AD 651 788

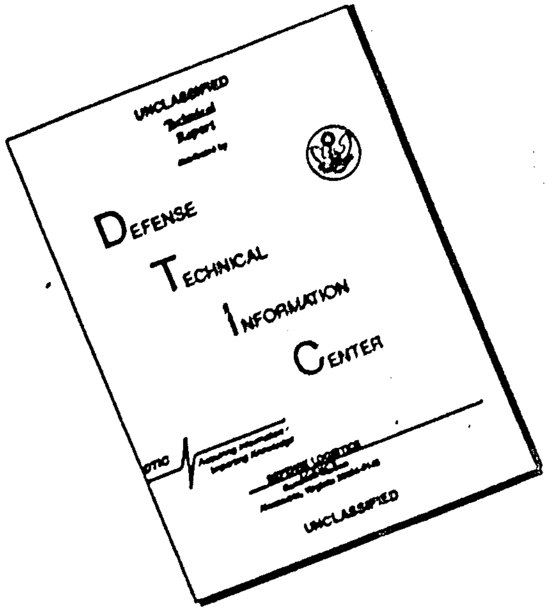
PSYCHOLOGICAL RESEARCH ON RADAR OBSERVER TRAINING

Stuart W. Cook

Army Air Forces
Washington, D.C.

1947

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

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.

Army Air Forces
Aviation Psychology Program
Research Reports

Psychological Research on Radar Observer Training

REPORT NO. 12

Edited by
STUART W. COOK
Research Director
Commission on Community Interrelations
of the American Jewish Congress
New York, N. Y.

1947

Preface

Operations of the Psychological Research Project (Radar) were carried out by a coordinated research group. From the point of view of recognizing individual contributions to this report, this fact has a number of implications. It is impossible, for example, to assign individual responsibility for critical stages of research planning which, for the most part, involved group consideration and decision. Even in project development, which was carried out by research teams, essential contributions of a critical nature must go unidentified.

With these qualifications, an effort has been made to footnote the names of those persons who carried central responsibility for various areas of the project's work. In addition, the writers of each chapter have been named, although, here again, critical contributions from the group notably altered original outlines and drafts.

Beyond this it did not seem wise to attempt in the text individual recognition for what was so effectively a cooperative and coordinated operation. On the other hand, brief note should be taken of certain individual services which were fundamental to the organization's work and to the preparation of the report. For example, Capt. H. Richard Van Saun and Sgt. Albert H. Hastorf did invaluable work in research coordination. Staff Sgt. Bernard C. Sullivan organized an extensive system of research records and supervised the Project's early statistical work. Staff Sgt. Roland E. Johnston carried these same responsibilities during a later period when extensive IBM analysis was being carried out. For a short but critical period the project profited from the statistical services of Lt. Sol M. Roshal and Capt. William F. Long.

Basic to all the research analysis carried out by the project was its research record system. Primarily instrumental in its development and maintenance were Cpl. Hyman Sofer, Sgt. Samuel D. Morford and Cpl. Arlene Babcock.

Graphic materials employed by the project in its research and used in this report were the work of Sgt. Alfred S. Arnott.

All development of physical measurement instruments and maintenance of testing apparatus was the primary responsibility of Technical Sgt. George M. Bollinger.

Principal credit for effective administration of the project's affairs during the major period of its work goes to Sgt. Harold I. Roth, Staff Sgt. Lester I. Foster, and Cpl. John D. Hennessy. During the emotionally trying stages of final editing, mimeographing, proofreading, and assembling this report, entire responsibility for administration was turned over to Capt. Gabriel D. Ofiesh. Coordination and administration of the project's field operations were carried out by Lt. Stuart Lottier and Lt. Lewis G. Carpenter, Jr.

In addition to the aid already mentioned the editor received invaluable assistance at other points in the preparation of the report. Lt. Stuart Lottier read all chapters from the point of view of improvement in style. A parallel review was made by Cpl. Harold H. Kelley from the point of view of technical content. The exacting task of reviewing the report for adherence to certain formal conventions established for the aviation psychology research series as a whole, was carried out jointly by Sgt. Ted P. Kisciras, Sgt. Harold I. Roth, and Capt. Gabriel D. Ofiesh.

Loyal and efficient secretarial service during the period of the project's research activities was rendered by Miss Mary Kingrea, Mrs. Laura Winter, Mrs. Rose Singer, Mrs. Virginia Van Saun, and Mrs. Mildred Flanagan. Mrs. Margaret Gage was invaluable in her role as research librarian. During the preparation of the mimeographed report the difficult clerical load was carried by Miss Nora Jenkins, Miss Phyllis Ashburn, and Mrs. Christine Glynn.

In addition to the full-time project personnel, Maj. B. von Haller Gilmer gave constant assistance to the research program in his supervisory capacity at Headquarters, AAF Training Command. Capt. Ike H. Harrison served as a valuable consultant on navigational and bombing problems.

Finally, attention should be called to the significant contribution of Col. William M. Garland. Such success as had been achieved in psychological research in the radar training program is due in no small measure to his vision and to his continued assistance and encouragement.

LANGLEY FIELD, VA.,
March 1, 1946

STUART W. COOK
Capt., A. C.

Contents

Chapter	Page

CHAPTER ONE

Scope of the Report[1]

This report describes the psychological research conducted in relation to the selection and training of radar observers (bombardment)[2] in the Army Air Forces. During the first 2 years of the war, operations of heavy bombers were conducted without radar[3] aids. Late in 1943, satisfactory airborne equipment became available in small quantities. A year later, production of radar sets had accelerated to such a point that it was necessary to initiate a large scale training program. The research activities described in this report took place largely in connection with this training program.

The report begins with an account of the use of radar as a device for blind bombing and navigation (ch. 2). The account includes a brief explanation of the basic principles on which radar operates, including the use of high frequency pulsating radio waves and the transmission and reception of these waves. A few of the major technical developments are noted, such as the early experiments with radar by naval scientists. The emphasis is placed on the development of airborne radar as a strategic and tactical weapon in World War II, starting with the first aircraft warning sets and culminating in the use of airborne radar in bombing and navigation in the European and Pacific Theaters. Finally, a survey is made of the development of radar observer training both in overseas installations and in those located in the continental United States.

Chapter 3 presents a chronology of research activities to serve as a framework for more detailed descriptions in later chapters. The major research undertakings of each of the three organizations which accomplished psychological research on the radar observer are enumerated and their interrelations pointed out. The survey begins with the research done by the Radar Project established in February 1943, at Camp Murphy, Fla., by the National Defense Research Committee

---

[1] Written by Sgt. Albert H. Hastorf.

[2] The full title of the aircrew specialist who operated radar as an aid to bombing and navigation in heavy bombardment aircraft of the Army Air Forces was radar observer (bombardment). For purposes of brevity, the shorter title, radar observer, is used throughout this report.

[3] A definition of this and other terms requiring explanation is included in a glossary to be found at the end of the report.

of the Office of Scientific Research and Development (Project SC-70, NS-146). Next presented are the research activities of the AAF Aircrew Evaluation and Research Detachment No. 1 on the selection of radar observers at the radar training station operated by the Eighth Air Force. Finally, there is given a description of the work of the Psychological Research Project (Radar). This project, officially established on 1 December 1944, concentrated its efforts on the psychological problems encountered by the AAF Training Command in selecting and training radar observers within the continental United States.

The major part of chapter 4 consists of a job description and analysis of the task of the radar observer based upon observations of ground and aerial training. A second section discusses job requirements in combat. Material for this discussion was obtained from personal interviews with combat experienced radar observers and from other reports from the combat theaters. The combat analysis differentiates the activities of the radar observer in the European and in the Pacific Theaters. Finally, there is a statement of probable requirements for the radar observer's task in the future; these predictions are based on current knowledge of the technical advances made in the development of new airborne radar equipment.

Chapters 5, 6, 7, and 8 deal with proficiency criteria developed for the radar observer training program by the Psychological Research Project (Radar). A battery of 11 printed proficiency tests and performance checks was used throughout training for evaluation of student proficiency. Chapter 5 discusses the proficiency tests, chapter 6 the performance checks, chapter 7 general problems in the measurement of performance, and chapter 8 the interrelationships between various proficiency measures.

The plan for proficiency test construction called for intermediate tests in each of the major curriculum divisions and one final comprehensive test. The tests described in chapter 5 represent each of these types of test. A brief account is given of the methodology of test construction, problems encountered in the standardization of test administration, and the difficulties in test construction resulting from an unstandardized curriculum. Among the descriptive materials will be found sample items, means, standard deviations, and, wherever available, reliability coefficients.

Chapter 6 presents a discussion of the three major groups of performance checks developed: The bench set trainer checks, the supersonic trainer checks, and the aerial checks. A statement is made of the rationale for the development of these checks. Their construction is described, including the selection of behavior to be evaluated and the development of items and of format. Sample items and statistical data are included. Also described in chapter 6 are the major steps in the training of examiners to administer the checks.

Chapter 7 discusses the validity and reliability of performance measures. A distinction is made between predictive validity and curriculum validity, which holds a measure to be valid if it furnishes a comprehensive test of achievement for a given area of instructional material. Types of criteria against which predictive validity may be determined are described. As a background for the discussion of reliability, performance check items are analyzed in terms of two components: The student's performance and the examiner evaluation of that performance. The reliability of different types of items is evaluated with reference to measurement errors associated with these two components. Statistical techniques appropriate to the measurement of performance check reliability are reviewed.

Chapter 8 analyzes the interrelations of the proficiency measures described in chapters 5 and 6. Among the matters discussed is the relation of proficiency test scores to scores on performance checks measuring the same skills. The findings are applied to the general problem of the relationship between verbal knowledge and actual job performance and the question of substituting printed proficiency tests for performance checks. Another section of the chapter is devoted to the interrelations of: (1) Performance checks which measure similar skills, and (2) proficiency tests which measure similar skills. Also, evidence is presented as to the degree of relationship between three areas of radar observer skill: Navigation, bombing and set operation. A comparison is made between the statistical findings reported in this chapter and parallel findings of AAF psychological research projects working on proficiency measurement in bombardier and navigator training.

Chapter 9 analyses the bombing error of radar observer students in training within the continental United States. The chapter includes a description of alternate methods for scoring the amount of bombing error. The reliability of the camera bombing circular error made by students at three training schools is presented. Data are presented also on the reliability of actual bomb drops; however, these are available from one training school only. Types of variable errors contributing to unreliability are discussed and suggestions are made for increasing reliability. Correlations are given between circular error and certain of the ground and aerial proficiency measures developed by the Psychological Research Project (Radar). Also discussed is the relationship between amount of practice and circular error. Finally, a constant error evident on most student bombing missions is analyzed at length.

Chapters 10 and 11 discuss selection research in radar observer training. Chapter 10 gives, first, an historical account of selection research prior to the establishment of the Psychological Research Project (Radar), and, second, an account of the methods through which the selection of radar students was actually accomplished. The chapter reviews the selection research accomplished by the Radar Project (SC–70, NS–146) of the National Defense Research Committee and the Air-crew Evaluation and Research Detachment No. 1 of the AAF Aviation Psychology Program. Tests developed by each of these groups are described. Results are reported for a validation study carried out by the latter organization.

Chapter 11 is concerned with selection test research conducted by the Psychological Research Project (Radar). Two validation studies were completed, each based upon a sample of bombardiers and a sample of navigators. The first study validates the Air-crew Classification Battery and the Radar Observer Selection Battery against a course grade determined by the training schools. The second study validates not only tests from these two batteries but also experimental psychomotor and printed tests. In the second study, the criterion for the bombardier sample was a course grade computed by the Project on the basis of standardized proficiency measures. The variables for the navigator sample were validated against course grades determined by the school and also against radar bombing error. In addition to validity coefficients, multiple correlation statistics are presented. Chapter 11 also includes a discussion of empirically-determined attributes of the successful radar student.

Chapter 12 evaluates the research accomplished to date and presents a prospectus for future investigation.

The report has one appendix and a glossary of technical terms; appendix A consists of descriptions of the selection tests validated by Psychological Research Project (Radar).

CHAPTER TWO

The Radar Observer in the Army Air Forces[1]

Radar was one of the outstanding technical developments of World War II. Used in detection and warning, aircraft interception, submarine hunting, bombing, navigation, fire control, and blind landing, it proved itself important in both offensive and defensive action. The purpose of this chapter is to describe the use of radar by the Army Air Forces as an aid to aerial bombing and navigation. The chapter will include an account of the training of the radar observer with emphasis upon the program within the AAF Training Command.

DEVELOPMENT OF RADAR

Radar is a contraction of the words *Radio Detection and Ranging.* Its basic principle is that of the echo, with high frequency radio waves substituted for sound waves. It is a familiar fact that a sharp noise or a shout near a cliff or high wall will be returned as an echo. The longer it takes the echo to return, the greater is the distance to the reflecting obstacle.

In radar, short pulses of radio energy, traveling at 186,000 miles per second, are sent out. If an echo returns, the radio pulse has reached some reflecting object. By measuring the lapse of time, the distance to the object can be found, and by determining the direction from which the echo returns, the bearing of the object can be determined.

Radar was an outgrowth of radio research conducted over a period of many years. The first step in its development came in 1922 when experimenters were working with high frequency transmitting and receiving equipment at the Naval Aircraft Radio Laboratories.[2] On one side of the Potomac River they had installed a transmitter and, on the other side, a receiver which converted the reflected radio energy to visual form on an oscilloscope screen. They noticed that a ship passing between the transmitter and the receiver interfered with re-

[1] Written by Sgt. Albert H. Hastorf.

[2] "Story of Radar" prepared under the direction of the Commanding General, AAF Eastern Technical Training Command, by Radio Publications Division, April 1945, p. 2.

ception. Realizing the possibilities of this discovery, they continued investigation of the phenomenon and found that, not only would a ship interfere with radio waves, but that it would reflect them as well. This suggested the desirability of placing the transmitter and receiver in the same place. By 1930, research had developed to the point where the presence of surface vessels hidden by fog, smoke, or darkness could be detected.

As research continued, methods of aircraft detection were developed. It was found that an aircraft passing between a transmitter and receiver also set up an interference pattern. By 1934, both the direction and distance of the aircraft could be determined accurately. In 1938, radar sets had been installed in some naval ships.[3]

AIRBORNE RADAR IN COMBAT OPERATIONS

The first important wartime application of airborne radar came in the air battle over Great Britain. In that battle, ground radar warning stations detected approaching enemy planes and informed fighter pilots by radio of the enemy's approximate position. The fighter pilots then proceeded to make contact visually. At night, however, or in the foggy weather often encountered over England, visual contact was difficult and often impossible. To remedy this situation the RAF turned to airborne radar.

RAF fighter aircraft were equipped with aircraft interception sets. Ground radar stations directed the fighters near enough to the enemy planes so that these short range sets could be used. It is reported that in one 24-hour period, radar-equipped fighters shot down 232 aircraft at a cost to themselves of only 40 aircraft and 12 pilots.[4]

The next major development in the use of airborne radar took place in the antisubmarine campaign. The fundamental type of search radar set was air-to-surface-vessel equipment, designated ASV. The earliest widely used set of this type was a long-wave set. This was soon replaced by more accurate microwave equipment. Another important development on recent ASV sets was that of the plan position indicator, or PPI scope, which presents a 360° picture of the area over which the aircraft is flying. With these sets, operators are able to pick up land targets, convoys, single ships and surfaced submarines at longer ranges than were previously possible.

Radar observer equipment consists of a search set similar to ASV, a precision ranging unit, and a bombing computer. With it, cities and other targets can be observed and bombed from high altitudes through a complete overcast. A city appears as a white patch on the scope at distances up to about 100 miles. As the aircraft approaches the city, the white patch takes the approximate shape of the city. Bomb-

[3] *Ibid.*, p. 2.
[4] *Ibid.*, p. 3.

ing can be accomplished in two ways. In coordinated bombing, the radar observer gives the bombardier information with which he synchronizes the bombsight to make an accurate release even though the target is not visible. In direct bombing, the radar observer makes the release independently of the bombardier, by use of the radar set alone.

Because the PPI scope presents a rough map of the area below the aircraft, the radar set is a very useful aid to navigation. From the scope, the radar observer can locate the position of his aircraft by determining its direction and distance from landmarks such as coastlines, lakes, rivers, mountains and cities. He is thus able to establish his position even though visual observation or radio contact is impossible.

The joint British and American seizure of the aerial offensive in the summer of 1943 gave airborne radar its first opportunity as an offensive weapon. Earliest use of radar for blind bombing and navigation was by pathfinder crews of the RAF. The AAF flew its first radar bombing mission in September of 1943, using British equipment. Improved equipment was soon developed at the Radiation Laboratory of the Massachusetts Institute of Technology, and mass production of radar sets was begun in the United States. A number of different sets were used by the AAF but, largely because of the highly secret nature of the sets, the official designations were rarely used, other terms being coined for radar equipment, observers, and operations. RAF terms include H2S and Stinky; AAF terms include H2X, Mickey, BTO for bombing through overcast, and Eagle for the AN/APQ-7 set.

Previous to October 1943, AAF crews had been trained on RAF radar equipment. In October, the first AN/APS-15 sets arrived from the United States for use by the Eighth Air Force. Four aircraft were equipped with these sets and crews were assigned to fly training flights over England and the North Sea. Eventually, the 482d Bomb Group of the Eight Air Force was selected as a central location for such aircraft. From this group, crews and aircraft were dispatched to fly with a different unit of the Eighth Air Force as pathfinder or lead crews. The job of a pathfinder crew was to lead a formation of bombers and make the bomb release which served as a signal for the entire formation to drop its bombs. Personnel of the 482d Group trained additional radar observers. As more radar equipped aircraft were received from the United States, the group began training personnel for permanent assignment to other units equipped with pathfinder aircraft. This decentralization made it possible for the radar crews to operate as more effective elements of the group to which they were assigned. Previously, efficient team work had been difficult because the pathfinders had led a different group on each mission.

703327—47——2

As the need for trained radar observers grew, the Eighth Air Force established a specialized training school with a formal radar observer course. The first class entered on 25 February 1944 and consisted of 42 navigators and bombardiers selected from combat groups in England. Each class received a month of ground and aerial training in radar navigation and bombing.

The inflow of students was increased so that by May of 1944, there were 75 students in a class. Approximately two-thirds of each class had received some radar training in the United States prior to embarking for England.

Shortly after the Eighth Air Force began extensive operations with radar, the Fifteenth Air Force in Italy procured its first radar observers. Combat-experienced navigators and bombardiers had been returned to the United States to receive radar training. With this personnel, the Fifteenth Air Force first began combat operations with radar and later activated its own training school for radar observers.

The use of airborne radar increased rapidly. In November 1943, radar equipment made it possible for the Eighth Air Force to fly more bombing missions than in any previous month. Throughout November and December, virtually all bombing missions were led by radar equipped pathfinder crews.[5] Gradually radar navigation and bombing techniques were refined and improved. On D-day, all heavy bombardment formations were led by radar aircraft and all coastline targets were bombed with the aid of radar.

In the Pacific Theater, in the early stages of the war, radar observers flew in B–24 bombers on sea-search missions against Japanese shipping. Two radar sets were used: The radar observer operated a search set such as the SCR–717 or AN/APS–15A, and the bombardier used a radar computer, called the AN/APQ–5. By coordinating the operation of these two sets, the radar observer and the bombardier were able to locate and bomb surface vessels with great accuracy.

The radar observer also played an important role in the bombardment of the Japanese home islands. During the last months of the war in the Pacific, all B–29 aircraft were equipped with either AN/APQ–13 or AN/APQ–7 sets. Early operation of these sets was unsatisfactory because of the lack of adequately trained personnel. This situation improved, however, early in 1945 as the expanded training program in the United States provided a greater number of qualified radar observers.

TRAINING OF RADAR OBSERVERS IN THE UNITED STATES

Until the fall of 1944, radar training in the continental United States was conducted on a relatively small-scale. Instruction in radar navigation and radar bombing was given regularly at only two schools:

---

[5] *Ibid.*, p. 7.

Langley Field, Va., and Boca Raton Army Air Field, Fla. At each of these schools, graduate bombardiers and navigators were given 4 weeks of radar observer training. The combined student flow was approximately 240 per month. At that time, Langley Field was an overseas replacement training center, and most of the graduates of both schools received additional training while awaiting transportation to the European Theater.

A third radar observer school was organized in April 1944 by the Second Air Force to train radar observers for B-29 operations in the Pacific Theater. This school was located at Smoky Hill Army Air Field, Salina, Kans., and trained classes of approximately 25 per month. The training consisted of both ground and aerial instruction in radar navigation and bombing. The curriculum stressed training in off-set bombing, a technique for bombing an obscured target by making computations with reference to a visible radar return located a known distance and direction from the target. Training continued until August 1944, when the school was disbanded and the instructional staff was transferred to the radar observer schools at Boca Raton and Langley Field.

In the fall of 1944, after almost a year of increasing success with radar navigation and bombing, it was decided to expand the training program in the United States. Because of a relative excess of pilots and shortage of bombardiers and navigators, a decision was made to attempt to train the former as radar observers. The training course was increased to 16 weeks and included instruction in non-radar navigation and bombing. After about 1 month, the 16-week course was abandoned because the motivation of pilots for such training was low. The selection of prospective students from bombardiers and navigators was resumed and the duration of the course was fixed at 10 weeks. At the time of entering radar observer training, bombardiers were required to have had navigation training in advanced bombardier school and navigators were required to take 4 weeks of preradar training in bombing.

In early 1945, a new radar observer school was established at Victorville Army Air-Field, Calif. The combined student flow of the three stations was soon increased to 500 students per month. With this rapid growth came many problems of curriculum standardization. These problems arose in part because the different training stations were equipped with different radar sets. The AN/APS-15 and the AN/APS-15A were used at Langley Field while the AN/APQ-13 was used at Boca Raton and at Victorville. However, most problems of curriculum standardization arose between schools using the same set, as to which operating procedures were best and how they should be taught. The question of which techniques were most appropriate to the different theaters of operation was particularly troublesome.

The demand for radar observers continued to increase. In the spring of 1945, a fourth radar observer school was activated at Williams Army Air Field, Ariz., and in the early summer of 1945, a fifth school was established at Yuma Army Air Field, Ariz. The combined monthly flow for the 5 schools was 1,000 students. Very few students graduated from either of the new schools, however, since, with the end of the war, radar training was immediately curtailed.

SUMMARY

This chapter gives a brief description of the training of the radar observer and his place in the war-time operations of the Army Air Force. It is divided into three sections. The fundamental principle of radar, it is pointed out in the first section, is similar to that of the echo. Short pulses of radio energy are sent out and a receiving unit presents in visual form the direction and distance of the object which reflects the pulses.

The second section is devoted to the use of airborne radar in combat operations. This account begins with the early uses of ground radar in detecting aircraft and the installation of the first airborne sets in fighter aircraft. Search sets were installed in aircraft as a part of the anti-submarine campaign; these sets were the forerunners to the sets operated by the radar observer. The RAF was the first air force to use radar as an aid to navigation and bombing. In time the Eighth Air Force had radar-equipped pathfinder aircraft and crews operating on many of its bombing missions. A school was also set up by this Air Force to train radar observers. Air forces operating in the Mediterranean and Pacific Theaters made increased use of radar aids as the war continued.

The final section of the chapter outlines the training program for radar observers set up in the United States. As combat operations increased it was necessary to expand training facilities in this country. Within a period of months the number of training stations was increased from 2 to 5 and the monthly student flow from 250 to 1,000.

CHAPTER THREE

Survey of Research[1]

The purpose of this chapter is to review, in chronological order, the major research studies accomplished in the selection and training of radar observers. The presentation is in the form of a survey; detailed accounts of specific research accomplishments may be found in subsequent chapters.

The studies to be mentioned will be introduced in relation to the development of three more or less independent research organizations. The three organizations, in the order in which they will be presented, are first, the National Defense Research Committee Project SC–70, NS–146, referred to as the NDRC Project; second, the AAF Aviation Psychology Program, Aircrew Evaluation and Research Detachment No. 1, the Eighth Air Force, to be referred to as the AERD No. 1, and third, the AAF Aviation Psychology Program, Psychological Research Project (Radar). In addition, mention will be made of the selection of radar observers by Psychological Research Project (Navigator) and Headquarters, AAF Training Command.

NATIONAL DEFENSE RESEARCH COMMITTEE—PROJECT SC–70, NS–146

The broad task assigned the NDRC Project upon its initiation in February 1943, was research upon psychological problems of radar operation in both the Army and Navy.[2] In carrying out this assignment personnel of the project conducted numerous investigations, some related primarily to ground radar installations, others primarily to sets employed by naval vessels or aircraft. In this review, only research bearing upon the problems of airborne radar will be mentioned.

Among the airborne radar assignments undertaken by the NDRC Project, one of the first was a job analysis of the operation of equipment designated as ASV (air-to-surface-vessel) a type of airborne

[1] Written by Sgt. Albert H. Hastorf.

[2] Applied Psychology Panel, NDRC, Final report in summary of work on the selection and training of radar operators, Research Report No. 19, 24 September 1945. The following personnel participated in the Project; Donald B. Lindsley, director, Irving H. Anderson, Alfred L. Baldwin, Charles H. Bridgman, Robert S. Daniel, John G. Darley, Robert H. Dreher, Edward P. Horne, Edward A. Jerome, William H. Lichte, Thomas L. McCulloch, Fred McKinney, Karl U. Smith, G. Raymond Stone, Edward J. Sweeney, Garth J. Thomas.

radar sea search equipment. ASV used either the SCR 717A or SCR 717B sets which, like the sets used by the radar observer, have a PPI scope and consequently require similar operational skills and abilities. Following this analysis personnel of the project prepared 15 selection tests dealing principally with perceptual aspects of the operator's task. These tests were believed to measure speed and accuracy of perceptual discrimination, alertness, persistence, and the ability to make quick judgments. To establish time limits and scoring methods, the tests were administered experimentally at Camp Murphy, Fla.; Boca Ration Field, Fla; Langley Field, Va.; and at 8 stations under the AAF Tactical Center, Orlando, Fla.

Several tests from this group of 15 were administered by AERD No. 1 as part of a selection research project to be described below.[3] Those found by AERD No. 1 to be most predictive of success in radar observer training were included later in a test battery administered to potential radar observer students in the United States by testing teams from Headquarters, AAF Training Command. Other tests of this group were used by Psychological Research Project (Navigator) during the period when that organization was responsible for radar observer student selection. Psychological Research Project (Radar) prepared machine scoreable forms of others for inclusion in a battery of experimental selection tests.

