57 | 57 | 51 | 62 |
| 31 | 56 | 21 | 05 | 23 | 20 | 19 | 16 | 07 | 42 | 63 | 51 | 43 | 61 |
| 32 | 05 | 02 | 23 | 13 | 09 | 16 | 12 | 06 | 08 | −04 | 16 | 19 | −07 |
| 33 | 26 | 20 | 07 | 15 | 24 | 10 | 23 | 14 | 51 | 51 | 36 | 33 | 54 |
| 34 | 10 | 17 | 20 | 13 | 18 | 11 | 14 | 15 | 31 | 35 | 22 | 22 | 23 |
| 35 | 01 | 14 | 14 | 02 | 03 | 07 | 11 | 08 | 13 | 09 | 10 | 13 | −01 |
| 36 | 07 | 18 | 19 | 12 | 13 | 09 | 13 | 12 | 23 | 12 | 17 | 18 | 11 |
| 37 | 07 | 03 | −02 | −02 | 05 | −02 | 05 | 05 | 23 | 17 | 05 | 03 | 24 |
| 38 | 23 | 12 | 03 | 11 | 20 | 06 | 18 | 08 | 40 | 40 | 30 | 22 | 48 |
| 39 | 40 | 16 | 05 | 16 | 24 | 10 | 27 | 07 | 43 | 51 | 45 | 36 | 65 |
| 40 | 41 | 40 | 11 | 31 | 37 | 19 | 31 | 13 | 68 | 63 | 62 | 53 | 69 |
| 41 | 39 | 05 | 03 | 08 | 13 | 09 | 09 | 03 | 28 | 30 | 26 | 20 | 39 |
| 42 | 12 | 10 | −02 | 01 | 08 | −03 | 10 | 07 | 34 | 35 | 13 | 05 | 41 |
| 43 | 37 | 16 | 03 | 13 | 20 | 06 | 20 | 09 | 47 | 49 | 37 | 23 | 61 |
| 44 | 04 | 14 | 15 | 11 | 10 | 06 | 06 | 05 | 12 | 05 | 13 | 16 | 04 |
| 45 | 25 | 20 | 11 | 14 | 22 | 13 | 21 | 12 | 44 | 42 | 36 | 32 | 45 |
| 46 | 00 | 13 | 14 | 07 | 09 | 05 | 09 | 07 | 13 | 07 | 09 | 13 | −01 |
| 47 | 05 | 18 | 15 | 11 | 12 | 03 | 11 | 07 | 19 | 14 | 15 | 18 | 03 |
| 48 | −03 | −02 | 10 | −04 | −03 | 00 | −01 | 02 | −03 | −04 | −02 | −02 | −08 |
| 49 | −11 | 05 | 09 | 04 | 08 | 03 | 10 | 06 | 01 | −04 | −03 | 05 | −18 |
| 50 | 03 | 07 | 13 | 04 | 01 | 06 | 05 | 04 | 02 | 00 | 06 | 09 | −05 |
| 51 | −04 | 03 | 10 | −03 | −01 | 04 | 02 | 05 | −03 | −01 | 02 | 05 | −11 |
| 52 | 09 | 00 | 14 | 03 | 03 | 10 | 02 | 04 | −01 | −04 | 12 | 13 | −05 |
| 53 | 11 | 02 | 09 | 12 | 06 | 15 | 04 | 03 | 01 | −02 | 14 | 14 | 02 |
| 54 | −06 | 07 | 08 | −05 | −02 | −05 | 00 | 03 | −07 | −07 | −05 | −03 | −13 |
| 55 | 02 | −03 | 01 | 04 | 00 | 06 | 00 | 00 | −07 | −06 | 02 | 04 | −08 |
| 56 | −04 | −04 | 01 | 01 | −01 | 06 | 00 | 02 | −08 | −09 | −02 | 00 | −11 |
| 57 | −13 | −05 | 13 | −08 | −05 | −05 | 04 | 00 | −04 | −01 | −08 | −04 | −13 |

[1] Variables identified in table B.1.

TABLE B.5.—*Intercorrelations of USMA grades, ratings, and physical measures for USMA cadets, class of 1946 (N = 815)*