The NDRC Project also contributed to radar observer research in the areas of proficiency measurement and of training methods. Two comprehensive printed proficiency tests were constructed, one for the AN/APS-15 set, and another for the AN/APQ-7 set. A film trainer was developed which used motion pictures of scope returns as briefing and reconnaissance aids. A manual was prepared to aid students in training as radar observers for low altitude radar bombing. A photo bomb scoring computer was constructed. One of the project's most extensive studies was an extended training experiment designed to ascertain the effect on proficiency of continued training beyond the normal duration of the AAF radar observer training course.

AIRCREW EVALUATION AND RESEARCH DETACHMENT NO. 1[4]

AERD No. 1 consisted of 6 officers and 15 enlisted men of the AAF Aviation Psychology Program who were detached from May to August 1944, and assigned to duty with the Eighth Air Force in England. At this stage in the development of the air war the increasing use of radar as an aid to nevigation and bombing had emphasized the importance of choosing well-qualified personnel for training as radar

[3] Tepley, W. M., ed., *Psychological research in the theaters of war.* AAF aviation psychology program research reports, No. 17. Washington, Government Printing Office, 1947.
[4] *Ibid.*

observers. As a result, AERD No. 1 assigned 6 of its members to specialized research in the selection of radar observer students.[5] A job analysis was made of the particular skills and abilities required by the airborne radar observer in training at the Eighth Air Force Radar Observer School. Following this analysis 20 printed selection tests were chosen for validation against course grades at the school. The tests validated include several of those developed by the NDRC Project, several originated by AERD No. 1, and a large group from the AAF Air-crew Classification Battery. Also validated were stanines for bombardier, navigator and pilot. Of the 20 tests, the 4 most valid were later to constitute the radar observer selection battery.

Another major undertaking of AERD No. 1 was the construction of a comprehensive printed proficiency test for radar observers. Included in this test was a section in which the student's task was to navigate through a simulated radar mission. Navigation was carried out with the aid of a full-size plotting chart of northwestern Germany and was dependent upon the correct interpretation of a series of photographs of the radar scope.

SELECTION OF RADAR OBSERVERS

As the importance of the radar observer in the European Theater increased, the desirability of screening potential students in the United States became apparent. In July 1944, the Psychological Research Project (Navigator), at the direction of Headquarters, AAF, assembled a selection battery to be administered to potential radar students at the advanced navigation schools. Three tests developed by the NDRC Project and a preference blank were chosen. The battery was administered to advanced students who either had navigator stanines of eight and above or ranked in the upper third of their class. The first administration took place in July 1944, at Hondo Army Air Field, Tex.; later administrations were carried out at three other navigation schools. Routine testing for screening purposes was conducted by personnel of these schools and continued until November 1944.

On 9 November 1944, Headquarters AAF Training Command established airborne radar observer selection teams for the purpose of administering a selection battery based upon the validation studies of AERD No. 1. In May 1945, the Navigation Proficiency Test, described in chapter 5, was added to this battery for selecting students in advanced bombardier training. The airborne teams continued selection testing until July 1945.

---

[5] Major B. von H. Gilmer, Captain Stuart W. Cook, Lt. William M. Wheeler, T/Sgt. Russell W. Bornemeier, T/Sgt. Robert B. Miller, Sgt. Philip H. Kreidt.

In the fall of 1944, after a year of increased success in strategic bombing with radar, the AAF embarked upon a greatly expanded program for the training of radar observers. The training program, until that time, had been concentrated at 2 training stations. Plans were made to increase the number of stations to 3, and later to 5, and to expand the total student flow from 250 to 1,000 students per month.[6] These developments made desirable the establishment of an AAF psychological project for specialized radar observer research.

Activation of the Project

Early in September 1944, Lt. Col. A. Paul Horst, Commanding Officer of AERD No. 1, accompanied by Maj. Beverley von H. Gilmer and Capt. Stuart W. Cook, reported to the Air Surgeon, Washington, to discuss the preliminary findings and plans for completion of the work undertaken by that detachment while with the Eighth Air Force. At this time the first conferences were held relative to the need for a psychological project in the selection and training of radar observers. Later this need was discussed with the Surgeon at Headquarters, AAF Training Command, Fort Worth, Tex., and with Col. William M. Garland, then deputy for training and operations at Langley Field, Va. It was decided at these discussions to activate a project under the direction of Captain Cook and, because of the urgency of the work, to assign personnel to Langley Field on temporary duty prior to official activation.

Arrangements were made, also, to enlist the assistance of established psychological units and projects. Two officers from the Psychology Department, School of Aviation Medicine, Randolph Field, Tex., were assigned on temporary duty to assist in the planning for experimentation with apparatus tests. The Director of Psychological Research Project (Bombardier) assisted with plans for proficiency measurement. An officer from the Psychological Research Unit, San Antonio, collaborated in the preparation of plans for experimentation with printed selection tests. By October 1944, a group of 9 officers and 15 enlisted men had been assembled at Langley Field. Working quarters were established for this group on the flight line where other flying and training activities were concentrated.

On 1 December 1944, the Psychological Research Project (Radar) was officially activated by Headquarters, AAF Training Command.[7]

---

[6] The first two schools were situated at Langley Field and at Boca Raton AAF. The former was responsible for AN/APS-15 and AN/APS-15A training while the latter taught AN/APQ-13 and AN/APQ-7. A third school, Victorville AAF was established for AN/APQ-13 training. It was supplemented later by Yuma AAF and Williams Field. The latter undertook all AN/APQ-7 training.

[7] Letter, Headquarters AAF Training Command, to Commanding General, AAF Eastern Technical Training Command, Subject: Establishment of Psychological Research Project (Radar), 25 November 1944, File 353 Radar.

The majority of personnel stationed at Langley Field on temporary duty were then permanently assigned.

Research Objectives and Priorities

The letter of activation states the mission of the project as follows: *a.* The development of aptitude tests for the selection of radar operators. *b.* The development of radar operator proficiency criteria against which to validate aptitude tests. *c.* The investigation of conditions of optimally efficient use of trainers and training methods. *d.* Conduct of research studies on other psychological problems to be directed by this headquarters.[8]

Priorities were assigned to research objectives on the basis of practical circumstances, under which the project began its work. First and highest priority was given to the development of proficiency tests and checks. Second priority was assigned to the validation of selection tests. Third priority was placed upon instructor selection and evaluation. Lowest priority was given to the investigation of trainers and training methods.

Several important considerations made the development of proficiency measures most urgent. The training program was new and there was an acutely felt need for acceptable methods of evaluating students. Supervisory training personnel under the leadership of Colonel Garland, were dissatisfied with the methods that had been hurriedly improvised and were receptive, consequently, to proposals regarding new types of proficiency measurement. In addition, it appeared probable that, until acceptable proficiency measures were constructed, there would be available no adequate criterion against which to validate selection tests. Of the possible criteria, one, the pass-fail criterion, was eliminated because the demand by the operational air forces for radar observers did not allow the failure of inferior students. Another, instructor grades, appeared likely to be of doubtful value because the rapid expansion of training necessitated the use of many instructors with no previous teaching experience and others with little or no motivation to teach. A third, bombing accuracy, was made impractical by the lack of sufficient photographic equipment at the radar training stations.

Work on the validation of selection tests, assigned second priority because of the necessity for immediate development of proficiency measures, was only slightly delayed by this emphasis. While it was not possible to begin the development of new selection tests, early attention was given to the assembly and administration of a battery of selection tests available from other sources. It was expected that development of new tests would not be long delayed since plans called for a standard training program in which identical proficiency

[8] *Ibid.*, paragraph 2.

measures would be used at all training stations. The fact that such standardization was not achieved multiplied greatly the work in proficiency measurement and unduly delayed selection test development.

Research on instructor selection and evaluation, while assigned third priority in the project's emphasis, was greatly needed and required a relatively small investment of personnel. Validation of the instructor selection tests was never accomplished because of the delay in accumulating suitable criteria of instructor proficiency. Lowest priority was given to research projects in training. While significant training problems were present, the project followed the policy of the Aviation Psychology Program in emphasizing research in personnel selection and evaluation.

Survey of Project Research

One of the first tasks of the project was to become adequately oriented to the radar observer's tasks and to the technical details of various types of radar equipment. Lectures, demonstrations, observations and orientation flights were specially arranged for this purpose. Members of the project were successively enrolled in the course at Langley Field throughout the duration of the training program. In this course project personnel participated in all ground and flying training as regular students, and accumulated sufficient numbers of hours of flying time to solo as radar observers.

The first 2 months of the project's activity were directed primarily toward the completion of a battery of standardized proficiency measures for the various types of radar equipment used. Early in January 1945, a battery of three proficiency tests and three performance checks, applicable to both AN/APS-15 and AN/APQ-13 equipment, was presented at the conference of the Radar Standardization and Advisory Board. Following the presentation, the board adopted a recommendation "that the phase checks and examinations which are set up by the Psychological Research Project (Radar) be regarded as the only examinations to be given and that no other special examinations be administered." Informal invitations were issued at that time by the Deputies for Training and Operations at all radar training stations to install proficiency measures at the earliest opportunity.

During January, the initial battery of proficiency tests and performance checks was further developed and refined. This work entailed conferences with training personnel at Langley Field, participation in training flights, and research with the AN/APQ-13 set at the Boca Raton radar observer school. Detailed memoranda were prepared dealing with procedures for administering, scoring, and safeguarding tests and performance checks. An experimental battery of six apparatus selection tests was installed at Langley Field under the supervision representative of the Department of Psychology, School of A ation Medicine. Plans were formulated for the first experimental battery of printed selection tests and for the validation of tests previously administered at air-crew classification centers and by the airborne radar observer selection teams.

Early in February 1945, a new phase of the project's work began. Administration of a battery of experimental selection tests was initiated at the three radar training stations then in operation and testing with the apparatus test battery began at Langely Field.[9] Simultaneously, the introduction of standardized proficiency tests and performance checks at the three radar observer schools was attempted. At two of the schools, the installation proceeded essentially according to plan and with considerable success. At the third, the battery was inapplicable because of unanticipated differences in curriculum and in operating procedure. Many similar incidents to follow introduced the project to problems which were never satisfactorily solved. Soon, for example the nature of aerial training at another of the schools was radically changed. The activation of additional radar training stations with new local variations in training techniques and proficiency requirements presented still other complications.

To these curricular and administrative problems were added questions concerning the manner in which performance checks were being administered. It soon became evident that checks administered by instructors could not be standardized and that a small group of trained and specialized examiners was necessary.

In the period between the development of these difficulties and the end of the war, the project's work involved concurrent attacks on all phases of its research assignment. The difficult task was undertaken of preparing, revising, and duplicating the extensive proficiency battery to fit each more or less unique local curriculum and operating procedure. Considerable effort by project personnel resulted in the establishment of specialized examiners at all training stations. A job description and job analysis were completed, plans for a second battery of experimental selection tests were formulated, administration of experimental selection tests was continued, an instructor evaluation program was installed, and refinements were made in the methodology of constructing proficiency tests and performance checks.

Following V-J Day existing research projects were terminated and project personnel on temporary duty at the various radar training

[9] Project personnel responsible at different times for the administration of these tests at the respective stations were: for Langley Field: Sgt. Hyman Heller; Cpl. James R. Holt; Cpl. James C. McClure; Cpl. Robert J. Patterson; Pfc. Gordon L. Puller, Jr.; Pvt. Donald C. Bennett; and Pvt. Martin S. Maltenfort; for Boca Raton AAF; Sgt. Norman Graff; Cpl. Nelson R. Nall; Pfc. Jack L. McCollom; and Pfc. Gordon L. Puller, Jr.; for Victorville AAF; Sgt. Michael Green; Cpl. Wilbert H. Schwotzer; for Yuma AAF: Cpl. Wilbert H. Schwotzer; and for Williams Field; Sgt. Gerald S. Blum and Cpl. Douglas W. Bray.

stations returned to Langley Field. The project then directed all activities towards completing the task of analyzing the data which had been obtained.

Personnel

In the following roster are listed personnel permanently assigned to the project. The names of persons associated with the project throughout the major portion of its research history are preceded by an asterisk (*).

OFFICER PERSONNEL

*Capt. Stuart W. Cook (*Director*)
Capt. Ike H. Harrison
Capt. William F. Long
Capt. Gabriel D. Ofiesh
*Capt. H. Richard Van Saun (*Assistant Director*)
*Lt. Lewis G. Carpenter, Jr.
*Lt. George S. Klein
*Lt. Stuart Lottier
Lt. Sol. M. Roshal.

ENLISTED PERSONNEL

*Technical Sgt. Geo. N. Bollinger
Technical Sgt. Sanford J. Mock
Technical Sgt. Hyman Schmierer
Staff Sgt. Trent E. Bessent
*Staff Sgt. Lester I. Foster
Staff Sgt. Roland E. Johnston, Jr.
Staff Sgt. Harold F. Kunsman
*Staff Sgt. Richard T. Mitchell
*Staff Sgt. Bernard C. Sullivan
Staff Sgt. William J. Woywod
*Sgt. Alfred S. Arnott
*Sgt. Gerald S. Blum
*Sgt. Stanley Blumberg
*Sgt. Nathaniel L. Gage
*Sgt. Norman Graff
*Sgt. Michael Green
*Sgt. John S. Harding [10]
*Sgt. Albert H. Hastorf
*Sgt. Hyman Heller
Sgt. Ted P. Kisciras
*Sgt. Philip H. Kriedt
*Sgt. William J. Mangan
*Sgt. Samuel D. Morford
*Sgt. Sheldon H. Nerby
Sgt. Albert Pepitone
Sgt. Harold L. Raush
*Sgt. Harold I. Roth
Cpl. Alfred D. Antilla
*Cpl. Arlene E. Babcock
*Cpl. Douglas W. Bray
Cpl. Irving Fudeman
*Cpl. John V. Hennessy
*Cpl. James R. Holt
*Cpl. Harold H. Kelley
*Cpl. Robert H. Koch
Cpl. Mary E. Loomis
*Cpl. John F. MacNaughton [10]
*Cpl. James C. McClure, Jr.
Cpl. Owen R. Munger
Cpl. Nelson R. Nail
*Cpl. Robert J. Patterson
*Cpl. Wilbert H. Schwotzer
Cpl. Vernon S. Scott
Cpl. James E. Skowronski
*Cpl. Hyman Sofer
Pfc. James H. Anderson
*Pfc. Donald G. Bennett
Pfc. Robert J. Blount
Pfc. Dwane R. Collins [10]

[10] Commissioned as 2d lieutenant, A. G. D. on leaving the project.

Pfc. Billy S. Elliot
Pfc. Sanford Goldstone
Pfc. Edmund W. Lyons
*Pfc. Jack L. McCollom
Pfc. Kenneth M. Mitchell
Pfc. Thomas B. Morgan
Pfc. Roger L. Murrel
Pfc. Gordon L. Puller, Jr.
Pfc. Hyman Rogosin
*Pvt. Martin S. Maltenfort

SUMMARY

This chapter has reviewed the research accomplished by this and other organizations in the selection and training of radar observers. The purpose of the survey has been to present a chronology of research activities to serve as a framework for the more detailed descriptions of the research accomplishments found in the subsequent chapters.

The survey is organized primarily around the development of three more or less independent research organizations. It begins with the work of Propect SC–70, NS–146, of the National Defense Research Committee in February of 1943. Although this project performed research on both air-borne and ground radar for the Army and the Navy, this chapter reviews only its research on the selection and training of air-borne operators. Following a job analysis of the operation of the air-borne radar equipment used in sea search, the project developed 15 selection tests dealing principally with the perceptual aspects of the operator's task. Among its contributions to research in the areas of proficiency measurement and training methods, were the construction of two comprehensive printed proficiency tests for radar observers and the conduct of an experiment on the effect of extended training.

The Air-crew Evaluation and Research Detachment No. 1, the second organization discussed, conducted its research at a radar observer school in the Eighth Air Force. Here a job analysis was made, and 20 printed selection tests were validated against course grades. The four most valid of these tests were later to constitute the radar observer selection battery, administered to all prospective radar observer students by testing teams under the direction of Headquarters AAF Training Command. Prior to the use of this battery, Psychological Research Project (Navigator) had administered 3 NDRC Project selection tests to prospective students.

In the fall of 1944 the Psychological Research Project (Radar) was activated. Its mission was to perform research on the psychological problems encountered in the training of radar observers in the AAF Training Command. Early emphasis was placed by the project on the development of a comprehensive battery of proficiency tests and checks. Although school differences in curriculum and operating procedure required numerous revisions of these measures, they were

eventually placed in general use at all training stations. The project conducted a selection research study in which all entering radar observer students were administered a lengthy battery of experimental printed tests. In addition, experimental apparatus tests were administered at one school. These tests along with scores from the air-crew classification and radar observer selection batteries were validated against course grades.

CHAPTER 4

Job Description and Analysis[1]

GENERAL CONSIDERATIONS

This chapter describes and analyzes the radar observer's job for the purposes of psychological research. It includes the job in training, the job in combat, and indications of probable future trends.

Reason for Describing a Job

Some description of a job forms the basis of all psychological research upon that job. More often than not this important first step has been carried out informally or not at all. Nevertheless, some definition of the job has been the foundation of every job research program even if it was only implied in the thinking of the research psychologist.

A systematic job definition can serve several highly important functions. First, it can describe the specific abilities or aptitudes required by the job. Aided by this description, psychologists may choose or construct selection tests which appear most likely to measure these particular abilities and aptitudes. Second, a job definition can describe various levels of proficiency on the job. Such descriptions are useful in developing measures of proficiency. Third, a job definition can provide a summary of the tasks and skills which comprise the job for the purpose of setting up the most efficient program of training personnel for the job. Fourth, a job definition can describe the actual techniques of doing the job and, where equipment is involved, show how the equipment is actually being used. This information often yields clues for improving techniques of using existing equipment, improving the equipment itself, or better adapting it to the abilities and characteristics of the average individual on the job.

Methods of Describing a Job

A job can be described using concepts at various levels of analysis. At one extreme is a description of what the individual does on the job in simple, nontechnical language. Such as description can be written by a person unfamiliar with psychological concepts. This

[1] Written by Cpl. Harold Kelley.

type of description, to be referred to hereafter as a "job description," employs terms which are for the most part specific to the job and not transferable from one job to another. At the other extreme is the type of description which will be referred to here as a "job analysis." A job analysis is a description of the job in terms of psychological functions or factors. These functions or factors represent abilities and skills which are found introspectively or statistically to be common to many jobs.

Whether a job description or job analysis is to be preferred in a particular research program depends upon which of the four purposes mentioned above is paramount to the job research. Job proficiency is most validly measured in terms of the specific tasks, skills, and contents of the job. Navigation proficiency, for example, is measured more validly by navigation problems than by general mathematics problems. Consequently, a job description provides the most adequate foundation for construction of proficiency measures. Training research is also based preferably upon a job description. Only if a job analyst believes in a great deal of transfer of training, will he base a training program upon a factor definition of the job and train individuals in general functions rather than in specific tasks.

In addition, a job description is more useful than a job analysis in furnishing the basis for research into techniques and equipment. However, even in such research, the job analysis is not completely without use. It is probable that an identification of pertinent factors, coupled with information as to the relative levels of ability in these factors possessed by the available population, could simplify the problem of adapting equipment to the abilities of the job performers.

A job description, job analysis, or definition at any level of generality between these extremes can be used as the basis for selection test research. The choice depends upon the extent to which factor theory is accepted. Factor theory rests in part on the hypothesis that performance on most of the tasks in contemporary technology can be explained by a limited number of independent functions or factors taken in various combinations and amounts. If this position is taken, a job analysis is made of the job and research proceeds with already available or newly constructed factor tests. Such tests usually have little face validity, that is, they have little specific content in common with the job. The basis of their construction is that they require tasks and operations which seem to have something fundamentally in common with tasks carried out on the job. In other words, they require use of one of the same psychological functions.

Rejection of factor theory leads to writing a job description and the development of job analogy tests. Such tests are construed so as to resemble the job as closely as possible. In actual practice, of course, most test construction falls somewhere between these two extremes.

The advantages and disadvantages of these two approaches to selection test research are as follows:

First, factor tests, taken singly, tend to have low validity while job analogy tests have relatively high validity.

Second, factor tests have low intercorrelations, while job analogy tests have high intercorrelations. Consequently, a combination of a number of factor tests may yield a composite score of useful validity even though the individual tests have a low validity. On the other hand, the combining of a number of job analogy tests adds little to the accuracy of prediction obtained by any one test.

A third consideration is that a number of factor tests have been constructed and refined and are immediately available for use in selecting trainees for jobs which require the corresponding abilities. Job analogy tests have to be specially constructed for every unique job. On the other hand, the factor approach is limited by the fact that for some years the number of factor tests available will be inadequate to the requirements of complex jobs. It is anticipated by factor theorists that eventually tests will be constructed to measure all the important factors, thus making it possible to set up a selection battery for any job by merely assembling the pertinent tests as indicated by a job analysis.

Fourth, part of the factor content of jobs which require operation of complex equipment can be measured with inexpensive printed tests. On the other hand, job analogy tests for such jobs require construction and maintenance of expensive testing equipment.

A final consideration is the theoretical objection to breaking a job into elements (factors) for the purpose of selection test construction. The contention is that measuring elementary, independent functions and additively combining the several scores yields something quite different than does measuring skill on the total task.

One method of evaluating this theory consists of a comparison of job analogy test validity with the composite validity of factor tests. If the theory is correct, a test adequately sampling the job should yield significantly higher validity than that obtained from the pertinent factor tests. Such an evaluation can be invalid either because the job analogy test does not adequately sample the job or because some of the factors involved in the job are not yet known.

Another method of evaluation consists of predicting the validities of job analogy tests (which are factorially complex) from the validities of the factors they have been found to contain. These predicted validities should fall short of the obtained validitites if the job analogy tests measure something more than merely the sum of their elements. Predictions made on this basis have usually been close to the obtained validities of the job analogy tests but have consistently been underestimates. Our inadequate knowledge of the factorial content of such tests limits the conclusiveness of this evaluation. As mentioned before, the number of factorially pure tests is inadequate and there is no doubt that many important factors have not been identified. Consequently, the factorial content of a relatively simple job analogy test, not to mention a criterion job, cannot be thoroughly defined.

Basis of Present Definition of the Radar Observer's Job

The delineation of the radar observer's job which follows consists of a job description paralleled by a job analysis. Both the description and the analysis were prepared to serve as a basis for selecting and constructing tests to measure aptitude for the job of radar observer. It was recognized that the selection research program could begin with a job description and apply job analogy tests, or it could begin with a job analysis and apply factor tests. Because of the theoretical interest in the relative merits of the two approaches, it was decided to follow both. The selection test program which resulted is described in chapter 11. The present chapter is concerned only with the application of the two methods of describing a job.

In writing the job analysis, the problem arose of choosing a psychological terminology. Where practicable, it was decided to make use of factor analysis studies and findings accomplished in the Aviation Psychology Program. These analyses, which were based on both aircrew classification tests and experimental tests, had produced unambiguous empirical evidence of the existence of a number of factors and yielded tentative evidence for others. The test batteries analyzed have not included all types of measures and have been deficient especially in personality and motivational tests. Consequently, it was believed unwise to depend wholly upon factors isolated from such batteries in attempting to explain the radar observer's job. When these factors appeared to be inadequate to explain a certain ability, the analysis was made in terms of hypothetical variables. In these instances, the attempt was made to describe hypothetical variables which were "testable" and independent.

In order to provide an understanding of the functions and abilities represented by the factor names used in the job analysis sections to follow, some of the factors that have been definitely or tentatively isolated in the Aviation Psychology Program are listed and defined below. Hypothetical factors will be defined in the text of the analysis when they are first mentioned. The tests referred to in the following factor definitions are described in appendix A.

1. *Length estimation*, a poorly defined factor, refers to the ability to estimate lengths without the aid of measuring devices. This factor is thought to be measured by Estimation of Length, CP631A.

2. *Mechanical experience* represents practical knowledge of mechanical devices such as might have been gained through experience with

ther. This well-defined factor is best measured by Mechanical Information, CI905B, which has a loading of about 0.75 and Mechanical Principles, CI903B, with a loading of approximately 0.60.[2]

3. *Memory I* (paired associates memory or rote memory), a poorly defined factor, is thought to be measured by Memory for Landmarks, CI510AX2. An almost identical form, CI510AX1, has a loading of 0.60 in Memory I.

4. *Memory II*, or visual memory, describes the ability to recognize previously seen patterns. Although there is little evidence for this factor, Visual Memory, CI514A, is thought to measure it adequately. A roughly similar test, Map Memory, CI505AX2, has a loading of 0.60 in Memory II.

5. *Numerical facility* describes the ability involved in carrying out simple arithmetic computations. It is also involved to a considerable extent in simply locating and observing numbers. The purest measure of this well-defined factor is believed to be Numerical Operations, CI702BX1. An older form, Numerical Operations, CI702A and B, has a loading of about 0.80. Mathematics A, CI702F, and Mathematics B, CI206C, have loadings of approximately 0.50.

6. The *perceptual speed*, or identification factor, is the ability to note quickly and discriminate details in visual patterns. Speed of Identification, CP610C, is probably the purest test of this factor with an estimated loading of 0.65. An older form, CP610B, using aircraft silhouettes as subject matter, has a loading of approximately 0.65. Spatial Orientation I, CP501B, has a loading of approximately 0.60.

7. *Pilot interest* refers to interest in aviation such as might have been gained through contact with it as a hobby, through model construction, or reading. General Information, CE505E, is the best measure, having a loading of approximately 0.40.

8. *Psychomotor coordination*, the only factor found to be characteristic of psychomotor tests alone, deals with gross motor coordination. It is best defined by Complex Coordination, CM701A, with a loading of about 0.40 and by Rotary Pursuit, CP410B. Form CM703A of Rotary Pursuit, which did not require divided attention, has a loading of 0.55 in this factor.

9. *Psychomotor precision* is best defined by Discrimination Reaction Time, CP611D, and Finger Dexterity, CM116A.

10. *Reasoning I*, or general reasoning, is one of the three reasoning factors, the others being too inadequately defined for use in the present analysis. Spatial Reasoning, CI211BX2, is thought to be an adequate measure of this factor with a loading of approximately 0.55. Mathematics B, CI206C, has a loading of about 0.50 but, in addition, has a similar loading in numerical facility factor.

---

[2] The factor loadings presented in this section are taken primarily from analyses of the July 1943 and November 1943 air-crew classification batteries.

11. *Space I*, or spatial relations, refers to the ability to move one's self mentally in space and predict the result of such movement in terms of position, view of terrain, etc. It is best defined by Instrument Comprehension II, CI616B, with a loading of approximately 0.50. Aerial Orientation, CP520A, and Flight Orientation, CP528A, are thought to have high loadings on this factor although this has not been tested in analyses. Discrimination Reaction Time, CP611D, and Complex Coordination, CM701A, have loadings near 0.40 and 0.50, respectively.

12. *Space II*, or rotational space, identified only tentatively, is thought to be measured by Position Orientation, CP526A. This test is a revision of Thurstone's Hands Test, CP512, which has a loading of 0.45 on this factor.

13. *Verbal comprehension*, the ability to understand printed verbal material, is best measured by Air Corps Vocabulary (1942) which has a loading of approximately 0.70 and Reading Comprehension, CI614H, which has a loading of about 0.60.

14. *Visualization* is the ability to predict the result of manipulating or moving objects in space by visualizing the manipulation or movement. Pattern Comprehension, CP803A, is thought to have a loading of approximately 0.50 on this factor, and Area Visualization, CP715A, is expected to have a loading of about 0.50. Mechanical Principles, CI903B, has a loading of about 0.50, but is highly loaded, also, on the mechanical experience factor.

JOB DESCRIPTION AND ANALYSIS OF RADAR OBSERVER STUDENTS IN TRAINING

Extent of Job Description and Analysis

This job description and analysis is based upon observation of radar observer students in training at Langley Field, Va. The radar sets used in this training were the AN/APS-15 and AN/APS-15A. Observers watched numerous training sessions, talked with both instructors and students, and personally completed the entire training course.

The application of this analysis to the tasks required of radar observers in combat or using different equipment has limited validity. This is not a serious limitation of the analysis since its primary purpose was to serve as a basis for selecting and constructing tests to predict success in training. A brief description of the job of the radar observer in combat follows later in the chapter. The rapidity of equipment changes and improvements indicates that both descriptions will be obsolete within a year. For this reason the last section of this chapter presents a discussion of changes which may be expected in the radar observer's job in the near future.

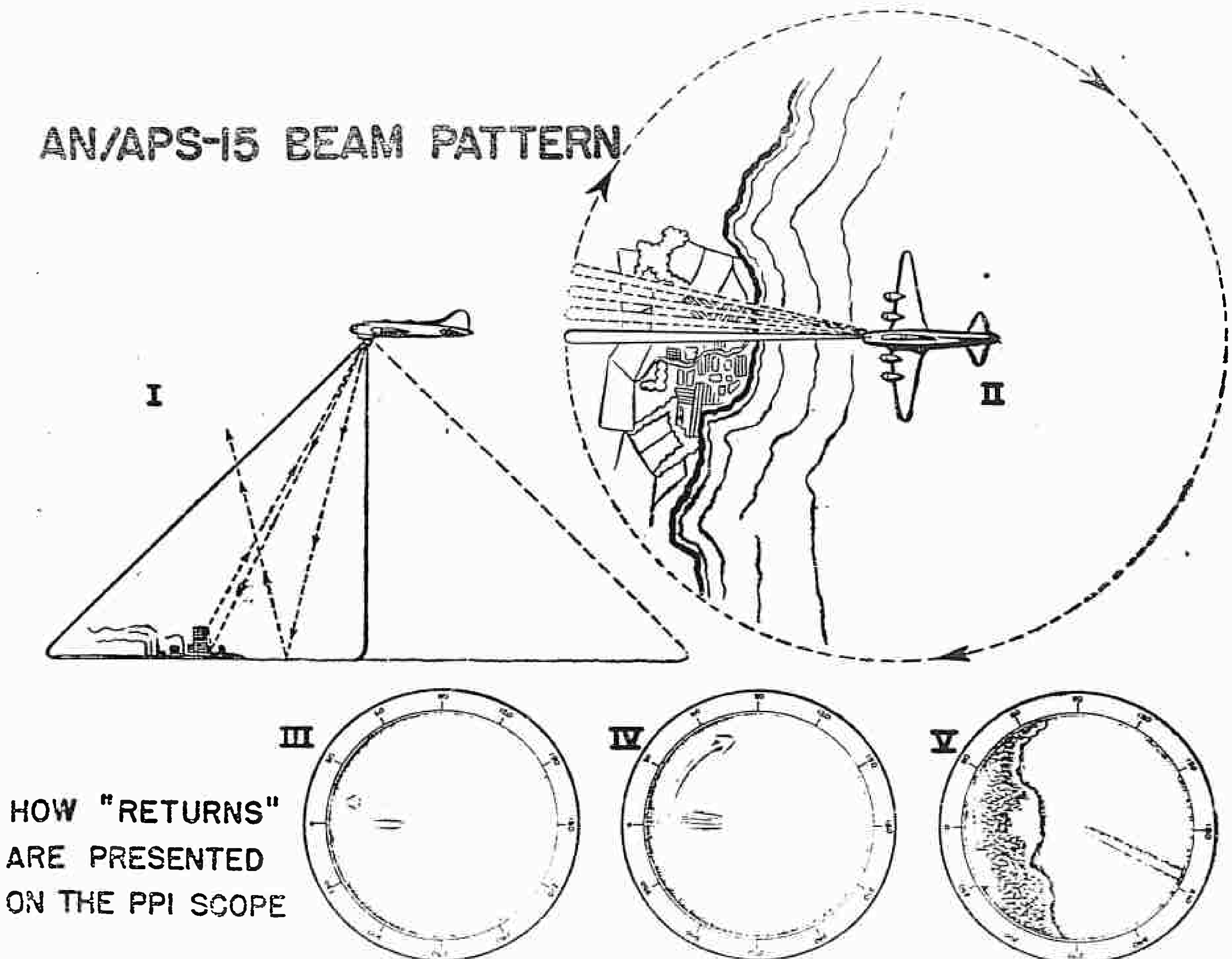

AN/APS-15 BEAM PATTERN
I
II
III
IV
V
HOW "RETURNS"
ARE PRESENTED
ON THE PPI SCOPE

FIGURE 42.

The radar observer's job will be discussed under four headings: Set operation, scope interpretation, bombing, and navigation. Each discussion will consist of a job description and a job analysis. These will be presented following a brief introduction to the equipment.

Introduction to Air Borne Radar Equipment

Training missions were carried out in B-17 and B-24 type aircraft: in both, the quarters were cramped and the table working surface provided the operator is very limited. Figure 4.1 shows a typical in-

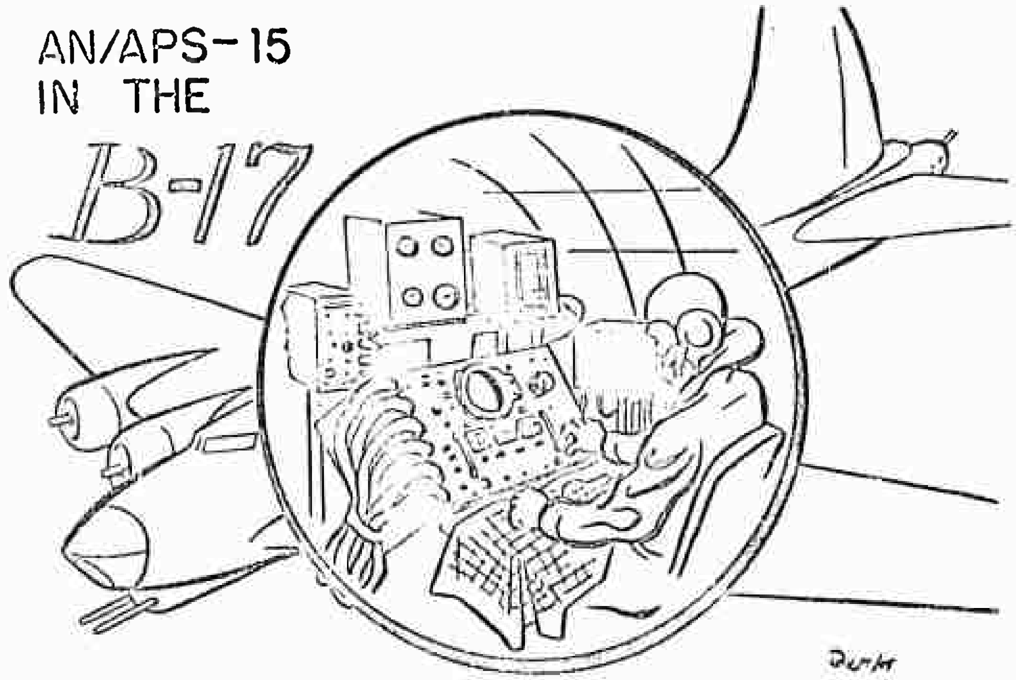


FIGURE 4.1.

stallation of the AN/APS-15 in the B-17. The AN/APS-15 and AN/APS-15A are alike except for relatively minor details. The "eye" of a radar set of this type is a directional antenna mounted in a protecting radome beneath the aircraft.[3] The antenna rotates at 26 revolutions per minute. As it rotates, a transmitting-receiving system alternates, at the rate of several hundred times per second, between transmitting short bursts of high frequency radio energy and receiving the reflected energy or echoes. Each pulse is beamed toward the earth in a narrow pattern so that as the antenna rotates, as shown in part II of figure 4.2, successive pulses cover adjacent narrow strips of the terrain radiating from the point beneath the aircraft. Thus, a huge circular area of the terrain is scanned as the antenna makes each rotation.

The energy reflected from the terrain is translated by the receiving and presentation circuits of the set into a picture of the terrain below

[3] The antenna and radome are shown in figure 4.3.

the aircraft. The picture is presented on an oscilloscope known as the plan position indicator or PPI scope. The echoes of reflected energy are received by the antenna and presented on the scope as an illuminated rotating radius, called the sweep. Actually the swing is a stream of electrons which rapidly plays across the surface of the scope beginning at the center and moving outward radially as shown in part III of figure 4.2. It appears as a constantly illuminated rotating radius because the movement of the stream is much too rapid to be seen. The position of the sweep is rotated in phase with the rotation of the antenna as shown in part IV of figure 4.2. As the electrons strike the surface of the scope, which is coated with a fluorescent substance, they create spots or returns which persist for a short time and vary in brightness depending upon the number of electrons striking the coating. In a complete rotation, the entire circular surface of the scope brightens as it is sprayed in rapid, successive, radiating movements, illustrated in part V of figure 4.2.

The various terrain features such as mountains, towns, and lakes, reflect characteristically different amounts of energy. Also, energy reflected from distant objects is received later and less strongly than returns from near objects. As the stream of electrons begins each outward radial movement, the number of electrons projected on the scope surface is governed by the strength of the echoes received from near objects. Reflected energy from more distant objects illuminates the sweep farther toward the edge of the scope. As the sweep rotates, always controlled by the amount of reflected energy and the relative time it is received, a circular picture of the terrain below the aircraft is produced with some terrain features appearing as bright areas and others appearing as dark areas. The appearance of the scope picture is shown in part V of figure 4.2 and figure 4.3. The center of the picture represents the point directly below the aircraft, and distant objects are presented toward the edge of the scope.

An auxiliary unit of the set can be made to project bright concentric circles on the scope marking off uniform and known distances from the center. These range marks, together with an azimuth scale, graduated in degrees around the edge of the scope, convert the picture into a polar grid map on which it is possible to express the position of any point in terms of distance from the center and bearing in degrees in azimuth.[4] An additional and more precise range measuring device, the range unit, enables the radar observer to set a bright circle, the bomb release circle, on the scope accurately at any given distance from the aircraft's ground position.

A second scope, the A scope, presents targets as vertical pips and is used for tuning and calibrating the set. The controls which the

[4] The appearance of the scope with range marks and azimuth calibration is shown in figure 4.4.

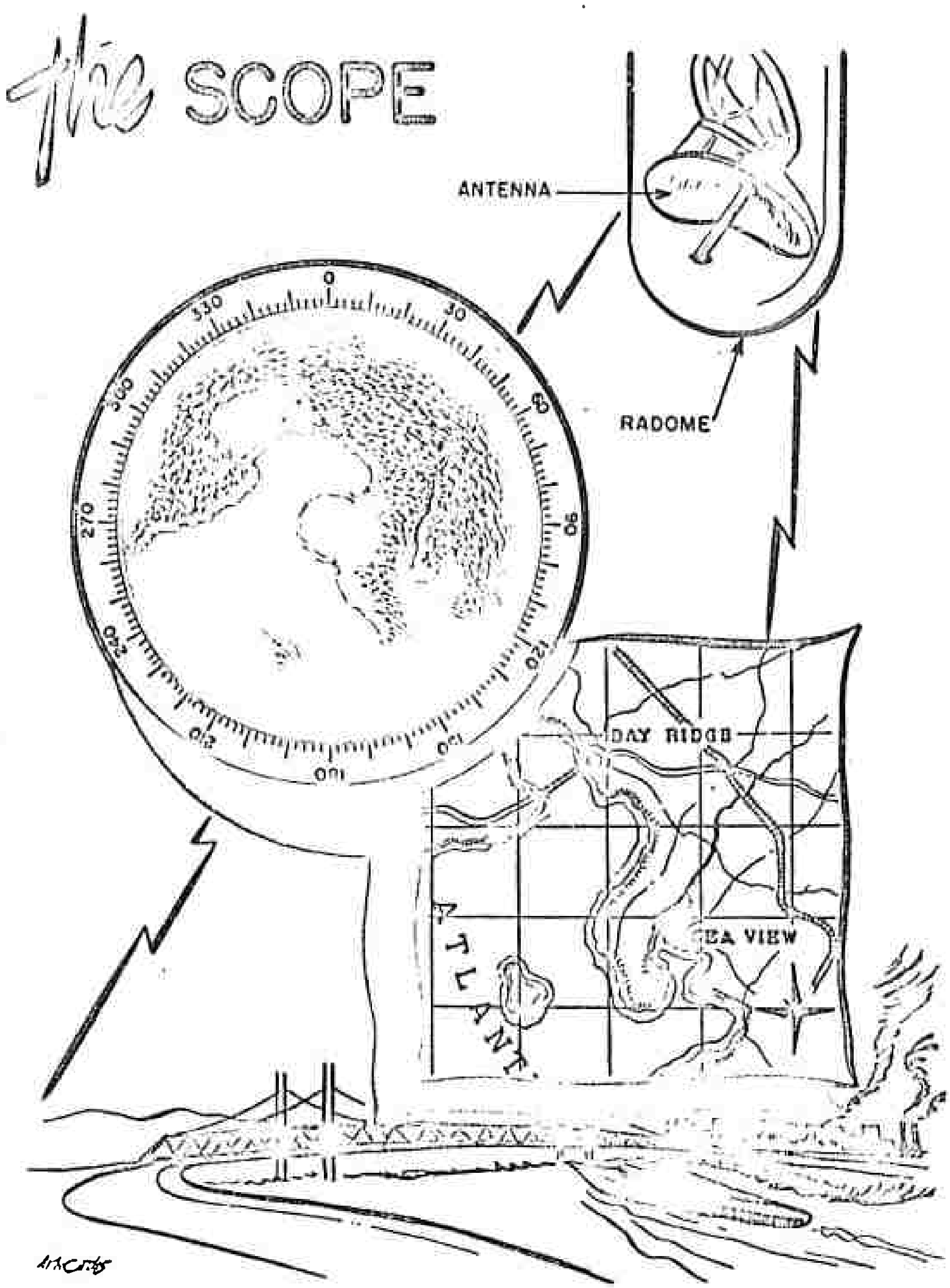


FIGURE 4.3.

radar observer must manipulate are distributed over the several units of the set.

Set Operation: Job Description

Preoperational check.—Prior to turning on the set, the controls must be positioned correctly. This prevents damaging parts of the set by the initial surge of electricity from the aircraft's power supply, prevents overloading the power supply, and places the controls in known positions that will facilitate the subsequent starting and tuning of the set. The preoperational check consists of checking for loose connections, checking for incorrect and defective fuses, and checking the positions of about 20 switches, knobs, and screw-driver adjustments. The check is ordinarily carried out before the aircraft takes off.

Starting and tuning.—To start the set properly and without damage, certain controls must be turned and adjusted in a prescribed manner. Starting includes applying power to the set, and properly brightening, focusing, and centering the sweeps on both the PPI scope and the A scope. This consists of about 20 steps, most of which have to be performed in a given sequence. The radar observer checks his adjustments by reading a current meter and by watching the bright lines presented on the PPI scope and the A scope.

Tuning consists of turning on the transmitter, picking up ground returns and adjusting the transmission, receiving, and presentation systems for maximum definition on the PPI and A scopes. About 20 steps must be performed in sequence to tune the set according to standard operating procedure. During the tuning, the radar observer adjusts for specific or maximum current meter readings, and for maximum contrast between target returns and other ground returns. On the A scope, this contrast is in terms of height of bright returns, or "pips" above a base line. On the PPI scope, target returns are brighter than other ground returns. Starting and tuning are carried out after the aircraft has taken off and reached an altitude of about 1,000 feet.

Calibration.—The range unit is a device for precise measurement of distance. Its accurate functioning is essential to bombing and precision navigation. Calibration consists of checking the range unit against a standard measuring scale built into the set and presented on the A scope as a line divided into 20 units. The radar observer manipulates knobs and screw-driver adjustments, making a very fine adjustment until all points on the two scales are lined up. The process involves 15 to 30 steps, depending on the set and the operator's skill. Most of the steps must be performed in prescribed sequence. Since the calibration of the set is affected by altitude and temperature, it is usually done after the plane has reached the altitude at which the mission is to be flown.

Flight maintenance.—The success of a radar mission may depend upon the radar observer's ability to correct certain set malfunctions during flight. The radar observer must be able to diagnose what is wrong with the set from the particular circuits that are inoperative, from the various meter readings, or from the picture presented on the scopes. Common troubles are blown-out fuses, loose cables, and variations in the aircraft's power supply. The radar observer must know what things to check when specific symptoms are noticed.

Maintaining set efficiency.—Continual adjustment is necessary to insure that the set remains properly calibrated and tuned. Changes in temperature, altitude, power supply, weather conditions, and terrain all have their effect. Usually such variations will not make navigation and bombing impossible, but will effect accuracy considerably. If he is to obtain maximum operating efficiency the radar observer should be aware of possible changes and make periodic checks on the power supply, tuning, and calibration.

Turning off set.—The process of turning the set off is simple, consisting of only seven steps. The only restriction is that four brilliance controls must be turned down before the power switch is turned off to prevent burning the scopes.

Set Operation: Job Analysis

In learning to operate the set, the radar observer student must study technical material, memorize the steps in the operating procedures, and know the location and function of the various set controls. This learning involves *memory I* [5] and *verbal comprehension*.

The student who develops a rationale for the operating procedures, the functions of the controls, and the relationship between the parts of the set will probably learn set operation more easily and retain the material longer. Even though his rationale is not technically correct, it will aid him in remembering which step is next in tuning, which fuse to check if a certain circuit goes out, and which controls to manipulate to improve the scope picture. This ability can probably be measured by tests of "scientific information" and tests of experience with electrical and mechnical contrivances. The aircrew classification tests loaded with the *mechanical experience* factor should also constitute fairly adequate measures of this aspect of the job. Some *reasoning I* (general reasoning) may be involved here also.

The radar observer who is thorough and systematic will operate the set more carefully and check the tuning and calibration more frequently. This quality of work may be related to habits of organization and thoroughness as shown in the individual's work and hobby history. The same quality has been designated "systematic diligence" elsewhere.

Some radar observers exhibit "finger trouble" in tuning the set. They constantly manipulate the controls trying to improve the picture and neglect other tasks. Such men are never sure of having obtained the best presentation possible under prevailing conditions. This may

[5] In this and later job analysis sections, all factor names have been underlined or placed within quotation marks. *Underlining* indicates that the factor has been isolated in factor analyses in the Aviation Psychology Program. All others, hypothetical or isolated elsewhere, have been placed within quotation marks.

indicate poor visual memory, *memory II*, since the radar observer must remember the best picture obtained during the control adjustments and finally reproduce it. He must also remember the appearance of scope pictures produced by his instructors under various conditions. "Finger trouble" may also indicate a lack of confidence in one's operating ability.

Psychomotor precision is involved in manipulating the switches, knobs, and screw-driver adjustments. Fairly fine adjustments of the rotatable controls are required but little speed is necessary.

"Visual acuity" and "brightness discrimination" are required in observing the scopes, dials, and neon light, and particularly in detecting the near-threshold length and brightness changes.

Scope Interpretation: Job Description

Interpreting returns.—The features of the terrain are represented as characteristic returns which constitute the picture on the PPI scope. The radar observer selects specific returns to determine his ground position, to find the target area, and to detect within the target area the aiming point to be bombed. Unless the returns are correctly interpreted and identified on the map, all the radar observer's skills in set operation, navigation, and bombing are of no value.

As shown in figure 4.3, returns on the scope differ in brightness, size, and shape. Water gives practically no return. Islands on water appear as bright areas against a dark background. Built-up areas on land, such as towns, appear as bright returns against a less bright background. Rivers, lakes, and inlets appear as dark areas with bright lines representing the far shores. Mountain ranges appear as long bright areas with shadows behind them. Towns behind mountain ranges or in valleys are hidden except when approached from certain angles. Dense clouds appear as bright areas and usually have complete shadows behind them. Bridges appear as sharp bright lines against the dark water. Towns, islands, rivers, lakes, etc. retain something of their shape, but details are usually lost or distorted.

The appearance of returns from the various terrain features must be learned and remembered by the radar observer. Because these differ considerably, the task of deciding whether a return represents a river, lake, town or mountain is usually not difficult. Only occasionally is there some problem, as for example, deciding whether a return is a dense cloud or a town. However, specifically identifying a return once it has been recognized as a town is more difficult. By noting the size, shape, and brightness of the return, the radar observer is sometimes able to identify or name it without reference to other returns. Frequently, however, none of the returns from towns have characteristic shapes or sizes and no other characteristic terrain returns are available. In such instances where the specific returns are homogeneous and without individuality, identification is possible only by the pattern which the returns form in terms of distances and bearings from each other.

Identification of aiming points within towns or target areas, which is crucial to accurate bombing, is a very difficult task. The various areas within large cities can usually be distinguished on the scope if the set is operated correctly. For example, factory and business areas ordinarily give brighter returns than residential areas. Railroad yards are brighter than surrounding areas. The radar observer must remember, from the scope photographs used in briefing, the shape of the aiming point, how it contrasts with the rest of the target area, and its position relative to other elements in the target area. He must operate the set skillfully enough to define the aiming point on the scope and he must recognize it from the size, shape, brightness, and pattern cues. This identification must be done rapidly since the target area on the scope breaks up into its elements only during the last 5 or 10 miles of the bomb run which is covered in from two to three minutes.

Difficulties in interpreting and identifying returns depend in part upon the radar observer's other skills. Poor set operation will cause returns to be poorly defined or lost altogether. If the radar observer does his navigation calculations rapidly, he will have more time to watch the scope. As new returns appear on the scope he will be able to identify them by referring to returns he already has identified. Conversely, the radar observer who is slow in computational work will frequently be faced with the task of interpreting and identifying a completely unfamiliar pattern of returns.

Scope interpretation is facilitated not only by rapid but by accurate navigation. Radar observers who do accurate navigation know their approximate ground position and course and can predict which towns and terrain features are about to enter the range of the set and where the returns will appear on the scope. From this information, they are able to identify individual returns when they are distorted and when parts of patterns of returns are not visible.

When the azimuth stabilization unit of the set is "on," the top of the scope picture is true north. The returns are in the same relative positions in which they appear on the map.[6] When azimuth stabilization is "off," the top of the scope represents the direction in which the aircraft is heading. If the aircraft is heading in any direction other than true north the pattern of returns is rotated from the position of the corresponding towns on the map. This rotation increases the difficulty of identifying patterns and targets.

Interpreting motion of returns. The radar observer can set an illuminated radius, the lubber-line, on the PPI scope to indicate the

---

[6] The relation between the scope picture and a map when azimuth stabilization is "on" is shown in figure 4.4. This relationship would be the same regardless of the heading of the aircraft.

direction the aircraft is heading. By observing the movement of returns in relation to the lubber-line, he can determine the direction in which the plane is drifting or being deflected from its heading by the wind. If the aircraft is not drifting at all, the returns will move parallel to the heading of the aircraft. If the aircraft is drifting to the right, returns will move from the right to the left of the lubber-line.

In navigation and bombing, the radar observer interprets the movement of returns on the scope in terms of direction in which the aircraft is drifting. In navigation, the direction of drift gives him a check upon his wind computation and, if measured accurately, can be used to compute a wind. In bombing, the radar observer "kills drift" on the target by giving heading corrections until the target is moving neither to the left nor to the right of the heading line. This interpretational problem is most difficult when the aircraft is heading toward the bottom of the scope. Then the radar observer's right is the aircraft's left and vice versa. Here it is important that he interpret the motion of the returns relative to the aircraft's heading and not relative to his view of the scope.

Scope Interpretation: Job Analysis

Identification of returns from shape, size, brightness, and pattern seems to be related most closely to the perceptual speed and memory II (visual memory) factors. Perceptual speed is important in the task of locating map features which are similar to the scope returns and/or in locating scope returns which are similar to points on the map. Memory II seems important in recognizing, on the scope, patterns of returns previously seen on the map or during the briefing session and in remembering the characteristic appearance of the various types of terrain. In remembering and matching patterns of returns, length estimation as well as the abilities to estimate sizes and angles are probably important. All patterns are composed of various sized returns at different distances and directions from each other.

It is likely that the student who has had scientific training and experience can understand more easily why the various terrain features yield characteristic returns. Therefore, "scientific background" would probably be an aid to remembering and identifying returns from individual targets.

Perceptual speed and memory II seem particularly important in identifying the aiming point. Since the identification must be done rapidly, the radar observer must rely on his memory of the target area; he has little time to consult scope photographs. The pattern of returns is often only dimly visible or is partially obscured. Fine brightness discriminations are often required even to perceive the pattern.

Perceiving and interpreting the motion of returns across the scope seem to depend upon ability in memory II and space I (spatial relations). The returns usually move across the scope at an almost imperceptible rate. The radar observer cannot take time to watch the scope for any length of time to determine the direction of motion of the returns. Instead he must look at the scope at intervals. Memory II is necessary, since each time he looks at the scope, he must remember where the returns previously were in order to determine the direction they moved.

When the aircraft is heading in a southerly direction, with azimuth stabilization "on," the radar observer must imagine himself rotated in space to make his right and left correspond to the aircraft's right and left. This will enable him to interpret correctly whether the aircraft is drifting to the right or left. When azimuth stabilization is "off," this same ability to imagine one's self rotated in space is necessary in all headings except true north. This operation is similar to that required in tests which have loadings in space I.