[Decimals before correlation coefficients omitted]

| Variable[1] | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | | 49 | 55 | 21 | 21 | 03 | 66 | 44 | 13 | 31 | 83 | 39 | 34 | 20 | 16 | 69 | 35 | 14 | 39 | 14 | 20 | 02 | 02 | 01 | -01 | -03 | 01 | 07 | 11 | 06 | 00 |
| 28 | 49 | | 51 | 68 | 50 | 05 | 92 | 44 | 21 | 36 | 50 | 81 | 77 | 61 | 31 | 53 | 74 | 18 | 62 | 20 | 29 | 02 | -04 | 06 | -01 | -01 | 01 | -03 | 02 | -05 | -04 |
| 29 | 55 | 51 | | 40 | 39 | -04 | 70 | 50 | 18 | 37 | 54 | 43 | 44 | 41 | 32 | 56 | 48 | 19 | 50 | 18 | 27 | -02 | -10 | -07 | -07 | -09 | -06 | -03 | 02 | -05 | -08 |
| 30 | 21 | 65 | 40 | | 59 | 05 | 72 | 41 | 20 | 33 | 30 | 59 | 69 | 84 | 30 | 40 | 66 | 18 | 58 | 16 | 26 | 01 | -03 | 05 | 00 | 02 | 07 | -02 | -01 | -05 | -07 |
| 31 | 21 | 50 | 39 | 59 | | 02 | 56 | 26 | 10 | 20 | 28 | 43 | 58 | 61 | 39 | 41 | 57 | 14 | 42 | 08 | 18 | -04 | -13 | 00 | -05 | 00 | 04 | -03 | 02 | -01 | -10 |
| 32 | 03 | 65 | -04 | 05 | 02 | | 08 | 27 | 40 | 38 | -03 | -02 | -05 | 02 | -09 | -10 | -09 | 34 | 12 | 41 | 40 | 40 | 44 | 46 | 41 | 30 | 45 | 31 | 45 | 39 | 36 |
| 33 | 66 | 92 | 70 | 72 | 56 | 08 | | 55 | 25 | 44 | 66 | 76 | 75 | 67 | 34 | 60 | 73 | 23 | 67 | 24 | 34 | 03 | -03 | 05 | -01 | 00 | 04 | 00 | 06 | -02 | -03 |
| 34 | 44 | 44 | 50 | 41 | 26 | 27 | 55 | | 53 | 84 | 41 | 32 | 32 | 38 | 12 | 31 | 31 | 45 | 55 | 51 | 58 | 12 | 14 | 14 | 09 | 10 | 05 | 04 | 03 | 05 | 11 |
| 35 | 13 | 21 | 18 | 20 | 10 | 40 | 25 | 53 | | 88 | 09 | 10 | 10 | 16 | -06 | 05 | 09 | 61 | 37 | 75 | 75 | 26 | 21 | 23 | 20 | 19 | 12 | 14 | 13 | 15 | 23 |
| 36 | 31 | 36 | 37 | 33 | 20 | 38 | 44 | 84 | 88 | | 28 | 22 | 23 | 30 | 04 | 18 | 27 | 60 | 53 | 73 | 76 | 27 | 20 | 21 | 16 | 16 | 09 | 10 | 08 | 11 | 20 |
| 37 | 83 | 50 | 54 | 30 | 28 | -03 | 66 | 41 | 09 | 28 | | 49 | 46 | 30 | 22 | 53 | 48 | 10 | 45 | 05 | 17 | -01 | 00 | -02 | -07 | -02 | -03 | 05 | 07 | 02 | -01 |
| 38 | 39 | 81 | 43 | 59 | 43 | -02 | 76 | 32 | 10 | 22 | 49 | | 80 | 54 | 29 | 51 | 76 | 12 | 61 | 12 | 21 | 00 | -07 | 04 | -04 | -05 | -01 | -03 | -02 | -05 | -05 |
| 39 | 34 | 77 | 44 | 69 | 58 | -05 | 75 | 32 | 10 | 23 | 46 | 80 | | 68 | 35 | 52 | 83 | 12 | 61 | 12 | 22 | 01 | -09 | 03 | -03 | -03 | -01 | -04 | -02 | -04 | -07 |
| 40 | 20 | 61 | 41 | 84 | 61 | 02 | 67 | 38 | 16 | 30 | 30 | 54 | 68 | | 36 | 43 | 64 | 16 | 55 | 14 | 24 | -02 | -04 | 03 | -02 | -01 | 03 | -04 | -05 | -05 | -09 |
| 41 | 16 | 31 | 32 | 30 | 39 | -09 | 34 | 12 | -06 | 04 | 22 | 29 | 35 | 36 | | 36 | 36 | 02 | 23 | -01 | 04 | -07 | -12 | -03 | -12 | -07 | -03 | -11 | 00 | -09 | -12 |
| 42 | 49 | 53 | 56 | 40 | 41 | -10 | 60 | 31 | 05 | 18 | 53 | 51 | 52 | 43 | 36 | | 57 | 08 | 43 | 06 | 13 | -05 | -10 | -07 | -07 | -09 | -07 | -02 | -01 | -02 | -12 |