Insofar as scope interpretation is dependent upon speed of computational work and quality of set operation, ability here will be related to the abilities described in the job analyses of set operation, above, and navigation, below.

Navigation: Job Description

Flight planning.—Before a mission, the radar observer obtains the following information: route, forecast or "metro" wind, altitude, temperature forecast for the altitude, and indicated airspeed. On a map, he draws in the courses for the navigation legs and bomb runs. He measures the direction of each leg (true course or track to make good) using the dividers and the protractor scale of the Weems plotter. He measures the length of each leg by first stretching the dividers over the leg and then laying it out along the degrees and minutes scale on the edge of the map, one minute of latitude being equal to one nautical mile.

The radar observer converts indicated airspeed from statute miles per hour to nautical miles per hour using the slide rule scale of the E-6B computer. Then setting pressure altitude opposite forecast temperature for that altitude on the E-6B computer, he reads true air speed. He draws the wind force and wind velocity on the vector face of the E-6B, centers the vector face on the true air speed, and determines what direction the aircraft would drift if headed on the true course. He then "juggles" the computer until the drift would cause the aircraft to travel along the true course. From the computer he reads the true heading which the aircraft should take in order to make good the true course and the ground speed that the aircraft will make on that heading. On the slide rule scale of the E-6B computer he determines how long it will take to fly each leg, having already found the distance and ground speed.

All of the information described above, both given and computed, is entered in the flight plan section of the log. From this flight plan, the radar observer can tell the pilot, on the basis of the metro wind and the other predicted data, what heading to take for each leg and how long it will take to fly the leg.

Maintaining record of ground position and flight data.—One of the most important functions of the radar observer is to provide ground position reports. This function is invaluable at night or when the aircraft is over clouds. The area represented on the PPI scope is a polar coordinate map, the center being the ground position or point directly under the aircraft. Ground position can be determined by identifying one or more returns on the scope, determining the spatial relationship between the center of the scope and the returns, and plotting this relationship on the map starting with the points corresponding to the returns and working back to the ground position corresponding to the scope center. Plotting ground position in this manner is called "taking a fix."

Three types of fixes may be taken: a range and bearing fix, a multiple bearing fix, and a multiple range fix. The first is more frequently used. It consists of measuring the range and bearing from the center of the scope to a single identified return. A multiple bearing fix consists of measuring only the bearing from the center of the scope to each of two or more identified returns. A multiple range fix consists of measuring only the range from the center of the scope to each of two or more identified returns.

To determine the bearing of a return, the radar observer makes use of a rotatable plexiglass face covering the scope. This plexiglass is bisected by an etched line which may be rotated until it passes through the return. He then reads the bearing at the point where the line intersects the azimuth ring at the edge of the scope. This value is the direction from the center of the scope to the return. In figure 4.4, the bearing of the point of land being used as a check point is 328°. The range of a return may be measured in two ways. In the most frequently used method the radar observer positions bright concentric circles on the scope at any one of the several intervals. The usual interval is 5 or 10 miles. Using these circles, he reads the distance to a given return by estimating to the nearest mile or half-mile between the marks. In figure 4.4, the check point is 40 miles out on the scope. In the second method, less frequently used, he measures ranges more precisely with the bombing circle. He turns on the bombing circle, turns a knob until the bright circle expands and touches the return, and reads the distance from a nautical mile scale or counter dial calibrated in hundreds of feet. In the latter case, he has to convert to nautical miles before plotting the fix.

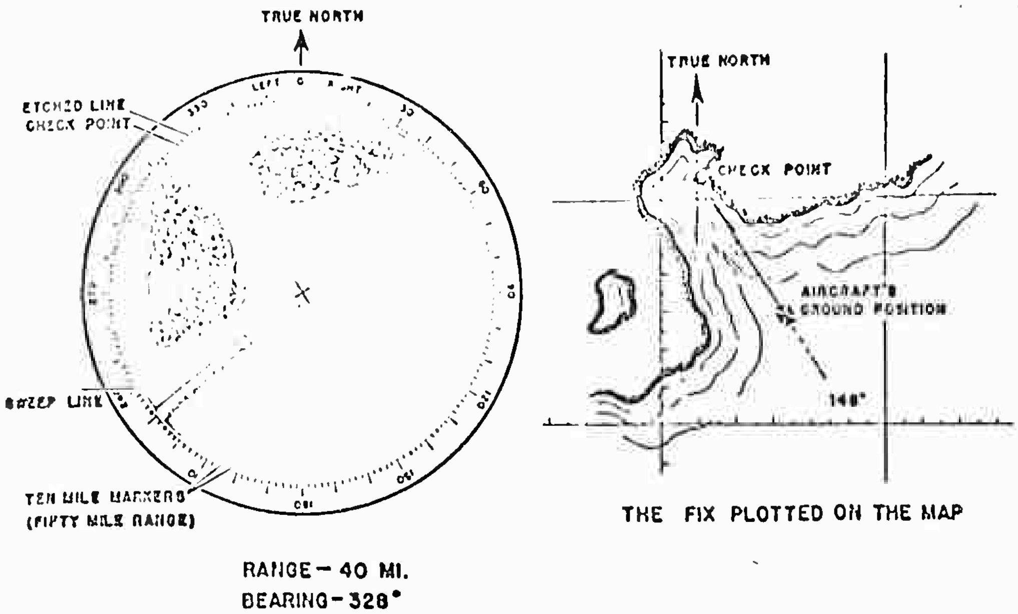


FIGURE 4.4—Fix-taking from the PPI scope.

The radar observer plots a range and bearing fix by first plotting the direction from the point corresponding to the return. Since his bearing reading from the scope was the direction from the ground position to the return, he must plot the reciprocal of this, the bearing plus 180°, as shown in figure 4.4. This directional plotting is done using the protractor scale of the Weems plotter. After positioning the plotter, he measures off the proper range on his dividers from the latitude scale at the edge of the map and lays this distance, 40 miles in figure 4.4, along the plotter. The resulting point is the ground position of the aircraft at the time of the fix.

To make possible further navigational work beyond the flight plan, it is essential that the radar observer maintain a continuous record of the aircraft's ground position and track made good. Consequently, he must take fixes at regular intervals, log them, and plot them. He enters them in the log by recording the time of the fix, the return used, and the range and bearing date. He also records flight data in the log at regular intervals, particularly noting changes in air speed, altitude, and true heading. To obtain this information, he must read the fluxgate compass, altimeter, air speed indicator, and free air temperature gauge.

Wind computation.—The winds predicted by the weather department have only a limited value for exact navigation since they are approximations, apply to limited areas, and cannot take into account sudden wind shifts. One of the important tasks of the radar operator is to compute winds using instrument data and data from his fixes. Three methods of wind computation will be described.

In the airplot method, from the fixes entered in the log and plotted on his map, the radar observer can determine the track made over a given interval of time and his ground position at the end of the interval. He plots on the map the position the aircraft would have reached if it had been traveling in still air. This position is called the air position and is shown in relation to the ground position in figure 4.5. To determine the aircraft's air position at the given time, the radar observer starts with a previous ground position and, using the Weems plotter, plots the true heading or course that would have been made good in still air. To determine the distance that would have been traversed in still air, he uses true air speed on the E-6B slide-rule scale. He measures this distance off the side of the map

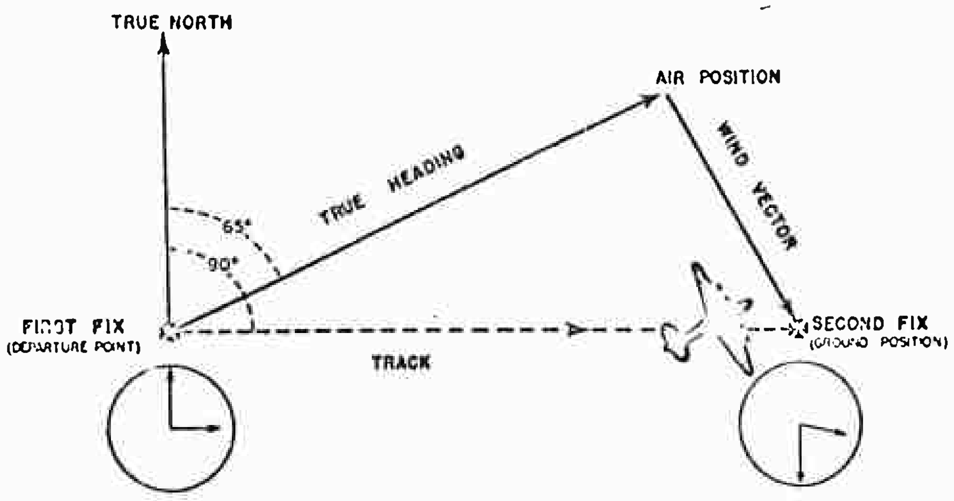


FIGURE 4.5—Determining wind by airplot.

with the dividers and lays it off on the true heading. The resulting point is the air position, the hypothetical point the aircraft would have reached flying in still air on the true heading at the rate of the true air speed for the given time interval. The line drawn, as in figure 4.5, from the air position to the ground position at the corresponding time, indicates the direction and distance the wind carried the aircraft during the time interval. The radar observer measures the direction of this line, in degrees azimuth, with the Weems plotter. He knows that the wind blew the aircraft from the air position to the ground position and that wind direction is indicated in terms of the direction from which it comes. He then measures the length of the line with the dividers and divides it by the time interval on the slide rule scale of the E–6B to get the wind force.

The second method of wind computation is the regular computer method. As previously stated, the radar observer can determine from his logged and plotted fixes the true course or track made good. By measuring the distance between two fixes and dividing it by the time interval between the fixes, he can obtain the ground speed for the true course. He places the vector face index of the E-6B computer at the true heading, centers the face on the true air speed, and plots the ground speed on the true course. He rotates the vector face, reads the wind direction from the azimuth scale and the wind velocity along the mile scale.

The third method of wind computation is the grid computer method, also called target timing. The radar observer takes two or more successive fixes on the same return and times the interval between the two with a stop watch. He then plots the fixes on the vector face of the E-6B computer over the square grid. Since returns travel across the scope parallel to the track the aircraft is making good, the line connecting these fixes is parallel to the aircraft's track. He rotates the vector face until the connecting line is aligned with the vertical grid lines and reads the direction of the track. At this point he must check that he is not reading the reciprocal direction by comparing the computer result with his general estimate of the track. He divides the distance between fixes by the time between fixes and computes an average or over-all ground speed. He then uses true course and ground speed on the vector face of the computer to determine the wind by the regular computer method described above.

Planning remainder of flight.—Frequent and accurate computation of ground position and wind enables the radar observer to navigate the rest of the mission accurately. Unless the metro wind is very accurate, the aircraft will likely fly off course and not make a track for the destination. In this case, the radar observer uses his best wind information to determine what heading he should give the pilot to correct the track. To this end he must do dead reckoning navigation along his track from the last ground position and determine an approximate point he will reach several minutes later. On the E-6B, he computes the distance he will travel in a certain number of minutes with his present ground speed. He lays this distance out along the track and plans to make a course correction at the resulting point. He draws the true course from that point to the destination and measures its direction. He juggles the computer, as described above, knowing true course, true air speed, wind velocity, and wind direction, and determines what heading will be necessary to make good the true course. When the course correction point is reached—that is, when the given time interval has elapsed—he gives the pilot the new heading. He also computes an estimated time of arrival for the destination so that he will know approximately when he is there. This information is particularly important when the point cannot be seen or picked up on the scope.

The radar observer uses his latest wind information in a similar manner to compute headings for the bomb run so that the aircraft will drift as little as possible away from the target.

Since any single wind computation may be inaccurate and the wind may change, the radar observer determines new winds as often as he has time. The latest or most trusted wind is used in computing course corrections or headings on new legs.

Navigation legs in training range in length from 60 to 100 miles. Consequently, the radar observer has a limited period in which to compute winds, course corrections, and new headings. After allowing for passage of sufficient time to collect accurate wind data he will have from 10 to 25 minutes depending, of course, upon the length of the leg and the wind. During this time he must also watch the scope, identify new returns, and maintain the efficient operation of the set. If the navigation leg is followed by a bomb run, he has to spend time near the end of the leg predicting ground speed and absolute altitude over the target and preparing the set for the bombing procedures to be described below.

Navigation: Job Analysis

One of the principal problems confronting the student on radar training missions is that of completing the necessary navigational work in time, without having to do hurried and inaccurate work. Good radar observers have systems for carrying out their work by means of which they obtain each bit of data as soon as it is available and spread out their computational work as much as possible. This planning ability depends partly upon already having thorough understanding of how the separate steps of navigational procedure are related to each other. Such an integrated view of navigational work is probably most readily obtained by a student who has a background of mathematics and science. "Scientific background" and reasoning I (general reasoning) are believed to be the abilities involved in this aspect of navigational work planning.

In his navigational work, the radar observer reads many dials and scales: E-6B computer scales, navigation instruments, and dials and scales on the set. Components of scale-reading ability seem to be numerical facility and space I (spatial relations).

In reading ranges on the PPI scope, the radar observer estimates distance within the interval between the range marks. He makes similar estimates in measuring along the nautical mile scale (minutes of latitude) on the map. A length estimation ability is believed necessary for these tasks.

After the radar observer obtains the navigational data and plans how to use them, he sets them into his computer, plots them on a chart, or sets them into the radar equipment. Psychomotor precision is involved in these operations, particularly in handling the dividers, plotter, and computer.

As the radar observer carries out his plotting, measuring, and computing, he should check the results of each step. He can do this by making approximate calculations mentally. Numerical facility and reasoning I are important in making these rough checks as well as in making the original computations.

In wind calculation, particular attention must be given to correctly interpreting the drift and the direction of the wind vector on the computer or airplot. Otherwise, reciprocal winds will be obtained or the drift will be given in the wrong direction. Space I is probably important to success in these tasks.

The radar observer must constantly attend to his instruments and set, keeping track of changes in heading, airspeed, ground position, etc. To have this data available without depending upon his memory, he must record it in his log. Diligent attention to the ever-changing scope picture will make his orientational problem much simpler. As previously stated, this "systematic diligence" will depend upon the existence of habits of doing tasks thoroughly, carefully, and systematically.

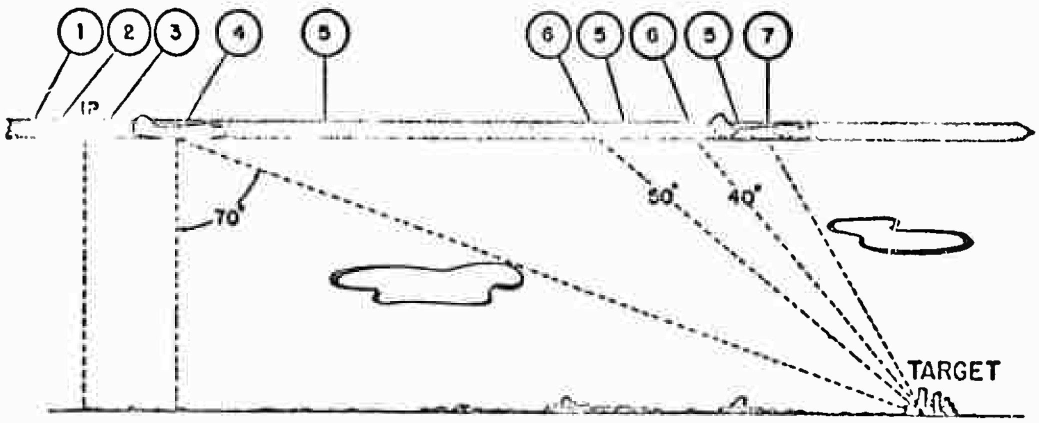
Two other abilities seem associated with navigation operations. One of these, verbal comprehension, is involved in learning the subject matter of navigation. The other, "scientific background" may facilitate learning the computational procedures, the use and calibration of instruments, the use of computers and maps, etc.

Bombing: Job Description

Setting up computer box.—When bombs are to be dropped on a target from a moving aircraft, they must be dropped at some point before the aircraft is directly over the target. Otherwise, the forward motion imparted to the bombs by the aircraft will carry them beyond the target during the time of fall. Among the factors which govern how far ahead of the target the bombs should be released, two are of concern to the radar observer. These are the aircraft's absolute altitude and ground speed on the bomb run. For accurate bombing, the radar observer must determine these factors and use them to prepare the set for bombing. He determines the expected ground speed and absolute altitude over the target before he reaches the IP at the beginning of the bomb run. He then predicts ground speed by putting his latest wind onto the E–6B and computing ground speed on the true course from the IP to the target. Absolute altitude is determined at some point before the IP with the SCR–718 Radar Altimeter, reading in terms of feet. An alternative method is to use the bombing circle on the AN/APS–15 or AN/APS–15A, setting it on the innermost return on the PPI scope and reading the altitude in terms of nautical miles from the computer drum on the AN/APS–15 or in terms of hundreds of feet in the counter window of the AN/APS–15A. The SCR–718 reading is more accurate. In either case the radar observer must correct this altitude for the difference between the terrain altitudes at the target and at the point where the absolute altitude was determined.

In direct bombing, where the radar observer releases the bombs, he sets the absolute altitude and ground speed into the computer box of the set. On the AN/APS-15 he adjusts a pair of cross hairs up a nautical mile scale, setting the horizontal line at the absolute altitude. He then rotates the computer drum until the proper ground speed line is under the vertical cross hair. On the AN/APS-15A, he twists a knob and sets the absolute altitude on the altitude counter in terms of hundreds of feet. He then twists another knob to replace the proper ground speed under the dial marker. In both sets, this positions the bright bombing circle on the PPI scope at a distance from the center such that when the target touches it, the bomb should be released.

In coordinated bombing, the radar set yields information which is put into the visual bombsight and the bombsight mechanism automatically releases the bombs at the proper moment. The procedures carried out on a typical coordinated bomb run are summarized in figure 4.6. Before the IP, the radar observer informs the bombadier of the altitude and ground speed. These are used to place rate and dropping angle into the bombsight. During the bomb run, the radar observer notifies the bombardier when the target is at certain specified angles from the aircraft. These "sighting angles," measured in degrees below the horizontal, are used to start the bombsight mechanism and to make adjustments in rate and dropping angle which are necessary because of inaccuracies in predicted altitude and ground speed. For each angle, the radar observer positions the bombing circle on



1. DETERMINES ABSOLUTE ALTITUDE AND GROUND SPEED, WHICH BOMBARDIER SETS INTO BOMBSIGHT.
2. SETS FIRST SIGHTING ANGLE AND ABSOLUTE ALTITUDE INTO COMPUTER BOX.
3. STARTS KILLING DRIFT ON THE TARGET.
4. CALLS OFF FIRST SIGHTING ANGLE (70°) WHEN TARGET TOUCHES BOMBING CIRCLE.
5. RESETS COMPUTER BOX AND IMPROVES TARGET PICTURE ON SCOPE.
6. CALLS OFF NEXT SIGHTING ANGLE.
7. RELEASE POINT— BOMBS AWAY.

FIGURE 4.6.—Coordinated bomb run.

the PPI scope and calls the angle when the target touches the circle. One of the advantages of coordinated procedure over direct bombing is that the radar observer has a chance to judge the coincidence of bombing circle and target four or more times. Consequently there is a greater chance that small errors in setting up the computer box or in the observing when the target touches the bombing circle will cancel out.

The radar observer sets up the computer box on the AN/APS-15 for the first sighting angle (70°) by adjusting the cross hairs on the nautical mile scale to the absolute altitude and rotating the range drum to a point where the cross hairs intersect the 70° sighting angle line. On the AN/APS-15A, he enters tables with the altitude and ground speed and finds the altitude counter reading (hundreds of feet) which he must set into the computer box for the 70° angle.

All of the initial preparation of the computer box described above is carried out at the end of the navigation leg immediately preceding the bomb run.

Killing drift.—In order for the bombs to fall directly upon the target, the aircraft must be making a track for the target at the time of release. If the aircraft is drifting to one side of the target, the bombs will fall on that side. The process of correcting the aircraft's heading until it is tracking toward the target is called "killing drift on the target."

At the beginning of the bomb run, only the target area as a whole is defined on the scope. Later, as the aircraft approaches within 5 or 10 miles of the target, the target area is seen as composed of smaller sections and the aiming point is defined. Consequently, drift is killed initially on the whole target area and finally upon the aiming point.

One method of killing drift on the target is called "homing." In this procedure, the radar observer sets on the target the etched line which bisects the PPI scope. If the aircraft is tracking toward the target, the target will move down this line to the center of the scope. If the target moves to the right of the line, the aircraft is drifting to the left of the target and a right correction in heading is necessary. As the radar observer notes in which direction the target moves, he gives small arbitrary corrections to the pilot. He continues this trial-and-error procedure until the target is moving directly down the line.

A more accurate drift-killing procedure is the "multiple-drift method." The radar observer reads the bearing of the target before and after it has moved one-fifth, one-fourth, or one-third of the way toward the center of the scope. He multiplies the difference between the two bearings by 5, 4, or 3, the reciprocal of the fraction used, and gives this number of degrees as a correction to the pilot. Again he must give the correction to the right if the target moved to the right of the first bearing or to the left if it has moved left.

The radar observer may use the range marks to delineate the portion of the distance the target moves. For example, he can follow the target as it moves from the 20- to the 15-mile range marks, in which case the target would have moved one-fourth of the distance toward the center. Some sets have lines etched on the face of the scope marking off one-fourth or one-fifth of the distance to the center. To use these most efficiently, the radar observer manipulates a control to move the target out to the outer mark when beginning a multiple-drift reading.

Killing rate and bomb release procedure.—As the aircraft moves closer to the target, the radar observer manipulates controls to reduce the area of terrain pictured on the scope. This increases the size of terrain features and finally results in definition of the various sections of the target area. He also adjusts controls concerned with tuning the set to improve the scope picture. Radar observers frequently turn on "sector scan" by means of which only a portion of the terrain below the aircraft is presented on the scope. The sweep covers this sector much more frequently than it does when traversing the entire scope, thus facilitating examination of the target. When the aiming point has been identified, the radar observer gives small arbitrary corrections to kill any residual drift.

In direct bombing, the radar observer switches on the bombing circle early in the bomb run. The aiming point becomes visible and tracks toward the aircraft. The radar observer toggles the bomb release switch at the exact moment he sees the near edge of the aiming point touch the outer edge of the bombing circle.

In coordinated bombing, when the target approaches the bombing circle positioned for the 70° sighting angle, the radar observer requests the crew to stay off the interphone during the bomb run, and warns the bombardier that the 70° sighting angle is coming up. When the near edge of the aiming point touches the bombing circle, the radar observer calls, "Mark." He quickly and accurately resets the computer box for the next sighting angle. As the aiming point approaches the circle a second time, he again warns the bombardier and specifies the angle. When the aiming point touches the bombing circle, he again calls, "Mark." He resets the computer box and repeats the procedure for as many angles as possible. A radar observer usually calls four or five angles on a single coordinated run. Meanwhile, he manipulates the controls of the set to reduce the area presented on the scope and to increase the definition of the aiming point. As the aiming point becomes defined, he usually has to give small last-minute drift corrections. The process of killing rate and drift on a coordinated bomb run is summarized in figure 4.6.

The release point corresponds to a 30° sighting angle for a ground speed of 180 miles per hour and an altitude of approximately 10,000 feet. The coordinated procedure from the 70° angle to the release point is carried out over a distance of 4 to 5 miles or a time of approximately 1½ minutes. The entire bomb run on training missions is about 30 miles long, which allows approximately 10 minutes for the drift killing and release procedure.

Bombing: Job Analysis

On the bomb run, the radar observer has much to do in a short period of time. It is essential that he plan the flight from before the IP to "Bombs Away." This plan should designate specific points at which to do all essential operations, such as obtaining absolute altitude, computing ground speed, preparing the computer box, and making multiple-drift corrections. Planning is especially important in resetting the computer box for successive sighting angles on a coordinated run. This process must be done quickly and accurately. Systematic planning of this sort, in all probability, is related to reasoning I (general reasoning), "scientific background," and "systematic diligence."

Setting and resetting the computer box also involves "immediate memory span," psychomotor precision, and scale and table-reading ability. Scale reading is involved in obtaining absolute altitude, computing ground speed, and setting the values on the computer scales. As previously stated, success in scale reading may be dependent upon numerical facility, space I (spatial relations), and "scientific background."

To kill drift, the radar observer must first select the proper section of the target area and later, when it becomes visible, identify the aiming point. The requirements involved here appear to be the same as those of scope interpretation described above, namely, perceptual speed, memory II (visual memory), and space I. The abilities required for interpreting direction of drift, memory II and space I, have also been discussed previously.

For the multiple drift procedure to be accurate, each of the two bearings must be taken on the same part of the target. When the target has moved toward the center of the scope the necessary distance, the radar observer must remember and use exactly the same section of the target that he used in reading the first bearing. This probably requires memory II.

On the bomb run, the radar observer is faced with a complex task that must be performed quickly and accurately. He must go through the multiple drift procedure, continuously give refining corrections, continuously expand the picture on the scope, manipulate the gain and tilt controls to improve the picture, give the bombardier the latest ground speed and drift information, warn the bombardier, call the first sighting angle, reset the computer box and call three or four further sighting angles, set the computer box for direct drop, make final drift corrections, and be prepared to release the bombs if necessary. The complexity and stress of this task indicate that "emotional control" or resistance to confusion are important in radar observer success.

Summary

Thus far in the job analysis, no direct statements have been made as to the relative importance of the various abilities or factors which have been described. In the following list, these abilities are grouped according to their estimated validity for predicting over-all success in radar observer training. A more quantitative prediction is not made since the reliability of the criterion will affect the absolute size of the validity coefficients. A summary will state the aspects of the radar observer's job for which each ability is thought to be important.

Abilities predicted to have highest validity.—"Scientific background" appears to be important in developing a rationale for the operating procedures, remembering the appearance of returns from various terrain features, and learning and planning the navigation work.

Space I appears to be important in interpreting the motion of returns across the scope, using the correct wind and drift direction, and reading dials.

Memory II is important in tuning the set, identifying target areas and aiming points, perceiving the motion of returns across the scope, and taking the two multiple drift bearings on the same part of the target.

Abilities predicted to have relatively high validity.—"Systematic diligence" is probably important in organizing and correctly carrying out operational procedures, organizing the navigational work to provide maximum time for every operation, keeping track of flight data, keeping an adequate log of navigational and flight data, and planning the bomb run.

Numerical facility is believed important in doing computations and making computational checks and in reading dials, scales, and tables.

Reasoning I is probably important in planning the navigation log and bomb run procedures, making computations, and checking them.

Perceptual speed is probably related to identifying individual returns and patterns of returns.

Verbal comprehension would seem to be essential primarily at the learning stages in acquiring technical information.

Abilities predicted to have relatively low validity.—*Length estimation*, together with size and angle estimation, is probably involved in remembering and identifying patterns of returns and in scale, dial, and table reading.

Psychomotor precision is probably important in manipulating the switches, knobs, and screw-driver adjustments on the set, handling the Weems plotter, E–6B computer, and dividers, and setting and resetting the computer box during the bomb run.

Mechanical experience may facilitate the development of a rationale for operating procedures.

Abilities predicted to have lowest validity.—Memory I is involved in learning and remembering the procedures for starting, tuning, and calibration.

"Visual acuity" and "brightness discrimination" are involved in tuning and perceiving patterns of returns. Individual differences are probably not great among radar observer students in these variables.

"Emotional control" is probably important in carrying out the complex task on the bomb run rapidly and accurately. Valid measures of emotional control are not known.

Abilities predicted to have no validity.—Abilities among those defined at the beginning of the chapter which probably have no validity are:: Psychomotor coordination, visualization, space II or rotational space, and pilot interest.

JOB DESCRIPTION OF THE RADAR OBSERVER IN COMBAT

Sources of Information

This sketch of the radar observer's job in combat is based upon the following sources of information: (*a*) Personal accounts of radar observers from the Eighth Air Force, (*b*) outlines of Standard Operating Procedures for radar navigation and bombing from the Twentieth and Twenty-first Bomber Commands of the Twentieth Air Force, (*c*) operational analyses from the Eighth, Fifteenth, and Twentieth Air Forces, and (*d*) articles in the magazine "Radar" which describe the use of radar equipment in the various theaters.[7] It must be noted that variability in tasks performed by the radar observer within and between the various units in a given theater is much greater than in training. An attempt is made below to describe the performance of a representative radar observer in each theater. Space does not permit summarizing all the known improvisations and variations.

European Theater

Equipment.—The B–17 and B–24 aircraft used in Europe were equipped with AN/APS–15 and AN/APS–15A radar sets. The operation of these sets was identical with that described in the job description of set operation in training.

[7] Radar: No. 7, 1 Jan. 1945; No. 9, 30 April 1945; No. 10, 30 June 1945.

Navigation.—Early radar observers were rated navigators who had little initial knowledge of the radar set. Only the lead ships were equipped with radar, there being approximately one such ship for each squadron of 12 planes. In the lead ship were two other navigators, the lead navigator who did pilotage navigation and a second navigator responsible for DR or dead reckoning navigation. The radar observer operated the set, gave fix information to the DR navigator every few minutes, and occasionally computed winds to serve as checks on the DR navigator. The airplot method of determining winds was used for the most part. The radar observer's primary responsibility in navigation was to lead the formation around flak areas when such areas were visible on the scope. This was carried out by radar pilotage, the purpose being to avoid such areas by at least 10 miles. The radar observer simply gave headings to keep the returns from the identified flak areas beyond the 10-mile range mark on the PPI scope.

Identification of check points for use in navigation was fairly easy over the coastal areas. It was more complicated over central Germany where identification had to be made largely on the basis of returns from towns. Check point identification was particularly difficult in the mountainous terrain of southern Europe covered by the Fifteenth Air Force operating from Italy.

Bombing.—The DR navigator was responsible for bringing the formation to the IP. The radar observer usually gave the heading to the target from the IP and made the initial drift corrections on the bomb run. The range of the set allowed him to pick up the target long before it could be detected visually. This advantage permitted him to give more accurate headings and initial drift corrections than the other navigators.

The radar observer set up his equipment for bombing regardless of weather conditions. If there was an undercast, he controlled the bomb release. Whether or not visual conditions prevailed, he prepared to do direct radar bombing, always being ready to take over if the bombadier could not make the release visually. Up to shortly before D-day, all radar bombing was direct bombing. After its introduction, coordinated bombing rapidly became the preferred procedure.

Identification of targets and aiming points proved to be a major difficulty. Most targets were well inland so that identification had to be made without aid from coastal returns. The sets provided poor definition of returns compared with later sets used in the Pacific Theater. Identifying the aiming point usually involved estimating its position within a large homogeneous return. In the early stages of operations with radar the quantity and quality of scope photos for briefing was not adequate to familiarize radar observers with returns from the target area.

Pacific Theater

Equipment.—While B–24's and B–17's with AN/APS–15 and AN/APS–15A sets were used in the Pacific for sea search, mine-laying, and regular bombing missions, the most frequent airborne radar installation in this theater was the AN/APQ–13 set in B–29 aircraft. A few AN/APQ–7 sets in B–29 aircraft were used. The AN/APQ–13, shown in a B–29 installation in figure 4.7, is comparable to the AN/APS–15 in function, controls, and operation, but provides

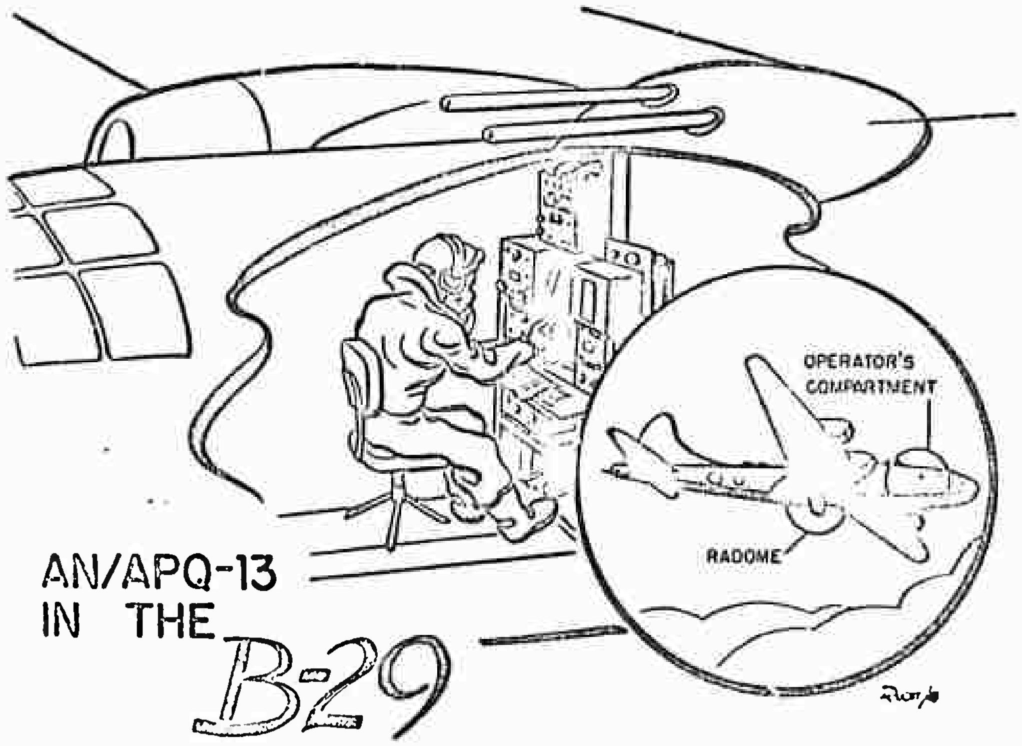


FIGURE 4.7

slightly greater definition of returns. The AN/APQ–7, a new development, is characterized by much greater definition on the PPI scope but the presentation is limited to an area of only 60° wide and directly ahead of the aircraft.

Navigation.—Airborne radar sets in the Pacific were first operated by specially trained enlisted aircrew members, designated as radar operators. Because they knew little about navigation, this phase of their work was supervised by the navigator. The radar operator simply controlled the set and made adjustments as directed by the navigator on the basis of returns on the auxiliary PPI scope. The navigator took fixes from the auxiliary scope and computed target-timing winds as described on page 39. Aiming point identification was carried out in cooperation with the navigator. The radar operator set up the bombing problem with information obtained from the bombardier. The navigator killed drift on the bomb run and either the navigator or radar operator called the successive sighting angles (coordinated bombing procedure), the radar operator resetting the computer box each time.

The enlisted radar operator was gradually replaced by the commissioned radar observer. As a result of his specialized training, the radar observer was able to assume more navigational and bombing duties. Besides operating the set, the radar observer took fixes at regular intervals, relayed the information to the navigator, computed winds from radar information, and gave them to the navigator as checks. Winds were determined largely by the target-timing metho , with the target being tracked for 6 to 10 minutes.

For a large part of the over-water flight to Japan, no radar returns were available and navigation was based upon Loran, driftmeter, and celestial data. However, because of the long range of the radar set, the radar observer was able to pick up the Japanese coast in time to aid the navigator in crossing the coastline at the briefed point. Accuracy of this first contact was vital to accuracy of the approach to the IP and turn onto the bomb run. The radar observer gave information as to when the formation should start to turn over the IP in order to come out of the turn on the briefed course. The high velocity winds encountered over Japan made it necessary to carry out all such navigational procedures quickly and accurately. All radar information was carefully checked against data from other sources.

The Japanese islands gave excellent radar returns. Once the formation came within 100 miles of the island chain, numerous navigational check points were available. The identification problem appeared simple and there was a tendency to abandon DR navigation and rely wholly upon radar pilotage. However, experience showed that identification of returns over Japan was not as easy as it first seemed. Many islands and inlets looked alike and the patterns made by the land and water returns were extremely complex. Missions were most successful when there was close cooperation between the DR navigator and the radar observer, each checking the other.

On the return legs of the mission, few radar returns were available, Under maximum range conditions, distant island chains were identifiable and finally the home islands provided fix information, Because of the paucity of navigational data on these missions, the radar observer had to maintain the set at its maximum performance to obtain all possible position data for the navigator.

Bombing.—The navigator was responsible for directing the plane to the IP. Under nonvisual conditions, the radar observer was responsible for identifying the aiming point, making course corrections, and calling sighting angles on the coordinated bomb run. Because of the large size and homogeneity of returns from the cities and built-up areas of Japan, the visual aiming point of the primary target often was not identifiable on the PPI scope. In such instances, under non-visual conditions, a secondary target which yielded characteristic and identifiable radar returns was used. However, since a sizable part of bombing over Japan had to be carried out by radar, it was impossible strategically to continue to neglect the primary target and bomb only the radar targets. Consequently, offset bombing systems were developed for bombing such targets by radar even though they were not visible on the PPI scope. These systems depend upon the use of an identifiable aiming point at a known distance and bearing from the target. Offset bombing requires extremely accurate headings from the IP to target and either timing the flight from the aiming point to the target or making already-computed adjustments in slant range settings. Because of these difficulties, offset methods were never completely satisfactory.

In general, in spite of the greater definition yielded by the AN/APQ-13 and AN/APQ-7, the radar observer usually has to estimate the position of the aiming point within a large homogeneous target return.

JOB ANALYSIS OF THE RADAR OBSERVER IN COMBAT

The information contained in the foregoing job description, for the most part, lacks sufficient detail to make possible a thorough analysis of the radar observer's job in combat. However, general estimates can be made of the importance to combat success of the abilities judged to be necessary for success in training. While these estimates are based upon all evidence available, they are advanced with full recognition of their tentative nature. No new abilities are introduced; the evidence at hand seems inadequate to justify this step.

Set Operation

Set operation formed a relatively larger part of the radar observer's task in combat than in training since he had fewer other responsibilities and was continually called upon to furnish position information. On the other hand, it is probable that skill of this sort had become semi-automatic through continued practice by the time the observer reached his overseas station. The correction of set malfunctions and set operation under atypical weather condition, however, became correspondingly more important. It seems likely that these changes tended to deemphasize the importance of memory I (rote memory), verbal comprehension, and psychomotor precision and increase emphasis upon mechanical experience and "scientific background."

Scope Interpretation

Identification of returns was difficult in all theaters. However, the the DR navigator afforded a check upon the radar observer in this respect since he plotted the radar fixes, computed successive ground speed and tracks, and had the radar observer retake fixes which were inconsistent with previous position data. In general, it appears that scope interpretation was equally as important and required the same abilities in combat as in training. It will be remembered that the factors thought to be most important for mastering scope interpretation in training were perceptual speed, memory II (visual memory), and space I (spatial relations).

Navigation

The radar observer shared his navigation tasks with the DR navigator so that his work on navigation legs of missions was limited mostly to set operation and fix-taking. In this respect, combat missions probably required less planning of activities, which is thought to entail "systematic diligence," reasoning I (general reasoning), and "scientific background," and less computational work, thought to require numerical facility and reasoning I and little wind calculation, thought to require space I. It is likely that individual differences in reading dials, scales, and the range marks on the PPI scope were less marked among radar observers in combat than in training. This would act to decrease the relative importance of numerical facility, space I, and length estimation. Verbal comprehension should also prove to be less important in combat than in training since it was ininvolved primarily in learning the navigational procedures.

Bombing

Aside from the emotional stress, the combat bomb run was closely simulated by training bomb runs. A relatively unimportant difference was that in combat, the radar observer often obtained necessary ground speed and altitude information from other crew members without having to determine it himself. This may have decreased the complexity of bombing duties required of the radar observer and hence simplified his planning problem. This would decrease the importance of reasoning I, "scientific background," and "systematic diligence" which are thought to be important in organization of duties. As in the case of navigation, we might expect overlearning to have minimized individual differences in reading dials and scales during the combat bomb run. However, the increased stress of the bomb run may have caused these differences to be reemphasized. The factors involved in scale reading, it is believed, are numerical facility, "scientific background," and space I. Identification of aiming points for killing drift and rate was probably more difficult in combat than training, although even in training it was customary to bomb unfamiliar targets. This task is thought to require perceptual speed, memory II, and space I.

Summary

In summary, it appears that the factors involved in the correction of set malfunctions and in operating the set under typical conditions, namely mechanical experience and "scientific background" and in interpreting the scope, perceptual speed, memory II (visual memory), space I (spatial relations), represent the most important abilities required of the successful radar observer in combat. Factors involved primarily in carrying out navigational computations, numerical facility and reasoning I (general reasoning), and in organizing the separate duties on the navigation legs and bomb runs, "systematic diligence," and reasoning I, probably become less important in combat than in training. Factors involved primarily in learning set operation, navigation, and bombing procedures, such as verbal comprehension, and memory I (rote memory) are thought to become relatively unimportant in combat.

It has often been pointed out that aircrew members were under great emotional stress in combat. Although no adequate test has been found to measure emotional control, it is undoubtedly true that it is more important in combat than in training.

THE JOB OF THE RADAR OBSERVER IN THE FUTURE

Trends in Equipment Development

The recent developments in airborne radar equipment indicate probable changes in the tasks required of the radar observer. The most important of the equipment trends as illustrated by the AN/APQ-13 and later sets are briefly as follows:

(1) Greatly increased definition which means that terrain features will be presented in greater detail on the PPI scope.

(2) Automatic solution of navigation problems by electronic computers integrated with the radar set. These computers will yield such information as track, ground speed, and drift.

(3) Computers permitting extensive use of offset bombing. These computers will make adjustments automatically in the bombing variables to permit aiming on any identifiable return at a known distance and bearing from the target without restricting the direction of approach to the target.

(4) Auxiliary systems which replace the bombsight in coordinated bombing or which provide for mechanically synchronizing the bombsight with the movement of returns across the PPI scope. By means of these units, the radar observer can kill rate, now accomplished by calling successive sighting angles, and drift by making adjustments to keep the target under appropriate markers on the PPI scope. Other

bombing aids include automatic sweep expansion by means of which returns are kept near the edge of the scope and automatic sector scan which reduces the presentation on the PPI scope to a sector which includes the target area.

Indicated Changes in the Radar Observer's Job

It is not clear what effect the increased definition obtained by new radar equipment will have upon the task of scope interpretation. Because more details of the terrain are presented on the PPI scope, it would seem that the problem of identifying check points, target areas, and aiming points will be somewhat simplified. However, experience in the Pacific theater with the AN/APQ-7 does not confirm this. While increased definition yields a picture which approaches an aerial photograph in terms of complexity or number of elements, the increase in complexity is not accompanied by an increase in the number and kind of cues that are available for distinguishing the elements from each other. The resulting picture, over most target areas, consists of an extremely complicated pattern of returns that differ from each other only slightly in terms of size, shape, and brightness.

The bombing identification problem will undoubtedly be simplified by the availability of simple offset bombing methods. The radar observer will not have to detect the aiming point or estimate its position in the target area. Instead he will be able to select any nearby sharp return which he can identify on his map, set up the automatic offset bombing computer, and aim on the sharp return. In general, however, it seems likely that the skills required in scope interpretation will retain their importance for the future radar observer. Automatic solution of navigation problems will greatly reduce the computational skills required of the radar observer. He will simply have to manipulate the controls on the computers and radar set, tracking returns across the scope. Use of the E-6B computer will be greatly reduced. Map-plotting will be reduced and simplified. Less over-all understanding of the relationships between the various navigational variables will be necessary. However, the addition of computers and auxiliary units to the basic air-borne radar set will increase the number of controls which the radar observer will have to use and the number of dials he will have to read and interpret.

In summary, it seems likely that the skills required in carrying out present radar navigation duties will decrease in importance and that skills similar to or identical with those involved in present set operation or radar bombing procedure will become more essential.

SUMMARY

The chapter is introduced by a discussion of the purpose and methods of describing a job. The distinction is made between a "job description," which is a description in terms specific to the job, and a "job analysis" or description in terms of general abilities.

Following a brief introduction to airborne radar equipment, the job of the radar observer in training is described under four headings: set operation, scope interpretation, navigation, and bombing. Each section is presented as a job description followed by a job analysis. The job analysis is made primarily in terms of factors isolated in factor analyses carried out in the Aviation Psychology Program. Where these factors are inadequate to account for the skills presented in the job description hypothetical abilities are called upon.

The description of the radar observer's job in training is concluded by a summary of the job analysis sections. In this summary estimates are made of the relative validities of the various factors for predicting success in radar observer training. Factors predicted to have the highest validity were "scientific background," space I (spatial relations), and memory II (visual memory). Those predicted to have relatively high validity were "systematic diligence," numerical facility, reasoning I (general reasoning), perceptual speed, and verbal comprehension. Factors estimated to have low validity were length estimation, psychomotor precision, and mechanical experience.

The job of the radar observer in combat is described for the European and Pacific Theaters of Operation. The combat job is briefly analyzed by comparing it with the analysis of tasks required in training. It is predicted that the following factors would have relatively higher validity for predicting success in combat than in training: mechanical experience, "scientific background," perceptual speed, memory II, and space I. Numerical facility, reasoning I, "systematic diligence," and verbal comprehension are estimated to have relatively lower validity for combat than for training.

A final section of the chapter summarizes recent trends in the development of airborne radar equipment. These trends are interpreted as indicating that, in the future, skills involved in present radar navigational tasks will become less important while skills important to present set operation and radar bombing will become more important.

CHAPTER FIVE

The Development of Printed Proficiency Tests [1]

INTRODUCTION

The Radar Project provided the radar observer schools with batteries of standardized proficiency measures including both printed tests and performance checks. It is the purpose of this chapter to describe the development of the printed tests. The chapter which follows this one will describe the development of the performance checks.

Before discussing the specific use of proficiency tests in the radar observer training program, the functions served by standardized measurement in such a program may be outlined briefly. Measurement under standard conditions yields, of course, more reliable results than instructor ratings of students and informal classroom quizzes. The benefits resulting from this greater reliability are many and warrant the effort required to initiate and supervise the administration of a standardized proficiency measurement program. Most important of these benefits is the possibility of maintaining constant and uniform standards for graduation from training. The reliable grades assigned students at graduation have the further effect of making possible more effective selection for subsequent operational assignments. In addition, such grades serve as a more reliable criterion against which to validate selection tests. Finally, reliable course grades constitute a criterion for use in studies of instructor effectiveness, training methods and devices, and student differences and weaknesses. Reliable grades assigned at intermediate stages in the training provide valuable information about student progress. One incidental adva. tage of the standardized measurement program is the basis it provides for promoting the attitude among students that their efforts are being accurately and fairly assessed.

In the outline of the development of printed tests which follows, it will first be pointed out where printed tests were particularly ap-

[1] This chapter was written by Sgt. Norman Graff with the assistance of Cpl. Harold Kelley and Sgt. Albert Hastorf.

plicable and, hence, were used in radar observer training. For each of the four classroom subjects of the course, set operation, radar navigation, radar bombing, and radar intelligence, printed tests served either or both of two functions: (1) Measuring proficiency in solving problems and (2) measuring verbal knowledge of technical information. The construction of printed tests will be discussed with reference to sources of information for items, preparation of items, provisions for subjecting items to expert criticism, weighting the several sections of a test, determining time limits, scoring formulas, and format. The procedures described for revisions included systematically compiling and reviewing criticisms, and statistically analyzing existing forms. The extent to which uniform testing was achieved is outlined in a discussion of standardized testing conditions. Certain difficulties encountered in using standardized proficiency tests are reported and interpreted as resulting from rapid equipment developments and curriculum changes in the newly-formed radar observer program. The remainder of the chapter describes eight representative printed proficiency tests administered in radar observer training. Each of these is described in terms of position in the course, subject matter tested, relation to earlier achievement tests, specific details of content and construction, and data from statistical analysis.

USES OF STANDARDIZED PROFICIENCY TESTS IN RADAR OBSERVER TRAINING

The several phases of radar observer training presented different problems in constructing proficiency measures. It was realized that the skills taught in some parts of the training could be adequately measured only by performance checks. Information and skills taught in other sections, on the other hand, could be measured by printed tests. In general, material taught in the classroom, which comprised about half of the 10-week radar observer curriculum, provided the basis for constructing tests. To measure acquisition of the technical information which made up a large part of the course, tests were constructed consisting wholly of verbal questions. Other tests were constructed with problem-solving items to measure the computational skills taught in other phases of the training.

Classroom instruction included four topics: set operation, radar navigation, radar bombing, and radar intelligence. Classroom instruction in set operation included the location and function of the controls and units of the air-borne radar set. Specific operating procedures included starting and tuning the set, calibrating the range unit, maintaining maximum definition under varying conditions, and locating set malfunctions. Such technical information provided a readily-available source of printed test items. Consequently, the set operation proficiency tests consist of information items testing verbal knowledge of functions and procedures. It was recognized, however, that discrete items, concerned with procedure present the student with an artificial situation. Such items require the recall of an element of procedure quite outside of the usual context where it is recalled as part of an actual operating sequence. Also artificial are the verbal descriptions of symptoms of set malfunction used in questions framed in such terms as "What should you do when ______________________?" It is entirely posible that a student could respond correctly to verbal descriptions of a malfunction and yet not be able to recognize it when actually operating the set. To measure set operating ability under less artificial conditions, the bench set performance checks and certain items in the aerial and supersonic checks, described in Chapter 6, were developed.

The radar navigation courses consisted, for the most part, of the solution of navigation problems using maps, the E-6B computer, Weems plotter, and dividers. The recognition of typical radar returns was also taught, as well as the application to the task of navigation of information obtained from the set. The radar navigation proficiency tests included both problem-solving and information items, primarily the former. The problems presented in these tests included airplot solutions of wind problems, determination of wind, heading, and track on the vector face of the E–6B computer, and solution of time-rate-distance and altitude problems with the slide-rule face of the E–6B computer. In addition to solving such problems as were included in these tests, it was felt desirable, also, to measure proficiency in navigating typical missions which required the student to obtain necessary data from navigation instruments and organize the use of these data. Supersonic trainer and aerial performance checks were constructed to measure performance on such missions.

In radar-bombing classes, students were taught theory of bombing, radar-bombing procedures, and the set operation procedures involved in radar bombing. Knowledge of these subjects was measured by multiple-choice technical information items. The application of formulas and principles to determine variables used in bombing was also taught; measurement, here, required the use of problem-solving items. Such items dealt with computing radius of turn, determining necessary drift corrections, and predicting absolute altitude over the target.

Again, however, the procedure on an actual bombing run is quite different from solving discrete items and providing answers to verbally presented problems. The radar observer's basic source of information on a radar-controlled bombing run is a complex pattern of returns on the PPI scope which moves across the scope and becomes increasingly detailed. This moving complex pattern can be represented only very roughly by a series of scope photographs in a printed test. Another important characteristic of actual bombing runs is that the student must integrate the various tasks essential to the bomb run into a smoothly functioning system which permits him to complete the multitude of tasks in a short time. Such integration of tasks probably cannot be tested by discrete test items. To measure these aspects of bombing proficiency which could not be measured by printed tests, the bombing sections of the supersonic trainer and aerial performance checks were developed.

The radar intelligence classes consisted of the presentation of factual information regarding radar countermeasures, priority targets, survival techniques, and target studies. This phase of radar training was probably measured adequately by the single type of proficiency measure used: the printed test comprised of technical information items.

On the basis of the foregoing considerations, a battery of printed tests and performance checks was prepared for the radar observer proficiency measurement program. This program was of interest in that it represented the first time in aircrew training that a complete battery of proficiency measures presented by a psychological research organization was approved and adopted for use in all the training stations of an aircrew specialty. The six performance checks included in this battery are described in chapter 6 of this report. Five printed tests completed the battery. These included a test given at an intermediate stage of training for each of the four subjects in the course and a final comprehensive examination sampling all four of the subjects. The intermediate tests were titled as follows: Set Operation Intermediate Test, Radar Navigation Intermediate Test, Radar Bombing Intermediate Test, and Radar Intelligence Intermediate Test. The comprehensive examination consisted of two test booklets called Final Test I and Final Test II, which were administered during a single testing period.

GENERAL PROCEDURES USED IN DEVELOPING PRINTED TESTS

Preliminary Drafts

The first step in constructing printed proficiency tests consisted of collecting the subject matter from all possible sources for each of the four subjects of the radar observer course: set operation, radar navigation, radar bombing and radar intelligence. Course outlines, individual lecture outlines, and lecture notes of personnel from the Radar Project enrolled in the course[2] were assembled. Information from

[2] Radar project personnel who completed requirements of the radar training curriculum were: Lt. William A. McClelland, Lt. George S. Klein, Tech. Sgt. Sanford J. Mock, St. Sgt. Richard T. Mitchell, Sgt. Nathaniel L. Gage, Sergeant Graff, Cpl. Nelson R. Nall, Cpl. Wilbert H. Schwotzer, and Pfc. Dwane R. Collins.

these sources was assigned to one of the four categories of subject matter and items were tentatively drawn up for elements of information in each category. Existing tests were examined and some items were revised and retained.