| 43 | 35 | 74 | 48 | 66 | 57 | -09 | 73 | 31 | 09 | 22 | 45 | 76 | 83 | 64 | 36 | 57 | | 12 | 60 | 11 | 21 | -03 | -12 | -01 | -05 | -06 | -02 | -04 | -01 | -05 | -09 |
| 44 | 14 | 18 | 19 | 18 | 14 | 34 | 23 | 45 | 61 | 60 | 10 | 12 | 12 | 16 | 02 | 08 | 12 | | 32 | 69 | 87 | 16 | 17 | 13 | 14 | 10 | 07 | 15 | 10 | 13 | 23 |
| 45 | 39 | 62 | 50 | 58 | 42 | 12 | 67 | 58 | 37 | 53 | 45 | 61 | 61 | 58 | 23 | 43 | 60 | 32 | | 38 | 50 | 05 | 04 | 11 | 04 | 05 | 03 | 00 | -02 | 02 | 02 |
| 46 | 14 | 20 | 18 | 16 | 08 | 41 | 24 | 51 | 75 | 73 | 05 | 12 | 12 | 14 | -01 | 06 | 11 | 69 | 38 | | 91 | 76 | 26 | 22 | 20 | 21 | 10 | 16 | 15 | 18 | 29 |
| 47 | 20 | 29 | 27 | 26 | 18 | 40 | 34 | 55 | 75 | 76 | 17 | 21 | 22 | 24 | 04 | 13 | 21 | 87 | 50 | 91 | | 23 | 23 | 22 | 18 | 17 | 10 | 17 | 14 | 16 | 26 |
| 48 | 02 | 02 | -02 | 01 | -04 | 40 | 03 | 12 | 26 | 22 | -01 | 00 | 01 | -02 | -07 | -05 | -03 | 16 | 03 | 26 | 11 | | 40 | 45 | 43 | 43 | 26 | 25 | 24 | 21 | 32 |
| 49 | 02 | -01 | -10 | -04 | -13 | 44 | -03 | 14 | 21 | 20 | 00 | -07 | -09 | -04 | -12 | -10 | -12 | 17 | 04 | 26 | 23 | 40 | | 36 | 33 | 33 | 20 | 22 | 20 | 21 | 33 |
| 50 | 01 | 06 | -07 | 05 | 00 | 46 | 05 | 14 | [illegible] | 21 | -02 | 04 | 00 | 03 | -08 | -07 | -01 | 19 | 11 | 23 | 22 | 45 | 36 | | 61 | 51 | 23 | 16 | 23 | 17 | 32 |
| 51 | -03 | -01 | -07 | 00 | -05 | 41 | -01 | [illegible] | [illegible] | [illegible] | -07 | -04 | -03 | -02 | -12 | -07 | -05 | 14 | 04 | 20 | 18 | 43 | 33 | 61 | | 48 | 26 | 22 | 28 | 22 | 29 |
| 52 | -03 | -01 | -09 | 02 | 00 | 50 | 06 | [illegible] | 19 | 16 | -02 | -05 | -03 | -01 | -07 | -09 | -06 | 10 | 05 | 21 | 17 | 43 | 33 | 51 | 48 | | 36 | 17 | 15 | 19 | 16 |
| 53 | 01 | 01 | -06 | 07 | 04 | 45 | 04 | 05 | 12 | 09 | -03 | -01 | -01 | 03 | -03 | -07 | -02 | 07 | 03 | 10 | 10 | 26 | 20 | 23 | 26 | 36 | | 17 | 50 | 36 | 07 |
| 54 | 07 | -03 | -03 | -02 | -03 | 31 | 00 | 04 | 14 | 10 | 05 | -03 | -04 | -04 | -11 | -02 | -04 | 15 | 00 | 16 | 17 | 25 | 22 | 16 | 22 | 17 | 17 | | 30 | 34 | 16 |
| 55 | 11 | 02 | 02 | -01 | 02 | 46 | 06 | 03 | 13 | 08 | 07 | -02 | -02 | -05 | 00 | -01 | -07 | 10 | -02 | 15 | 14 | 24 | 20 | 23 | 28 | 15 | 50 | 30 | | 55 | 11 |
| 56 | 06 | -05 | -05 | -05 | -01 | 39 | -02 | 05 | 16 | 11 | 02 | -05 | -04 | -05 | -09 | -02 | -05 | 13 | 02 | 18 | 16 | 21 | 21 | 17 | 22 | 19 | 36 | 34 | 55 | | 10 |
| 57 | 00 | -04 | -08 | -07 | -10 | 36 | -03 | 11 | 23 | 20 | -01 | -05 | -07 | -09 | -12 | -17 | -09 | 23 | 02 | 29 | 26 | 32 | 33 | 32 | 29 | 16 | 07 | 16 | 11 | 10 | |

[1] Variables identified in table B.1.

# Index of Names

## M



## N



## O



## P



## R



## S

# Subject Index

## D



## E



## F

## G



## I



## J



## L



## M



## N

☆ U. S. GOVERNMENT PRINTING OFFICE: 1947—703323