The question of the relative weight to be assigned to each of the four subjects in the course was decided by training authorities. The assigned weights were followed in constructing the tests by apportioning more items to one subject than another. A difficulty in the use of this method, arising primarily in the construction of technical knowledge sections, was the lack of sufficient test item material for the more heavily weighted topics. This was especially true of two subjects, radar bombing and radar navigation, which required many instructional hours in the course for the presentation of testable subject matter. Less difficulty was encountered in expanding problem-solving sections to increase their weight. Emphasis was placed on the important navigation and bombing techniques by presenting each problem a number of times but with altered data from item to item.

Review and Criticism

After preliminary drafting, all items were submitted to radar experts for review and criticism. First, the individual items were reviewed by instructors who had prepared the lectures from which item material was taken. Revisions were made on the basis of resulting criticism and a first draft of the proposed test was constructed. This first draft was then presented to an assembly of all instructors teaching the pertinent subject. It was later found to be more efficient to submit proposed tests to school authorities rather than to instructors. This had the advantage of obtaining criticism from individuals not involved in teaching the source material and the administrative criticism insured an acceptable relation of test content to graduation standards.

The methods just described for obtaining criticism were sufficient for technical information items. For problem-solving items, it was necessary to obtain further judgments from expert radar observers regarding correct answers and desirable tolerances for incorrect alternatives. To facilitate obtaining these decisions, a system was developed which made use of three panels of five experts each. The members of each panel individually solved the test items after which the five experts compared results and agreed upon answers and mislead tolerances. When the decisions of the three panels were compared, they usually reflected sufficient agreement to confirm the adequacy of the separately determined values. As a final check to insure accuracy, the Radar Project test construction teams carefully reworked each problem. An incidental benefit derived from the panel system was that the working time required by the experts in the original solution of the items provided an approximate time limit for the first administration of the test.

Test Format

Because the alternative choices in both the problem-solving and technical information items were brief and required little space, items were arranged on the pages of the test booklets in two columns. This accomplished a saving in space and gave a desirable compactness to each item. Answers were marked on a separate answer sheet which allowed for machine scoring. In assembling the various sections of each test, timed problem-solving sections—the only sections requiring the use of equipment—were placed first in the test booklets. This arrangement facilitated proctoring by making it difficult for students to work unnoticed on problem-solving sections after time limits had elapsed. It also facilitated administration since few students finished the speeded sections before the time limits were up.

Time Limits

It has been mentioned that time limits for initial administrations of problem-solving sections were determined on the basis of the time required by panels of experts to solve the problems. These time limits were used on the assumption that they would allow very few students to complete the problem-solving sections. In assigning time limits for technical information sections, on the other hand, the attempt was made to allow all students to try every item. The practice of speeding the problem-solving sections met with resistance from some school authorities, who thought that tasks measured in these sections were not solved under time pressure in actual radar bombing and navigation. Whereas some authorities felt that accuracy, not speed, was most important in solving bombing and navigation problems, others believed that speedy, systematic handling of the problems was paramount. Among members of the Radar Project it was felt, on the basis of observations of trained radar observers at work, that both speed and accuracy should be measured. Another justification for speeding some sections was that a section scored only in terms of accuracy would provide less differentiation among students than would the same section given with an appropriate time limit. The latter would discriminate between those students who require a great deal of time to get accurate answers and those who work both quickly and accurately.

Scoring Formula

All tests were scored simply in terms of the number of correct responses. The raw scores were converted to centile scores by means of conversion tables constructed from the data accumulated from a number of classes. No attempt was made in the scoring formula to correct for guessing. The use of scoring formulas for this purpose would have been complicated because not all items had the same number of alternative responses, the number ranging from two to five.

Bases for Revisions

The urgency of installing the proficiency testing program made it impractical to administer any of the tests experimentally prior to their publication. Consequently, first forms of the printed tests were put into use without being subjected to the customary statistical analysis to detect item defects. Revisions were based upon statistical analysis of existing forms, systematically gathered subjective criticisms, and changes in the curriculum.

Inspection of the means and standard deviations of the various sections of a test and of the test as a whole often indicated necessary revisions. For example, these statistics for a speeded problem-solving section of a test, part A of Final Test I, indicated that this section did not contribute to the total test score in proportion to the time spent in its administration. To increase the number of items in the section wihout adding to its testing time, existing items which required intermediate steps for their solution were broken down into separate items, each requiring an answer. Also, answers which were given in terms of two independent quantities were divided between two items. For example, instead of asking in a single item for a wind solution in terms of wind direction and wind force, the student was asked in one item for wind direction and in another for wind force. The results of using the two methods on part A of Final Test I are described on page 81.[3] The technique of making separate items out of the intermediate steps in the solution of a complex problem was first employed by Psychological Research Project (Navigator) where 13 problem-solving items were expanded to 119 items.

The distribution of raw scores for speeded sections also served as a guide for determining the adequacy of time limits. The distributions from early administrations of most of the speeded sections were negatively skewed, most of the scores being concentrated at the upper end of the scale. This lack of discrimination among students at the upper levels of proficiency was taken to mean that the time limits were too long. To increase differentiation among the higher scoring students, the time limits were shortened and the distributions became more symmetrical.

Item analyses in terms of difficulty level and internal consistency were employed to yield data for test revisions. Item difficulty was measured by the percentage of students attempting an item who also got it right. The correlation of each item with the total test score was found by computing a phi coefficient. This coefficient was based

[3] In table 5.8 are presented the means and standard deviations of part A of Final Test I before and after splitting two-answer items into two separate items.

upon the relative percentages getting the item right in the upper and lower 50 percent of the group divided on the basis of total test score. Items which were shown to be too difficult or too easy were eliminated or revised, as were items having zero or negative correlation with the total test score. In some instances, however, such items were retained because of the importance of their content to an adequate sampling of the subject matter.

An attempt was made to compile systematically the criticism made by instructors and students during the period of administering a given form. These criticisms, usually dealing with item clarity or the adequacy with which a test sampled the subject matter, were considered and items were adjusted where necessary. Changes and developments in airborne radar equipment resulted in changes in the subject matter taught in the course, primarily in set operation. These changes frequently made it necessary to delete obsolete material and incorporate new material in the tests.

STANDARDIZATION OF TEST ADMINISTRATION

Several steps were taken to promote the standardized test administration which is essential to the success of a proficiency evaluation program. The examiners who administered the tests were selected for the task by the school authorities. They were given indoctrination and instruction in methods of test administration by personnel of the Radar Project. Detailed standardized directions for administration were provided for each test. These directions prescribed approved procedures for distributing and collecting test materials and timing the various sections of the test, and included directions to the students which the examiner read verbatim. All testing was carried out in specially designated rooms at each training station. Adequate working space was provided each student and necessary precautions were taken to insure independent work during the test. All tests were scored under the immediate supervision of the Radar Project, and conversion tables and rosters of raw and converted scores were prepared for the training authorities.

As in all testing situations, motivation played an important part in the test results. Radar observer students had already received their commissions and aircrew specialty ratings either as bombardiers or navigators and it appeared that many of them did not care whether or not they graduated from radar observer training. Illustrative of the motivation problem is the following incident which occurred after the cessation of hostilities in the Pacific on 14 August 1945.[4] On 23 August, class 45-34 at Langley Field, consisting of 39 students, took the final examination and 13 students failed. The following week, the succeeding class, 45-35, which included six of the previous

[4] Statistical analysis of this problem was conducted by Cpl. Hyman Sofer.

failures, was promised a week-end leave for "passing" the examination. With this goal, only 2 men in 45 failed. The data for the two classes are presented in table 5.1. In addition, data for class 45–33

TABLE 5.1.—*Effect of motivation on raw scores from speeded and nonspeeded tests*[1]

Class	N	Final test I, P1a-B (speeded)		Final test II, P1b-B (nonspeeded)		Total of final tests I and II	
		Mean	SD	Mean	SD	Mean	SD
45-33 (pre-VJ-day)	31	62.55	10.92	69.16	9.25	131.65	16.12
45-31 (post-VJ-day, unmotivated)	39	51.31	10.65	67.41	7.92	118.97	12.53
45-35 (post-VJ-day, motivated)	45	61.38	11.34	(Not administered)			

[1] Based upon administrations at Langley Field.

are presented to show the level of performance for students who completed their training before the Japanese surrender. It will be noted that the specially motivated class attained the same level as the presurrender group and that the low score of the poorly motivated group was due primarily to poor performance on the speeded section of the test.

DIFFICULTIES EXPERIENCED IN THE USE OF STANDARDIZED PROFICIENCY TESTS

The usual difficulties encountered in installing and administering a battery of standardized proficiency tests were intensified by the urgency and acceleration of the radar training program. At the time the program was initiated, progress in the development of airborne radar equipment was rapid. Equipment improvements which required new operating techniques were being constantly introduced. These changes inevitably produced difficulties in applying a given standardized test to students over any considerable period of time. In several instances, to eliminate obsolete subject matter, it was necessary to delete items without replacing them. This was particularly true of the radar intelligence course.

Because of the changes in equipment, students were often trained on sets that differed from those for which proficiency tests had been constructed. Further, the training literature often lagged behind the appearance of the new equipment. The development of lectures explaining new equipment and the attainment of proficiency by instructors in new procedures was consequently slow. These things combined to yield inconsistencies between the subject matter taught and the content of the printed tests.

On several occasions, there was disagreement among the schools as to correct operating procedures and the emphasis that should be placed on various phases of the course. For example, one school trained its students in navigation with the view of graduating students who were prepared to take over the navigation function in emergency combat situations. Another school emphasized set operation and bombing and required, as the only navigation function, the plotting of fix information for the dead-reckoning navigator. At several schools students were trained to obtain winds by the target-timing method when this method did not appear in the curriculum of the other schools.

Unfortunately, training authorities tended to desire test items which could be passed by most of the students. This feeling arose from a conception of the minimum information required of a radar observer in combat and, apparently, also from a desire to make the results from a particular school seem to indicate a high level of proficiency. However, it was possible to include difficult items in the tests by explaining their function in discriminating among students at the upper levels of proficiency.

DESCRIPTION OF TYPICAL PRINTED TESTS USED IN THE RADAR TRAINING PROGRAM

Eight proficiency tests which were used in the radar training program are described below. Each test will be discussed in terms of its function in the training program, its development, item content, details of administration, important revisions, and statistical findings. First will be presented the four tests given at intermediate points in the training, measuring proficiency in each of the four subjects of the course. These intermediate tests, in order of administration in the course and presentation below, are: Set Operation Intermediate Test, Radar Navigation Intermediate Test, Radar Bombing Intermediate Test, and Radar Intelligence Intermediate Test. Next are presented the two parts of the comprehensive examination, Final Test I and Final Test II, administered at the end of the course. These are followed by a description of a comprehensive examination, final test for AN/APQ-7 set, prepared to measure proficiency on a recent airborne radar set. Finally described is the Navigation Proficiency Test, not a radar test, which was used as part of the selection battery given to candidates for radar observer training.

Set Operation Intermediate Test

The set operation intermediate test [5] consisted primarily of technical information items measuring knowledge of the operation of radar equipment, ski'' [illegible] analyzing equipment malfunctions, and knowledge of auxiliary equipment used in radar navigation and bombing. The various forms of the test contain approximately 100 multiple choice items and require about an hour for administration. A set operation

---

[5] First forms of these tests were developed by Sgt. Graff, and Sgt. Krieldt. Revised forms were prepared by S/Sgt. Mitchell and Sgt. Gerald S. Blum.

intermediate test was administered to radar observer students at approximately the middle of the ten-week course after the classroom study of set operation had been completed.

Most of the item material for the first form of this test was obtained from the radar operator achievement examination for AN/APS-15, prepared at the request of the Psychological Section, Medical Research Division, Air Surgeon's Office, by a project of the National Defense Research Committee. This test was given by the Aircrew Evaluation and Research Detachment No. 1 in the selection of pathfinder crews for the Eighth Air Force. Revisions were made by the Detachment and the resulting form, after supplementation by the radar project, became form A of Set Operation Intermediate Test, P1k-A and P1L-A.

Various forms of the test developed by the radar project contained from 76 to 120 items. None of the forms was speeded. Each required 50 to 75 minutes testing time. No variations from the general procedure of developing the test items were employed. Although most of the items are of the verbal knowledge type, some may be classified as problem-solving items, since they require the application of theory to the solution of problems.

Samples:

17. If there is no sweep or spot on the scope, the trouble is a blown fuse in the

- 17–A Synchronizer.
- 17–B Range unit.
- 17–C High voltage rectifier circuit.
- 17–D Low voltage rectifier circuit.
- 17–E Main control box.

18. Given:

Range unit "on."	Altitude 2.2 miles.
Sweep delay at 30 miles.	Computer drum reads 4.9.
Bombing circle on target.	

Find: What is the slant range to the target?

- 18–A 30.9.
- 18–B 32.7.
- 18–C 32.9.
- 18–D 34.9
- 18–E 37.1.

In the set operation intermediate tests developed for the Eastern Training Command schools, the booklets were divided into sections on the basis of the following areas of subject matter: normal operating procedures, locating malfunctions, and use of auxiliary equipment. For Western Training Command schools, no division into parts was made.

Revisions of set operation tests were necessitated primarily by the introduction of new sets and equipment into the training program. When these revisions were made, statistical analyses of previous forms were employed. In all, six set operation intermediate tests were developed. The first form produced by the Radar Project consisted of two sections combined into a single booklet, Plk–A and PlL–A for the AN/APS–15 and AN/APS–15A, respectively. This test was replaced by Form B of Pll–B when instruction on the AN–APS–15 was eliminated from the course. Two forms were constructed for the AN/APQ–13, Plm–A and V Plm–A. When the AN/APQ–7 program needed a set operation test, the Radar Project was so absorbed with meeting the demands for test revisions from other schools and with supervision of performance check administration that aid from the NDRC Project was requested. The temporary test produced, at that time, Radar Operator's Proficiency Examination for AN/APQ–7 was later replaced by radar project revisions, W Plt–A and W Plt–B, which incorporated new test items from the materials in the growing curriculum.

With one exception, the distributions of raw scores from the various forms were negatively skewed. The distribution statistics given in table 5.2 indicate that the tests were fairly easy.

TABLE 5.2.—*Means and standard deviations of distributions of raw scores from various forms of the set operations intermediate test* [1]

School	Classes	Form	Number Items	Time	N	Mean	SD
				Minutes			
1	All cases	VPlm–A	120	50	275	76.15	12.67
2	All cases	Plm–A	100	60	457	56.09	10.32
3	45–18 through 45–20	Plk–A and Pll–A	101	75	413	70.95	10.66
3	All cases	Pll–B	80	60	700	70.86	10.07

[1] School 1 is Victorville, 2 is Boca Raton, and 3 is Langley Field.

The odd-even reliability for form Plk–A and Pll–A is 0.60 which corrects by the Spearman-Brown formula to 0.74. This form has 104 items and a time limit of 75 minutes. The mean and standard deviation of the odd scores are 38.11 and 4.96 respectively. The mean and standard deviation of the even scores are 35.45 and 5.03 respectively.

Item analysis of difficulty level based on 434 cases for form Plm–A, which yielded an approximately normal distribution of raw scores, shows adequate item difficulty level. The median item difficulty for the upper half of the group was 68 percent passing and for the lower half was 49 percent. This form, which has 100 items with a time limit of 90 minutes, yielded a median phi of 0.10 when analyzed for internal consistency using the same 434 cases. Computation of phi was based upon the upper and lower 50 percent of the group.

Radar Navigation Intermediate Test

The radar navigation intermediate test[6] was designed to measure proficiency in solving typical radar navigation problems, and verbal knowledge of the techniques of radar navigation. The various forms of the test contain from 65 to 90 multiple-choice items and require from 65 to 90 minutes for administration. A navigation intermediate test was administered upon completion of classroom training in radar navigation, which was given during the first three-quarters of the course.

The materials for the items of the early forms were obtained by the radar project from existing instructors' quizzes, navigation manuals and lectures, and suggestions of aerial and ground school instructors. Helpful ideas for adapting navigation problems to multiple-choice items were gained from the proficiency tests developed by the navigator project.

Most of the items in the radar navigation intermediate tests were of the problem-solving type. Since navigation problems require the use of a rather complex set of data, it was arranged, in order to save testing time and space in the test booklet, to have two or more items based upon a single set of basic data. To simulate inaccuracies which commonly enter into navigation instruments and data, erroneous materials were occasionally introduced into the basic data. Also, it was felt that by incorporating erroneous information, the student would be required to use judgment comparable to the judgment he would employ in the air when rejecting a poorly plotted fix. These techniques are illustrated in the following sample item. It will be noted that item 28 is dependent upon items 26 and 27 insofar as the student must use the selected alternatives to the latter items to solve item 28.

Sample:

Some of the fixes are incorrect, consequently all six fixes should be plotted for each problem. The times for the first and last fixes are always accurate. Questions 26–28 are parts of the same problems.

Given: Time, bearing, and horizontal range for six wind run fixes.

102804, 199°, 14 n. m.
102950, 193°, 11 n. m.
103430, 071°, 8 n. m.
103515, 062°, 11 n. m.
103603, 056°, 14 n. m.
True heading, 212°.
True air speed, 152 K.

26. Find: True course.

26–A 038°.
26–B 042°.
26–C 208°.
26–D 218°.
26–E 222°.

---

[6] Developed by Sgt. Graff, Sgt. Kriedt, and Cpl. Koch.

27. Find: Ground speed.

27-A 179 K.
27-B 189 K.
27-C 199 K.
27-D 209 K.
27-E 219 K.

28. Find: Wind (use the GS and TC selected as correct for question 26 and 27).

28-A 016°/50 K.
28-B 021°/44 K.
28-C 051°/50 K.
28-D 051°/44 K.
28-E 056°/50 K.

Early forms included items on scope interpretation, radar unit functions, radar beacon navigation, and E-6B computer problems. Later forms were more definitely divided into separate sections. The most heavily weighted section was the theory and technique of radar navigation, course determination, and drift determination. In the various forms, this first section included from 28 to 72 items given in 30- to 70-minute unspeeded sessions. Next was presented a speeded section of from 15 to 18 items requiring 10 to 18 minutes and measuring ability to use the E-6B computer in determining wind direction and force, true heading, true course, ground speed, ground range, and ETA. Ten airplot problems similar to those described on page 80/for Final Test I were presented in another speeded section, with a time limit of 20 minutes. For the airplot section, this test incorporated a standard expendable Mercator chart the size of the test booklet upon which the student plotted his own latitude and longitude data. An additional 12-item section,[7] employed only in the Western Flying Training Command schools, was used for measuring ability in target timing wind determination. This speeded section had a time limit of 12 minutes.

Original forms of the radar navigation intermediate test emphasized the airplot method of determining winds as was done in the schools of the Eastern Technical Training Command. Important revisions were necessitated by the introduction in Western Flying Training Command schools of the target timing method of determining winds. Three forms were constructed for the eastern schools: P1h-A and P1h-B for the AN/APS-15 set and P1i-A for the AN/APS-15A set. Three more forms were constructed for the AN/APQ-13 used at the western schools: P1j-A and V P1j-A for use at Victorville, and W P1j-A for use at Williams Field.

---

[7] Preliminary work on this section was conducted at Boca Raton, Fla., by Sgt. Graff and Sgt. Krieldt. The latter completed the section in the Western Flying Training Command station at Williams Field.

The means and standard deviations for the part and the total scores of form P1h–B are presented in table 5.3. The extent to which each section contributed to the total score may be estimated from the standard deviations.

TABLE 5.3.—*Means and standard deviations of distributions of part and total raw scores from form P1h–B of the radar navigation intermediate test, administered at Langley Field*

Classes	Part	Number items	Time	N	Mean	SD
			Minutes			
45-9 through 45-27	I	45	35	709	31.38	7.37
45-9 through 45-27	II	15	10	675	8.70	2.85
45-9 through 45-27	III	10	20	675	4.63	1.92
45-9 through 45-27	Total	70	65	675	4[illegible].18	9.26
All cases	Total	70	65	959	43.24	8.84

Two reliability studies were made of form P1h–B. The first consisted of an odd-even reliability computed for part A, an unspeeded section containing 45 items with a time limit of 35 minutes. The odd-even coefficient based on 179 cases, Langley Field classes 45–19 through 45–33, is 0.61 which corrects by the Spearman-Brown formula to 0.76. The mean and standard deviation of the odd scores are 14.66 and 3.18, respectively. The mean and standard deviation of the even scores are 14.48 and 3.12, respectively. Odd-even reliability coefficients were not computed for the other two parts of this test because they were speeded.

The second reliability study consisted of determining a test-retest reliability for the parts and total score of form P1h–B. The results, presented in table 5.4 are based upon only one class of 27 students. The reliability coefficient for the total test was found to be 0.64.

TABLE 5.4.—*Test-retest reliability coefficients for part and total raw scores of radar navigation intermediate test, P1h–B* [1]

Part	N	Mean-test	Mean-retest	SD-test	SD-retest	r_{11}	$SE_{r_{11}=.00}$
I	27	27.09	32.74	6.97	3.31	0.59	0.20
II	27	8.06	11.52	3.40	2.23	.33	.20
III	27	3.33	5.55	1.05	2.10	.39	.20
Total	27	39.09	49.18	9.06	4.55	.64	.20

[1] Based upon Langley Field classes 45-17; 94 percent were bombardiers and the remainder were navigators. Test and retest were separated by an interval of 1 week.

Item analysis yielded the difficulty level and internal consistency statistics presented in table 5.5. For each part of form P1j–A the median difficulty level and the median phi coefficient are given for the upper and lower halves of the group divided on the basis of total test score. Statistics concerning difficulty level, which ranges from 49 to 87 percent passing, indicate that the items in both the technical-information section and problem-solving sections were answered correctly by most of the students. Of considerable interest in table 5.5 is the discrepancy in difficulty level between school 1 and school 2. The students from school 1 were mostly navigators, while the students from school 2 were mostly bombardiers. Despite the fact that most of the developmental work on this form of the test was conducted at school 2, its students consistently scored lower than students from school 1.

TABLE 5.5.—*Median difficulty level for upper and lower scoring groups and median phis for the parts of radar navigation intermediate test, P1j–A* [1]

School	Part	N	Median difficulty upper 50 percent [2]	Median difficulty lower 50 percent [2]	Median phi [3]
1	I	1,050	87	70	0.19
2	I	526	61	49	.11
1	II	1,050	83	72	.10
2	II	526	77	52	.14
1	III	1,050	68	61	.13
2	III	526	63	49	.11

[1] Data from school 1, Victorville, are from classes 45-12 through 45-27. Data from school 2, Boca Raton, are from classes 45-7 through 545.
[2] Based upon the number attempting the item.
[3] Based upon the upper 50 percent versus the lower 50 percent.

Radar Bombing Intermediate Test

The radar bombing intermediate test [8] consists both of technical knowledge and problem-solving items measuring knowledge of the theory and techniques of radar bombing. The various forms of the test, which include from 52 to 77 multiple-choice items, are not speeded and require somewhat less than an hour to complete. A radar bombing intermediate test was administered to radar observer students near the end of their training after the lectures covering radar bombing had been completed.

The radar bombing test was developed entirely by the Radar Project, its only antecedents being daily quizzes based upon bombing lectures. Item material was also garnered from bombing manuals, lectures, and recommendations of instructors and supervisors. Items were distributed randomly with respect to subject matter coverage which included such topics as formation bombing, off-set bombing, coordinated bombing procedures, set operation during the bombing run, drift determination for collision course to target, and procedure turns.

All forms of the test except one were without part divisions; the exception was a two-part test developed for use in the Western Flying Training Command schools. The problem-solving items were constructed in interdependent groups. Care was taken in selecting mislead values for the problem-solving items to make them close enough to the correct value to reduce to a minimum the likelihood that a student could do careless work and still select the right answer.

[8] Developed by Sgt. Graff, Sgt. Kriedt, and Cpl. Koch.

Sample: Interrelated problem-solving items. Questions 1–6 are parts of the same problem.

Given:

True course to I. P.	085°.
True course from I. P. to target	035°.
True air speed	170 K.
Wind	230°/70 K.
Absolute altitude	24,320 feet.

1. Find: Radius of turn.
   1–A 3.0.
   1–B 3.2.
   1–C 3.4.
   1–D 3.6.
   1–E 3.8.
2. Find: True heading to I. P.
   2–A 093°.
   2–B 095°.
   2–C 097°.
   2–D 099°.
   2–E 101°.
3. Find: True heading from I. P. to target.
   3–A 024°.
   3–B 026°.
   3–C 028°.
   3–D 030°.
   3–E 032°.
4. Find: Wind correction vector.
   4–A 1.2.
   4–B 1.4.
   4–C 1.6.
   4–D 1.8.
   4–E 2.0.
5. Find: Turn allowance (horizontal range).
   5–A 1.9 n. m.
   5–B 2.1 n. m.
   5–C 2.3 n. m.
   5–D 2.5 n. m.
   5–E 2.7 n. m.
6. Find: Turn allowance (slant range).
   6–A 4.1 n. m.
   6–B 4.3 n. m.
   6–C 4.5 n. m.
   6–D 4.7 n. m.
   6–E 4.9 n. m.

Technical information item.

25. With azimuth stabilization ON, the target drifts to the left. The radar observer should correct the pilot to the—

25–A right and move the track line to the right.
25–B right and move the track line to the left.
25–C left and move the track line to the right.
25–D left and move the track line to the left.
25–E left and not move the track line.

Original forms of the radar bombing intermediate tests included P1e–A for the AN/APS–15, P1f–A for the AN/APS–15A, and P1g–A for the AN/APQ–13. Curricular revisions, local school requests, and the results of item analyses created the need for several modifications. Form P1e–B was a revision for the AN/APS–15. P1g–A, applicable to the AN/APQ–13 set which was used in the Western Training Command schools, was revised when the E–6B computer solution of procedure turns was introduced into the curriculum and given considerable emphasis. The resulting test had three alternate or revised forms for use at Victorville, V P1g–A, V P1g–B, and V P1g–C, and W P1g–A, for use at Williams Field. The problems in this test required the inclusion of two tables giving data for bombing computations.

The means and standard deviations for three forms of the radar bombing intermediate test, each given at one of the three largest training schools, are presented in table 5.6.

TABLE 5.6.—*Means and standard deviations of distributions of raw scores from various forms of the radar bombing intermediate test*

Form	Number items	Time	N	Mean	SD
		Minutes			
P1e–B	55	40	829	41.98	4.97
P1g–A	57	40	568	31.21	4.82
VP1g–B	65	40	200	46.88	6.57

The results of item analyses of difficulty level and internal consistency for two forms of the radar bombing intermediate test are presented in table 5.7. The difficulty level statistics indicate that the

TABLE 5.7.—*Median difficulty level for upper and lower scoring groups and median phis for two forms of radar bombing intermediate test* [1]

School	Form	N	Median difficulty upper 50 percent	Median difficulty lower 50 percent	Median phi [2]
1	P1g–A	846	74	48	0.11
2	P1g–A	410	71	53	.15
3	P1e–B	754	90	76	.18

[1] Data from school 1, Victorville, are from classes 45-16 through 45-27. Data from school 2, Boca Raton, are from classes 45-5 through 45-35. Data from school 3, Langley Field, are from classes 45–11 through 531.
[2] Based upon the upper 50 percent versus the lower 50 percent.

items were easy since a high percentage of students answered the items correctly.

The odd-even reliability computed from 180 cases for form P1c–B is 0.42 which corrects by the Spearman-Brown formula to 0.59. This form has 55 items and a time limit of 40 minutes. The mean and standard deviation of the odd scores are 20.12 and 2.56, respectively. The mean and standard deviation of the even scores are 21.86 and 2.81, respectively.

The reliability of 0.59 is lower than an odd-even reliability of 0.75 found for part B of Final Test II, P1b–A, which consists mainly of radar bombing items similar to those in the intermediate test.[9] A value of 0.69 is shown in the same table for a similar section of a later test, part B of Final Test II, P1b–B. The lower reliability of the intermediate test may be partially explained by its shorter length.

Radar Intelligence Intermediate Test

The radar intelligence intermediate test[10] consists wholly of technical information items measuring verbal knowledge of scope interpretation, radar countermeasures, scope photography, current theatres of operation and radar targets, escape and survival techniques, and mission briefing and interrogation. The various forms contain from 83 to 120 multiple-choice items and require from 75 to 85 minutes for administration. The test was usually administered near the end of the course.

The radar intelligence tests were constructed wholly on the basis of lecture material and classroom quizzes already in use. Three radar intelligence tests were developed before rapidity of change from one area of bombing operations to another forced the abandonment of this test in the proficiency battery. P1n–A and P1n–B were developed for the schools using the AN/APS–15 and AN/APS–15A sets and P1p–A for the AN/APQ–13 schools.

Means and standard deviations obtained for each class taking the various forms of the radar intelligence intermediate test are not presented because the constantly changing subject matter render the results meaningless. After several administrations of a given form, the particular subject matter tested was often found to have been eliminated from the radar intelligence curriculum.

One phase of the radar intelligence curriculum for which no adequate proficiency measures were developed is scope interpretation. Repeatedly, aerial instructors and students found that errors in radar bombing accuracy were due largely to failure to correctly interpret the returns on the PPI scope. However, the complexity of the returns presented on the scope and particularly their movement and increasing

---

[9] See table 5.12 for reliability statistics of Final Test II, P1b–A.

[10] These tests were prepared by Sgt. Graff with the supervision and aid of Sgt. Kriedt.

complexity on the bombing run, appear to limit the extent to which printed tests can measure scope interpretation. It is probable that aerial and supersonic check items can be developed to measure this important function more adequately. An additional testing technique which promises to yield more adequate measures of scope interpretation ability is the motion picture test.

Final Test I[11]

Final Test I is a radar navigation test consisting wholly of speeded problem-solving items. It measures proficiency in use of the E-6B computer, the Weems plotter, and the dividers, with emphasis upon air plot wind determination, and navigation on the basis of ground positions obtained from photographs of the PPI scope. The test consists of 4 parts with a total of 85 or 105 items, depending upon the form, and requires 145 minutes for administration. It is the first of two test booklets which comprise the final examination given to students upon completion of radar observer training.

The chief source of items in Final Test I was sections 5 and 6 of the Radar Operator Achievement examination, prepared by the National Defense Research Committee and used by the Aircrew Evaluation and Research Detachment No. 1 in the Eighth Air Force.[12] The AERD No. 1 revision of section 5, dealing with position and direction of flight, was used in part B of Final Test I, in which the student determines ground position from simulated scope photographs. Section 6 of the NDRC tests, a navigation problem, was revised by AERD No. 1 and became the simulated mission presented in part A of Final Test II. The revisions carried out by AERD No. 1 were toward closer simulation of the materials used in actual radar navigation and consisted of employing an actual combat zone mercator map and simulated scope photographs, the latter being more realistic than the NDRC mimeograph presentation of the PPI scope. The motivational effects of these devices were apparently great, since students taking the test seemed to feel that they were practicing a mission they might fly some day over Germany. In its revision of the AERD No. 1 version, the radar project supplied misleads for the mission and scope plotting sections by the panel system described on page 61. Several of the test photographs were improved and the E-6B computer and airplot sections were edited and expanded. The simulated mission was placed first in the test booklet to take advantage of the greater student interest in this section.

Part A has 25 items in form P1a-A and 45 items in form P1a-B. Both forms require 70 minutes to administer. Part A consists of a simulated radar bombing mission over Germany, the separate items

[11] Developmental work on this test was conducted by: Capt. Horace R. Van Saun, Sgt. Graff, Sgt. Philip H. Kriedt, and Cpl. John F. MacNaughton.
[12] See chapter 10.

of which call for solution of navigation problems that would typically be met on such a mission. The student is provided with an expendable mercator map of the area over which the bombing mission is flown and a book of 10 simulated scope photographs representing the scope at given times during the mission. Using the E-6B computer, Weems plotter, and dividers, the student is required to determine ground position from the photographs, compute winds, determine new headings to make good a given course, compute estimated times of arrival at various points, and solve similar basic navigation problems. The speeded items, in conformity with usual test construction practice, increase in difficulty through the test and are necessarily interrelated as is evident in the following item:

Sample: A problem-solving item related to an earlier item.

19. In order to avoid flak around Munster you alter course and fly directly from the last ground position (questions 13 and 14) to 52°10′ N, 07°50′ E. Using the wind in questions 17 and 18, which is closest to the wind that you actually computed and the same true air speed of 215 K, find the new true heading.

19-A 052°.
19-B 056°.
19-C 060°.
19-D 164°.
19-E 068°.

The responses to many items in form P1a-A contained two independent variables. In form P1a-B such items were split into two, each having single-variable responses. This change, which amounted to increasing the number of items in part A, was most commonly made in items requiring wind information and coordinate readings.

Sample: Old form:

5. Photo No. 5 shows the screen at 0957. What is the wind?

5-A 345°/50 K.
5-B 350°/44 K.
5-C 360°/57 K.
5-D 005°/42 K.
5-E 010°/64 K.

Sample: New form:

7. Photo No. 2 shows the scope at 0957. What is the direction of the wind?

7-A 165°.
7-B 175°.
7-C 335°.
7-D 345°.
7-E 355°.

8. What is the wind force?

8–A 10 K.
8–B 19 K.
8–C 57 K.
8–D 67 K.
8–E 77 K.

Tolerances were determined by a panel of instructors and by test construction teams. Some of the considerations were accuracy of plotting and of use of computing equipment. Included in alternative choices were answers based on reciprocal plots, careless settings and inaccurate readings of the E-6B computer, and common misunderstanding of required technique. It was believed that the simulated mission requires much the same selection and integration of data that is required by actual radar navigation. The simulated mission is to be contrasted with the remaining parts of Final Test I, which, like the usual test, require solution of independent items.

Part B consists of 20 items and has a time limit of 30 minutes. Each item requires the student to determine ground position from a simulated scope photograph. The radar returns appearing on the 20 photographs represent areas on the Mercator map used for part A. A dead-reckoning position is given with each photograph which is within approximately 50 miles of the represented ground position. The student must identify one or more returns on the photograph, take a fix on these returns, and plot it accurately on the map. The answer is recorded by selecting the one of five pairs of latitude and longitude values that corresponds most closely to the coordinates of the plotted point. Selection of the correct pair by elimination was made unprofitable by the enforcement of a severe time limit. Several conditions govern the selection of misleads. Mislead values were chosen which were close enough to the correct values to measure plotting accuracy. At the same time, however, misleads were given values such that reasonable plotting deviations would not force a student to select at random one of two choices equidistant from his plotted point. Another consideration was that a degree of latitude, as represented on the Mercator map, was 60 percent longer than a degree of longitude. Because of this scale difference, a vertical plotting error of, for example, 2 minutes represented poorer performance than a horizontal error of 2 minutes. On this basis, for each item, the longitude values in the misleads differed from the true longitude value slightly more than the mislead latitude values differed from the true latitude.

Sample: Scope plotting, part B (Each question is supplemented with a scope photograph.)

46. Approximate position at time of photo 46:
51° 40′ N, 06° 20′ E.

Find: Ground position at time of photo 46.

46-A 51°13′ N., 05°24′ E.
46-B 51°25′ N., 05°30′ E.
46-C 51°32′ N., 05°40′ E.
46-D 52°20′ N., 07°00′ E.
46-E 52°30′ N., 06°50′ E.

Part C consists of 20 items requiring 12 minutes for form P1a–A and 15 minutes for form P1a–B. Each item requires the student to use the E–6B computer to solve a navigation problem, such as finding wind direction, wind force, track, heading, and ground speed. The student selects the answer to each item from among five alternatives.

In setting up mislead alternatives, attempts were made to rule out, as far as possible, the chance selection of the correct answer. Several of the four misleads were given values near the correct value in a number of the items. The problem then arose of the size of errors that were due to variation in the accuracy of the E–6B computers. Guidance on this problem was found in an unpublished study conducted in January 1945 by the Navigator Project. In order to determine the magnitude of computer differences, the Navigator Project used eight each of four common types of computer to solve a number of typical navigation problems. The distribution of the answers about the arithmetic mean of all solutions indicated that computer variability was negligible. Only 7 percent of the computers had errors greater than 1 knot for the computation of true air speed from indicated air speed, pressure altitude, and temperature. Drift determination, using a 40-knot wind 45° from the true heading, produced an error of ½° or greater in only 3 percent of the cases. Ground speed in the same problem was in error by 1 knot or more for only 4 percent of the computers.

Misleads for the E–6B computer problems also incorporated common errors in procedure. For example, when problems involve high velocity winds, it is necessary to enter the wind velocity on the computer at a fraction of its value. At the same time, the true air speed must be entered at the same fraction of its value and the resulting ground speed must be multiplied by the reciprocal of the fraction. In such problems, the values that would be obtained if a student overlooked one or more of the enumerated steps were given as misleads.

Sample: E–6B computer problems.

74. Given:

Wind	080°/70 K
True heading	075°
True air speed	250 K.

Find: Ground speed:

74–A 148 K.
74–B 180 K.
74–C 188 K.
74–D 205 K.
74–E 208 K.

Part D consists of 20 airplot problems with a time limit of 30 minutes. The problems require the use of the E–6B computer, Weems plotter, and dividers in determining winds by the airplot method or carrying out dead-reckoning procedures. The plotting is done on the unused portions of the mercator map used for parts A and B. The student is required specifically to determine air position, wind velocity and direction, and estimated time of arrival at a proposed point in the flight. Several groups of interrelated items are included. The mislead alternatives for the five choice items were obtained from the computations of the panel of experts and from a consideration of common errors in procedure. Consideration was given to the difference in latitude and longitude scale, and to the greater inaccuarcy of short compared to long wind legs.

Sample: Airplot.

Questions 88 and 89 are parts of the same problem.

88–89 Given:

Point of departure	53°32′ N., 05°38′ E.
Time of departure	0829.
True heading	060°.
True air speed	192 K.
Time of turn and PPI fix	0843.
Coordinates of PPI fix	54°04′ N., 06°42′ E.
New true heading	210°.
New true air speed	198 K.

88. Find: Wind.

88–A 160°/50 K.
88–B 168°/12 K.
88–C 168°/43 K.
88–D 177°/12 K.
88–E 177°/43 K.

89. Find: Best estimate of ground position at 0856. (Use the wind in question 88 that is closest to the one you actually computed.)

89–A 53°31′ N., 06°09′ E.
89–B 53°36′ N., 06°06′ E.
89–C 53°39′ N., 06°00′ E.
89–D 53° 40′ N., 06°10′ E.
89–E 53°41′ N., 06°10′ E.

Final Test I is applicable to curricula built around either the AN/APS-15, AN/APS-15A, or AN/APQ-13 sets since it is concerned primarily with elementary navigational techniques. Two forms of Final Test I were prepared, P1a-A and P1a-B. P1a-B differed from the earlier form primarily in that the simulated mission was expanded from 25 to 45 items without an increase in testing time. The process by which this was accomplished as mentioned earlier, consisted of splitting items with two-variable answers into two items, each having one-variable answers. Also complex problems were broken down into intermediate steps and an answer was required for each step. The items in P1a-A were analyzed and items were deleted on the basis of lack of relation with total test score or because of extremely high or low difficulty level. The items in the E-6B and airplot sections were re-arranged and the time limit for the E-6B section was shortened from 15 to 12 minutes to reduce the number of students completing the section. The latter change was necessary to increase the discriminatory power of this test at the upper levels of proficiency.

The means and standard deviations for the part and total scores of both forms of Final Test I are presented in table 5.8.[13]

TABLE 5.8.—*Means and standard deviations of distributions of part and total raw scores from two forms of final test I* [1]

Part	Form	Number items	Time	N	Mean	SD
			Minutes			
A	P1a-A	25	70	236	15.96	4.06
	P1a-B	45	70	685	29.96	7.52
B	P1a-A	20	30	236	9.41	3.29
	P1a-B	20	30	685	10.44	3.72
C	P1a-A	20	15	236	13.09	3.67
	P1a-B	20	12	685	13.61	3.77
D	P1a-A	20	30	236	8.14	2.41
	P1a-B	20	30	685	9.39	3.12
Total	P1a-A	85	145	236	46.72	10.25
	P1a-B	105	142	685	63.93	12.52

[1] Data for Form P1a-A are based upon Langley Field classes 44-12 through 44-16 and for Form P1a-B, Langley Field classes 45-8 through 45-24.

In table 5.9 are presented the means and standard deviations for the part and total scores of form P1a-B for three schools.[14] It will be noticed that most of the critical ratios of the differences between the means are significant at the 1 percent level. The differences are explainable in terms of population differences and the varying emphasis upon navigation procedures. Students from both the highest and low-

[13] All statistical work was conducted under the supervision of S/Sgt. Bernard C. Sullivan. He was assisted by Cpl. Robert H. Koch, Sgt. Samuel D. Morford, and Cpl. Sofer. For Final Test II and subsequent tests, Cpl. Koch discontinued work with statistical analysis and contributed directly to test construction with item development.

[14] Critical ratios were computed by Cpl. Sofer, and checked by Cpl. Wilbert H. Schwotzer.

TABLE 5.9.—*Comparison of schools in terms of means and standard deviations of distributions of part and total raw scores of Final Test I, P1a–B*[1]

Part	School	N	Mean	SD	Critical ratio of mean differences		
					1	2	3
A	1	685	29.96	7.52		14.21	4.22
	2	496	20.92	7.60	14.21		12.80
	3	625	26.73	7.18	4.22	12.80	
Total		1,806	26.61	7.53			
B	1	685	10.44	3.72		9.63	6.65
	2	496	8.38	3.58	9.63		3.27
	3	625	9.09	3.64	6.65	3.27	
Total		1,806	9.45	3.66			
C	1	685	13.63	3.77		16.54	1.73
	2	496	10.02	3.62	16.54		15.40
	3	625	13.27	3.38	1.73	15.40	
Total		1,806	12.60	3.60			
D	1	685	9.39	3.12		18.49	4.64
	2	496	6.32	2.60	18.49		14.02
	3	625	8.62	2.88	4.64	14.02	
Total		1,806	8.37	2.90			
Total	1	685	63.93	12.52		21.06	8.80
	2	496	45.56	13.26	21.06		15.26
	3	625	57.67	13.14	8.80	15.26	
Total		1,806	57.19	12.94			

[1] Data from school 1, Langley Field, are from classes 45–8 through 45–24. Data from school 2, Boca Raton are from classes B–6 through B–425. Data from school 3, Victorville, are from classes 45–[illegible] through 45–24

est scoring schools, No. 1 and No. 2 respectively, were predominantly bombardiers. This tends to emphasize the effect created by curricular differences.

No reliability statistics are available for this test since all parts were speeded and no alternate form was available to use in determining a test-retest reliability.

Difficulty level and internal consistency statistics for form P1a–B from administrations to Langley Field classes 45–8 through 45–30 are given in table 5.10. It will be noted that the airplot section, part D, is the most difficult, followed by the ground position section and the simulated mission, with the E–6B computer section being the easiest.

TABLE 5.10.—*Median difficulty level for upper and lower scoring groups and median phis for the parts of Final Test I, P1a–B*[1]

Part	N	Median difficulty upper 50 percent[2]	Median difficulty lower 50 percent[2]	Median phi[3]
A	866	88	71	0.22
B	866	78	56	.25
C	866	91	71	.26
D	866	77	51	.22

[1] Based on Langley Field classes 45–8 through 45–30.
[2] Based upon the number attempting the item.
[3] Based upon the upper 50 percent versus the lower 50 percent.

The difficulty figures are, of course, based upon the number attempting the item. The interrelations of the four parts of Final Test I are presented in chapter 8.

Final Test II.—Final Test II [15] measures knowledge of the air-borne radar set, radar bombing, and radar navigation. The various forms of the test include from 80 to 110 technical information items and are administered in a 1-hour testing period. None of the forms is speeded. Final Test II is the second of two booklets which comprise the final examination given to students upon completing the course.

As was true of Final Test I, the items for Final Test II were obtained primarily from items developed by the Air-crew Evaluation and Research Detachment No. 1 which were, in turn, suggested by material in the National Defense Research Committee Operator Achievement Examination. New items were constructed from course quizzes, lectures, and training manuals.

Most of the forms of Final Test II contained two parts: the first was concerned with knowledge of set operation and particularly the functions of the components of the set used on a radar bombing run; the second covered the theory and procedures of radar bombing and navigation with the emphasis upon bombing. Questions involving the student's knowledge of the appearance of specific targets on the PPI scope were also included in the second section.

The main consideration that guided the construction of misleads for the five alternative items was that of plausibility. Many of these incorrect choices included typical student errors arising from lack of understanding of required techniques. Sample items concerned with navigation and set operation are given below.

Sample item: Navigation technique.

27. Azimuth stabilization is ON. At the IP the pilot has been given a true heading of 170° to make good a true course of 173°. After the turn is made the target appears at 176°. What is the proper correction?

27–A 6° right.
27–B 3° right.
27–C 3° left.
27–D 6° left.
27–E 9° left.

Sample item: Set operation (technical information):

93. The spinner should not be stopped pointing directly toward beacon because

93–A beacon will stop transmitting.
93–B beacon will overload and cut out other aircraft.

---

[15] This test was developed under the supervision of Capt. Van Saun. The chief contributors were: Sgt. Graff, Sgt. Kreidt, and Cpl. Koch.

93-C the radar set will not pick up any signal.
93-D other aircraft will pick up your signal.
93-E all beacon signals will fuse.

Time limits were assigned to the test which allowed most students to complete all items. For the two-part forms, time limit markers were placed through the test to indicate to the student the adequacy of his working rate.

Four two-part forms of Final Test II were constructed, P1b-A and P1b-B, for the AN/APS-15, and P1d-A and P1d-B for the AN/APQ-13. An additional four-part form, V P1d-A for the AN/APQ-13, was prepared only for the radar schools of the Western Flying Training Command. The revisions were made necessary by fluctuations in the content of the curricula which varied between schools and from time to time at the same school. For example, at one school lectures on off-set bombing were eliminated from the curriculum only to be reestablished later. V P1d-A was made necessary by the western schools' emphasis upon target timing, wind determination, off-set bombing, and procedure turns.

The means and standard deviations of the part and total scores of forms P1b-A and P1b-B for classes at Langley Field are given in table 5.11. These classes consisted of both bombardiers and navigators with bombardiers becoming predominant in classes 45-10 to 45-35.

TABLE 5.11.—*Means and standard deviations of distributions of part and total raw scores for two forms of Final Test II* [1]

Part	Form	Number items	Time	N	Mean	SD
			Minutes			
A	P1b-A	50	27	236	31.17	7.01
	P1b-B	60	35	686	38.91	6.74
B	P1b-A	50	28	236	34.88	6.54
	P1b-B	40	20	680	32.77	6.20
Total	P1b-A	100	55	236	66.02	10.93
	P1b-B	100	55	321	71.66	9.79

[1] Data for form P1b-A are from Langley Field classes 44-12 through 44-16. Data for part scores of form P1b-B are from Langley Field classes 45-8 through 45-24. Data for the total score of form P1b-B are from Langley Field classes 45-8 through 45-12 and 45-22 through 45-24.

In table 5.12 are given the results of reliability studies of two forms of Final Test II. Form P1b-A appears to have a reliability of about 0.80 while P1b-B, given to later classes, seems to have a reliability of approximately 0.70. One possible explanation for the apparent lower reliability of the second form is that the students were believed to be more homogeneous in ability in later classes. As will be noted from the table, the standard deviations for the second form are low as compared to those for the first.

TABLE 5.12.—*Odd-even and Kuder-Richardson reliability coefficients for two forms of Final Test II*

Form	Part	N	Mean-odd	Mean-even	SD-odd	SD-even	r'_{II}	r_{II}[1]
P1b-A[2]	A	236	14.53	16.62	4.16	3.30	0.73	0.84
	B	236	16.40	18.53	3.43	2.76	.69	.75
P1b-B[3]	A	190	19.85	20.68	3.07	2.88	.43	.60
	B	189	13.93	15.00	2.76	2.62	.52	.69
Total		189	33.90	35.53	4.85	4.20	.58	.73

Form	Part	N	Total mean	Total SD	r_{II}[4]
P1b-A[2]	A	236	31.2	7.0	0.78
	B	236	34.9	5.5	.67
Total		236	66.0	10.9	.82

[1] Odd-even coefficient corrected by the Spearman-Brown formula.
[2] Data for form P1b-A are from Langley Field classes 44-12 through 44-16.
[3] Data for form P1b-B are from Langley Field classes 45-10 through 45-33.
[4] Kuder-Richardson reliability coefficient.

Item difficulty and internal consistency data, computed for Victorville classes 45–12 through 45–23, are presented in table 5.13. Although this test is more difficult than Final Test I, it is still relatively easy, having an average item difficulty level of approximately 70 percent passing.

TABLE 5.13.—*Median difficulty level for upper and lower scoring groups and median phis for parts of Final Test II, P1d–B*

Part	N	Median difficulty upper 50 percent[1]	Median difficulty lower 50 percent[1]	Median phi[2]
A	600	79	60	0.12
B	600	73	66	.10

[1] Based upon the number attempting the item.
[2] Based upon the upper 50 percent versus the lower 50 percent.

Final test for AN/APQ–7.—The final test for AN/APQ–7 consists of technical information and problem-solving items measuring knowledge of the operation and functions of the AN/APQ–7 set and its use in radar bombing and navigation. The test consists of 120 items and requires 97 minutes to administer.

The AN/APQ–7 is an air borne radar set that is especially valuable in bombing. Students in the AN/APQ–7 course had already graduated from the 10-week AN/APQ–13 course and were given 4 weeks of additional training. Only one school in the radar training program taught the use of this equipment. It was necessary to develop a special final examination because the operating procedures and functions of the AN/APQ–7 were different from those of any other set for which the Radar Project had constructed a test.

The item material in the final test for AN/APQ–7 was compiled entirely by the Radar Project. Test construction teams participated in sample training missions, attended classes, and examined lectures and training outlines. As items were constructed, they were discussed with instructors in the AN/APQ-7 course.

The final test for AN/APQ-7 [16] consists of five sections. The first three are made up of problem-solving navigation and bombing items, while the final two measure verbal knowledge of equipment and procedures. Section A is speeded and includes 20 problem-solving items given with a time limit of 30 minutes. The items require the student to determine wind force and direction, true headings for bombing runs, range-wind factors and ground speeds to set into the AN/APQ-7 computer for the bombing run, and drift angles to set into the bombsight. Auxiliary equipment required to solve these problems includes a Mercator projection, an E-6B computer, an N-1 ground speed computer, a Weems plotter, and dividers. The technique used for form P1a-B of Final Test I of splitting wind force and wind velocity into separate items was employed to increase the number of items without increasing the necessary testing time. The primary consideration in constructing the mislead alternatives for the five-choice items was to penalize students for inaccuracy while allowing for reasonable deviations. Several sample problems are shown below.

Sample: 1-4. Given:

Point of departure	42°19′ N., 135°20′ E.
Time of departure	0928
True heading	135°
True air speed	210 m. p. h.
Time of GPI fix	0946
Target of GPI fix	41°00′ N., 136°10′ E.
Bearing range of GPI fix	13° R/30 S. M.
Initial point	40°40′ N., 136°20′ E.
Target	40°20′ N., 135°50′ E.

1. Find: Wind direction.
   - 1-A 050°.
   - 1-B 060°.
   - 1-C 075°.
   - 1-D 235°.
   - 1-E 245°.

2. Find: Wind force.
   - 2-A 37 m. p. h.
   - 2-B 43 m. p. h.
   - 2-C 49 m. p. h.
   - 2-D 55 m. p. h.
   - 2-E 69 m. p. h.

---

[16] Developed by Sgt. Gage.

3. Find: True heading from initial point to target.

3–A 036°.
3–B 043°.
3–C 050°.
3–D 227°.
3–E 237°.

4. Find: Ground speed to be set into computer for bombing run.

4–A 157 m. p. h.
4–B 162 m. p. h.
4–C 278 m. p. h.
4–D 290 m. p. h.
4–E 301 m. p. h.

Section B is a speeded section which includes eight items with a time limit of 15 minutes. These items measure proficiency in computing procedure turns and require the use of an E–6B computer and three tables besides the navigation data given in the item. The items are multiple choice with no regularity of tolerances between choices. A sample item is given below.

Sample:

Wind (statute miles)	*True heading to target*	*True heading to I. P.*	*True air speed (m. p. h.)*	*Absolute altitude (feet)*	*Slant range from IP at which procedure turn should be started (statute miles)*				
					A	B	C	D	E
23. 190°/40	240°	180°	180	10,000	1.4	2.1	2.3	2.9	3.3

Section C is a speeded section of 24 items with a time limit of seven minutes. The items require the student to solve drift problems on the radar bombing run, making corrections for drift and determining amount and direction of drift on a given heading. No equipment is required and answers are multiple-choice. A sample item of this section is given below.

Sample:

	Distance of first bearing (statute miles)	*Distance of second bearing (statute miles)*	*First bearing*	*Second bearing*
31–33	30	25	4° L.	2° L.

Answer:

	A	B	C	D	E
31. Direction of correction	L	R	------	------	-------
32. Amount of correction	2°	4°	6°	8°	12°
33. Precorrection drift	14° L.	10° L.	8° L.	8° R.	10° R.

Section D, not speeded, consists of 33 technical information items and has a time limit of 20 minutes. The items measure verbal knowledge of equipment function and operation, especially emergency calibration procedures and location of burnt-out fuses. In constructing misleads, the primary concern was for plausibility.

Sample: 58. Scan zero start and sector sweep should be adjusted so that—

58-A the sweep is of uniform intensity from 0 miles to 30 miles.
58-B the sweep is of uniform intensity from 30° left or right through 0°.
58-C there is a prominent bright line down the center of the scope.
58-D there is a definite dark line down the center of the scope.
58-E there is a thin bright wedge down the center of the scope.

Section E, not speeded, has 33 items and a time limit of 25 minutes. The items measure knowledge of radar bombing theory and procedures. Both problem-solving and technical information items were used, as illustrated by the two sample items below.

Sample: 89. On a synchronous bombing run using the Eagle-Norden Sighting Angle Computer, the bomb is dropped by the—

89-A Eagle-Norden Sighting Angle Computer.
89-B operator's indicator computer.
89-C radar observer's toggle switch.
89-D bombardier's toggle switch.
89-E bombsight.

109. If it takes 20 seconds for a bomb to reach the ground from an airplane traveling at a ground speed of 180 m.p.h., what is the whole range?

109-A 4,020 feet.
109-B 4,180 feet.
109-C 5,090 feet.
109-D 5,280 feet.
109-E 6,080 feet.

Because the equipment and associated training course was developed late in the radar program, only one form of this test, W Plg-A, was prepared. This form was installed just before the end of the war, so very few statistical data were accumulated for it.

Navigation Proficiency Test

The Navigation Proficiency Test [17] measures knowledge of nonradar navigation information and techniques that are considered to be essential for potential radar observers. The test does not include items pertinent to radar operation or radar navigation, but is merely a general navigation proficiency measure.

The Navigation Proficiency Test was first used in the 16-week radar observer curriculum, being given as a final examination at the close of the 4-week navigation phase. When this phase was eliminated

[17] This test was developed by Capt. Ike H. Harrison. He was assisted by Sgt. Kriedt.

and the curriculum was reduced to 10 weeks, the test was occasionally given to bombardiers entering radar observer training to indicate any specific weaknesses in navigation. Finally, it was incorporated into the radar observer selection battery to eliminate bombardiers with insufficient navigation proficiency, as discussed in chapter 10.

Much of the item material for the Navigation Proficiency Test was obtained from tests developed by the Navigator Project. This material was supplemented by navigation items from the Aircrew Evaluation and Research Detachment No. 1 examinations. The informal quizzes used in the 4-week navigation phase of the 16-week curriculum and various navigation manuals provided further suggestions for items.

The two forms of this test, P5–A and P5–B, contain 148 and 135 multiple-choice items, respectively, and require 135 and 125 minutes to administer. The test is divided into seven parts, including three sections which involve primarily problem-solving items and four which consist mainly of technical knowledge items. The seven parts as they appear in form P5–B are described in the following paragraphs.

Part I is a speeded section including 40 E–6B computer problems with a time limit of 40 minutes. The items include such problems as determining true air speed, ground speed, wind force and direction, track, and time to destination.

Part II is a speeded section measuring skill in using the air-plot method to determine winds. It contains 20 items with a time limit of 35 minutes. The items require determination of either air position, ground position, or the wind force and direction. Part of the air-plot items are solved on Mercator charts included in the test booklets, most of the information is given for the solution and no plotting is required. The rest of the air-plot items require plotting on a separate Mercator sheet and are similar to the air-plot items used in Final Test I.

Part III, not speeded, consists of 25 technical information items with a time limit of 12 minutes. The items measure knowledge of maps and associated terminology. Part IV, also not speeded, contains 10 items with a 10-minute time limit and evaluates understanding of wind and its effect on drift. The items require interpretation of information included in vector diagrams.

Part V, not speeded, contains seven problems measuring knowledge of compass and drift-meter calibration and alignment and has a time limit of 8 minutes. Part VI, not speeded, has 22 items, a time limit of 10 minutes, and measures knowledge of navigation instruments such as the compass, altimeter, drift-meter, and air speed indicator. Part VII contains 11 items to be answered in 10 minutes. The items are concerned with technical knowledge and interpretation of the flux-gate compass, which is the important unit of the azimuth stabilization circuit of the air-borne radar set.

The revision of form P5-A of the Navigation Proficiency Test which resulted in form P5-B was made necessary when it was decided to give the test to bombardiers for the purposes of diagnosis of navigation weaknesses and selection for radar observer training. The revision consisted of eliminating from the original form all items which contained information not taught in the navigation phases of bombardier training.

The means and standard deviations of selected part and total scores, based upon administrations to entering bombardiers, bombardiers who had completed the 4-week navigation phase of radar observer training, and previous navigation instructors are presented in table 5.14. As would be expected, the entering bombardiers had the lowest

TABLE 5.14.—*Means and standard deviations of distributions of raw scores from Navigation Proficiency Test, P5 form B, for differently trained groups* [1]

Group	N	Part I		Part II		Parts III-VII		Total	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	28	31.35	4.38	13.92	1.33	49.04	7.77	94.23	11.51
2	90	25.61	6.21	11.46	4.73	50.27	12.24	86.59	19.37
3	32	32.22	4.77	14.72	2.73	60.25	5.22	107.19	9.69

[1] Group 1 consists of bombardiers completing 4 weeks of navigation training, Langley Field class 45-4R. Group 2 consists of bombardiers entering radar observer training, Langley Field, 25 February 1945. Group 3 consists of previous navigation instructors entering radar observer training, Langley Field, 25 February 1945.

scores in all sections of the test. The scores of the bombardiers who had completed the 4 weeks of navigation training indicate that this training was adequate to bring their E-6B performance almost to the level of well-trained navigators. The critical ratio of the difference between the means of the 4-week-trained bombardiers and the navigation instructors for part I, the E-6B section, is 0.72. The training did not bring the bombardiers up to the navigation instructors in air-plot proficiency, the difference between their means for part II having a significant critical ratio of 4.57. Also, the training did not appear to improve the bombardier's performance on parts III to VII, the technical information sections. The difference between the means of the entering bombardiers and 4-week-trained bombardiers for the total of these sections is insignificant, having a critical ratio of 0.62.

SUMMARY

In the development of a standardized technique for determining relative proficiency of students in radar training, a battery of 11 measures was prepared. Five of these measures were printed proficiency tests and six were performance checks. In this chapter only the printed tests have been discussed.

Printed tests were developed for the four instructional areas of radar observer training: Set operation, radar navigation, radar bombing, and radar intelligence. Two types of items characteristic of printed tests were used: Verbal questions measuring technical information, and problem-solving items duplicating parts of a student's actual job as a radar observer. Problem-solving items were constructed as isolated problems and as series of interdependent problems, the latter use simulating the sequential activity required in actual radar operation. It was realized that many of the complex operations required of the radar observer could not be effectively reduced to printed question form but required performance checks for their evaluation. In addition to handling the equipment and solving problems sequentially, operators must interpret moving patterns on the PPI scope and must recognize and adjust for peculiar symptoms caused by set malfunctions. It is impossible to represent moving patterns and malfunctional symptoms on a printed page.

An outline of the usual procedure followed in constructing tests and making revisions was presented. Material for items was obtained from existing informal classroom quizzes, tests constructed by the National Defense Research Committee and the Air-Crew Evaluation and Research Detachment No. 1, course outlines, lecture notes, and training manuals. Items were reviewed by expert radar observers and instructors in pertinent courses before being incorporated into a test. Problem-solving items were solved by panels of experts who determined correct answers and desirable alternative choices. Initial time limits were set on the basis of the time required by experts to answer all items and were revised later on the basis of raw score distributions from administrations to students. Problem-solving sections were speeded for the purpose of discriminating between students on the basis of speed as well as accuracy. Technical information sections were given adequate time limits to allow every student to answer all problems. All tests were scored in terms of the number of correct responses. Revisions were made on the basis of statistical analysis, reliability studies, and examination of raw-score distributions, and on the basis of systematically gathered criticisms and changes in the curricula.

The principal difficulties in applying proficiency tests arose from the rapid changes in air-borne radar equipment which resulted in curricular revisions and changing emphases on various procedures. The radar intelligence proficiency tests were particularly affected by the changing subject matter so that the attempt to apply standardized measures to this area of instruction was finally abandoned.

Since it was not feasible to describe all forms of the test constructed, discussion was limited to typical tests.

For each type of test a statement of the general information measured was first presented. This was followed by locating the place in the course of study where the test is administered. A developmental history of the measure was presented and reference was made to antecedent tests in the same general field of study. Following this information was a detailed description of a typical form of the test: Number of items, time limits, and specific information measured by the items. Items are classified either as problem-solving or technical information. The description of the type of response required was presented with the method of ascertaining tolerances. A sample item was presented with each of these item discussions to illustrate the principles involved. For reference, all forms of each test were presented with their respective code numbers. With this list was given the reasons for the various revisions.

Wherever possible, statistical data were presented for the major forms of each test. These data included means, standard deviations, reliability coefficients, and difficulty level and internal consistency statistics.

Several recommendations may be made which stem from the experience of the Radar Project in constructing standardized proficiency tests. First, the nature of equipment and technique advances in airborne radar requires that a systemized test revision program be adopted. Already new and improved equipment has made obsolete a large proportion of existing measures. For example, the original estimate of necessary tests for existing equipment differences was 16. Independent operating procedures and curriculum revision mushroomed test construction to 29 forms. Coordinated changes in procedure and curriculum might have reduced this number.

Second, the success of a proficiency program depends heavily upon standardized test administration. It is recommended that examiner boards be trained and established wherever such a program is initiated. The board will reduce to a small group the individuals who must be convinced of the value of standardization.

When using groups of interdependent problem-solving items, care must be taken to guard against unreliability. The interdependence of items may in effect shorten the length of a test by automatically penalizing students for a whole series when they miss the first item. The organizational and sequential aspect can be retained to some extent by periodically breaking the sequence at the end of a typical series of items and having the students start another group.

Experience with wind problems and problems answered in terms of latitude and longitude suggests that velocity and force or latitude and longitude be presented as separate items. Another method of lengthening a test without appreciably increasing time limits consists of making separate items out of intermediate steps in the solution of a complex problem. Wherever possible, maps, diagrams, pictures, or any other device that adds reality to a problem should be used.

A clear understanding of the relative efficiency of printed tests and performance checks should be kept in mind when selecting the testing device for a given area of study. It is recommended that correct answers for problem-solving items be determined by a panel of authorities rather than depending upon average results of trial administrations. In determining differences between alternative choices of problem-solving items, it is recommended that rule-of-thumb differences be avoided and that the proper deviations be determined empirically.

The time required for instructing students in any single topic should not be considered as the index of the importance of that topic. Relative importance of subject matter should be determined by training authorities.

Experience with problem-solving items suggests that speed as well as accuracy be measured. For determination of reliability of these speeded forms, alternate forms must be prepared.

CHAPTER SIX

Standardized Performance Checks[1]

INTRODUCTION

The necessity of supplementing printed proficiency tests with measures of actual performance with radar equipment soon became clear. In determining the student's course grade, radar school authorities weighted heavily student performance on ground trainers and aerial missions. If course grades were to be made more reliable, standardized measures of performance had to be provided.

Considerable contrast is evident in comparing the classroom situation in which printed tests are given with the actual job conditions of the radar observer. The classroom is usually comfortable, with plenty of space to work, and the student has no responsibility other than attempting to do well on the test. The working situation in the air is vastly different. The student is usually hampered by cumbersome clothing; he must wear an oxygen mask; his working space is cramped and inadequate; and he is subject to all the usual distractions and anxieties of flying. Moreover, he has the responsibility of directing the aircraft and cooperating with the other crew members.

Performance in the air and on the ground trainer differs further from a written test situation in that it is continuous and paced. Regardless of what the student does, the aircraft keeps moving, requiring new computations and procedures. He cannot, as he may on a printed test, answer the easy problems first and delay the solutions of difficult problems. Since the various aspects of radar performance in the air are interdependent, failure to perform adequately on a difficult aspect will likely detract from performance on other aspects which may be easier.

Standardized performance checks were indicated for three aspects of radar observer training: Performance in the air, at the supersonic trainer, and at the bench set trainer. It was decided in consultation with training authorities that aerial performance would be checked at two levels. The first aerial check would be an intermediate check to be administered on a mission approximately half way through the

[1] Written by Cpl. Douglas W. Bray.

student's aerial training. The second would be a final check given just prior to completion of the course. Supersonic trainer performance was to be evaluated at three points. An intermediate navigation check was to be given after instruction in basic navigational procedure; an intermediate bombing check after bombing techniques were taught; and a final check near the completion of supersonic training. Performance on the bench set trainer was to be covered by one final check to be given relatively early in the course after the elements of set operation had been taught.

In this chapter an illustrative form of each of these three types of checks will be described and available statistical data will be presented. In addition, a brief discussion will present some problems encountered in constructing performance checks and in training examiners to administer them. Consideration of systematic aspects of the measurement of performance will be deferred until chapter 7. The present chapter is intended only as a history of the development of the performance checks in a specific training situation.

GENERAL PROCEDURE USED IN DEVELOPING PERFORMANCE CHECKS

Before outlining the procedure followed in constructing the specific performance measures developed by Psychological Research Project (Radar), a general characterization of "performance checks" is in order. A performance check is usually administered to one individual at a time by a trained examiner. The student is required to perform a standardized series of tasks which is the same or equivalent for all students. The conditions under which each task is to be accomplished are also standardized. For each task, standards of success and other instructions for administration are prescribed.

The first step in constructing each check for the radar observer program was the selection of the critical aspects of performance. The personnel charged with the development of a particular check began by systematically observing the behavior to be measured. For example, before starting work on the aerial performance checks, project members accompanied student and graduate radar observers on numerous flights. In addition to observation, the task was also studied by participation; those responsible for check construction learned and took part in the performance themselves. To this end, many members of the project took the entire radar training course, while others concentrated on only those phases of the course with which they were mainly concerned. Frequent conferences were held with instructor consultants and department heads. Their personal experience as radar observers as well as their familiarity with curriculum and instruction were utilized to great advantage.

Once the items of behavior to be checked were selected, it was necessary to decide on the method of measurement for each item. However, such decisions were so dependent on the task to be checked that their discussion will be postponed until the specific checks are considered. After the methods of measurement were selected, the check items and standards of success were stated in preliminary form and trial administrations were begun. On the basis of these trial administrations, revisions were made in both the choice of items and methods of measurement.

ARRANGEMENT AND FORMAT

An effort was made to standardize the organization and appearance of all performance checks developed by the radar project. The items of performance to be checked were preceded by a uniform introductory section. This section contained notes on standardized performance checking, a statement of required conditions for the check, and directions for administration and for student briefing. The notes on standardized performance checking stated briefly certain principles which were expanded during examiner training lectures. The section on standard conditions enumerated the training prerequisites which the student must have completed before being given the check; it called attention to the requirement that only trained examiners should administer the check; it listed the equipment needed by both student and examiner; and it described the conditions under which the check was to be given. For each of the aerial and supersonic checks, standard missions were prescribed. For the bench check, the positions were listed in which the controls of the radar set were to be placed before starting the check.

The prescribed procedures for administration included directions to the examiner for using the special scoring sheet, acquainted him with his special duties in administering the particular check, and cautioned him to brief the student carefully. The student briefing section, to be read by the student, defined the task for the student and stated the special requirements to be fulfilled.

The introductory section was followed by the items of performance to be checked. Each was accompanied by standards of success and methods of evaluation. The items appeared on cut-back pages which were narrower than the other pages in the check booklet. A single answer sheet for all the items followed the item pages and was of normal width. The right-hand side of the answer sheet contained spaces in which a check mark (√) or a zero (0) were to be entered, depending upon whether a student did or did not satisfy the standards of success for the particular item. Because of the narrow item pages, both the items and the check spaces were visible at the same time. This made it unnecessary to turn pages to check an item. Such an arrangement was of advantage also in that only the scoring sheet was expendable. Two different arrangements involving cut-back pages were employed. On some performance checks, the item pages decreased in width on succeeding pages so that there was provision for several columns of check spaces on the scoring sheet. On other performance checks, the item pages were of uniform width and required only one column of check spaces. The latter arrangement was more compact and easier to reproduce but was limited to performance checks in which the number of items did not exceed the number which could be checked in one column.

Several performance checks included a further section which contained supplementary instructions to the examiner. The instructions were numbered to correspond with individual check items and elaborated the more general statement of methods of checking. This section saved valuable space on the item pages and avoided encumbering them with material that was useful but not always essential in checking the items.

EXAMINER TRAINING

The effectiveness of a performance check depends heavily upon proper administration. Recognizing this, the project recommended the formation of examiner boards to be responsible for all proficiency measurement. It was urged that competent radar observers, interested in proficiency measurement, be assigned as members of such boards.

The training of examiners was initiated with a lecture covering relative grading, the importance of standardization, the rationale of performance checking, and the distinctions between testing and teaching. Specific checks were discussed in detail. Each item was read, explained and illustrated, and the directions for administration were reviewed. Examiners were warned against common errors in administering each check. This lecture-discussion was followed by trial administrations of each check by pairs of examiners, who alternated as subject and check administrator. Each trial check was carefully supervised by project personnel. For practical reasons, trial administrations of the aerial check were given on the supersonic trainer.

Since it was impossible for project members to supervise all subsequent routine check administrations, certain examiners were given supervisory responsibility. It was their duty to see that conscientious, standardized checking was carried out. In addition, however, project personnel made frequent observations of check administration and reported discrepancies and difficulties to the chief examiner.

CHECK REVISIONS

Revision of the battery of six standardized performance checks was to be expected in the new, expanding radar training program. An evident reason for new forms of the checks was that the types of radar equipment varied from school to school. Langley Field, where the first set of checks was developed, taught the AN/APS-15 and the AN/APS-15A sets. Boca Raton Army Air Field, Victorville Army Air Field, Williams Field, and Yuma Army Air Field all used the AN/APQ-13 radar set. AN/APQ-7, a newer set and less important so far as student flow was concerned, was the subject of a separate course, centered first at Boca Raton and later at Williams Field.

The necessity for check revision was not limited, however, to dealing with equipment differences. Even when two schools used identical radar sets, there was no assurance that the procedures taught for tuning the set or for navigating and bombing with the set would not differ from school to school. For example, although the schools at Boca Raton and Victorville both used the AN/APQ-13 equipment, the method by which the student was taught to compute a wind from his radar data differed markedly between the two schools. To understand this, it should be remembered that radar observer training had developed without any centralized instructional authority. Difficulties in developing standardized performance checks under such circumstances were unavoidable.

Not only did the curriculum vary between schools, but in any one school it was subject to frequent change. This was a consequence of the fact that radar training was new and better ways of doing things were constantly being discovered. Fortunately, such curriculum changes were not usually radical, and revised forms of the checks could be produced by changes in several items. Occasionally, however, a curriculum change of such magnitude occurred that the use of radar project performance checks had to be abandoned temporarily.

Equipment and curriculum changes were of course not the only forces motivating check revisions. Experience with the initial forms of the checks sometimes showed that the selection of critical items could be improved or that the method of measurement for items retained could be revised for more accurate evaluation. Also, item analysis occasionally demonstrated that the tolerances set for precision items was too small or too large. Revisions attempted to incorporate all promising improvements.

STANDARDIZED BENCH SET PERFORMANCE CHECKS

The Bench Set and the Student's Task

The apparatus upon which the standardized bench set check was administered was simply the radar set installed in a classroom, rather than in the airplane.[3] The parts of the equipment with which the check was concerned were the various controls involved in tuning

[3] For a description of the set, see chapter 4.

and calibrating the set, the PPI scope, the "A" scope (except on the AN/APQ-13 set), the computer box, and various voltage controls.

The student's task on the bench set was to make the preoperational check, start and tune the set, calibrate it correctly, and turn it off. The preoperational check included the adjustment of approximately 20 controls. Starting the set included turning on the power, brightening, focusing, centering, and in other ways adjusting the systems which presented the ground returns on the scopes. Tuning consisted in picking up ground returns and adjusting the transmission, receiving, and presentation systems for maximum definition on the scopes. Turning the set off included the adjustment of several controls to avoid damage to the equipment when the power supply was cut.

The objectives of bench set instruction were often a source of disagreement among instructional personnel. Some were convinced that the student should merely learn "procedure" on the bench set: the controls to adjust, the direction and amount of adjustment, and the sequence in which the adjustments should be made in order to produce the desired results on the scope. Others were convinced that the mastery of procedure was relatively unimportant and that the student's proficiency should be evaluated in terms of the end result, i. e., the quality of the scope picture resulting from his precision in tuning.

When the first forms of the bench set performance check were being constructed, it was decided that the emphasis of bench set instruction should be on the learning of a standard operating procedure. The first bench check was therefore built to measure the procedure type of proficiency. For example, the item which refers to centering the trace on the scope was:

> Center *Sweep* horizontally and vertically to reduce size of hole in center of scope.

No attempt was made to evaluate the quality or precision of the student's adjustment. It was considered sufficient that he knew which controls to manipulate to work toward the desired result. A further condition for receiving credit on the above item was that the student made the adjustment in prescribed sequence with other items which preceded and followed it.

After several months, the desire of training authorities to measure precision as well as procedure became strong enough to incorporate into the bench check some items which considered the quality of the student's adjustment. For example, the item on centering became:

> Center *Sweep* horizontally and vertically in exact center of scope. (If sweep is not exactly centered, zero (0) this item.)

The student's knowledge of standard operating procedure still accounted for the majority of items in this check, but, wherever possible, items were stated with accuracy requirements. It is probable that the early preference for procedure items was associated with the use of the older form of the check as a teaching aid. The check, however, apparently improved and standardized instruction to a point where most students could perform the standard operation procedure with almost no error. The check had become "to easy," and the desire to increase its difficulty partly accounted for the later decision to evaluate precision.

The bench set check was administered during the last period allotted to bench set instruction when the student had completed training in the various elements of the bench set task. Students could be reverted to a lower class if deficient in set operation, thus preventing the loss of valuable aerial instruction time.

Bench Check Construction

The selection of items for the bench set check was relatively simple. There was no necessity for sampling behavior; all the steps in the task could be included. Towards the end of the war the construction of the initial bench set check at a school consisted primarily in stating the standard operating procedure in check form. Of course, certain changes in phrasing were made, and standards of success for each item were stated. At the beginning of the training program, however, check construction was handicapped by a lack of uniform instructional practice even within a school. For example, at the school where the first bench set check was constructed, no standard operating procedure existed and the procedure a student learned reflected the practice of the instructor to whom he was assigned. The efforts to establish a standard measure of proficiency speeded the formulation and publication of a standard operating procedure. Since this procedure was to be integrated into an aerial check, standardization between flight line and ground school instruction was also promoted.

The Typical Bench Set Check

The typical bench set performance check[3] contains the conventional introductory material as outlined above. The most important section of the introductory material directs the examiner to set about 30 controls in designated positions. Most of these settings were designed to put the controls out of adjustment so that the student would be required to readjust them. The controls were to be set in the same initial positions for every student.

The number of items in the various forms of the bench set check varied from 79 to 136. As indicated above, they included all items necessary to accomplish the tuning and calibrating tasks. The student earned a check on an item if he performed the task as stated and if he performed it at the proper time. Sequence, therefore, was very important. Most of the steps could be performed correctly only between

[3] See appendix B.

two other specific steps. Small groups of items were sometimes bracketed which indicated that the items within the brackets could be performed in any sequence.

Types of Bench Check Items

As mentioned above, procedure items constituted the bulk of the bench check. An example of a pure procedure item follows:

Receiver Gain: Counter-clockwise.

No attempt was made to give more credit for important procedure items than for minor ones. In later forms of the check, as already noted, examiners were required to judge the precision of adjustments. For example:

Turn up *PPI Brilliance* until trace is barely visible. (If trace is invisible or too bright, zero (0) this item.)

In addition to items of actual performance, most of the bench checks contained a few items that were administered as oral questions during the course of the check. For example:

Instructor asks: "In the air what must you do immediately after turning the power on?"

Answer: "You must select inverter No. 1 or No. 2 by means of the inverter selector switch."

One reason for such questions was that certain essential steps in aerial procedure could not be checked on the bench set because the appropriate equipment was not at hand. For instance, in reference to the quoted item, there were no inverters on the bench set. Other questions were directed to asking the student what he would do if certain expected results did not occur. The student was not required to answer in the exact words given for the item, but he had to include all essential elements of the answer in his reply. Since the early forms of the bench set check were used for instruction, the questions and answers were just another item of procedure which the student could memorize.

Bench Check Administration

In administering the check, the examiner was instructed to say nothing to the student except that which appeared in quotation marks in the check. Except for the items in question form, the examiner gave only general infrequent directions such as "Go through the preoperational check." The student was briefed to perform all the steps called for by the examiner's directions and, as an aid to checking, to verbalize all steps as he did them. When the student performed an item correctly and in the correct sequence, the examiner entered a check mark in the appropriate space on the scoring sheet. If the item was performed incorrectly, omitted, or not performed in the correct sequence, a zero was entered. When the student performed the item incorrectly or omitted it, the examiner corrected him at once and told him the adjustment to make. The administration of the check required about one-half hour.

In spite of the apparent simplicity of the checking procedure, there was some difficulty in securing standard administration. Differences in the functioning of equipment were common. Some sets on which checks had to be given would not pick up targets. This reduced certain steps in the tuning to artificial procedure instead of actual work with a radar return. Meter readings on the bench set were frequently outside the acceptable range, and the student could not adjust them because of the nature of the power supply on the ground.

Examiner differences were also encountered. Halo effect, accentuated by some students' apparent confidence and ability, led examiners to give credit to supposedly good students whose errors were explained away as only oversights. Sometimes less conscientious examiners who disagreed with the standard operating procedure would give the student credit for following a procedure which suited the examiner rather than the check. Some examiners were noticed giving unintentional hints or leading questions in a well-meaning effort to "get the best out of the student." Operating in the other direction, a common mistake was failure to stop the student at once for an error or omission and thus permit his errors to multiply.

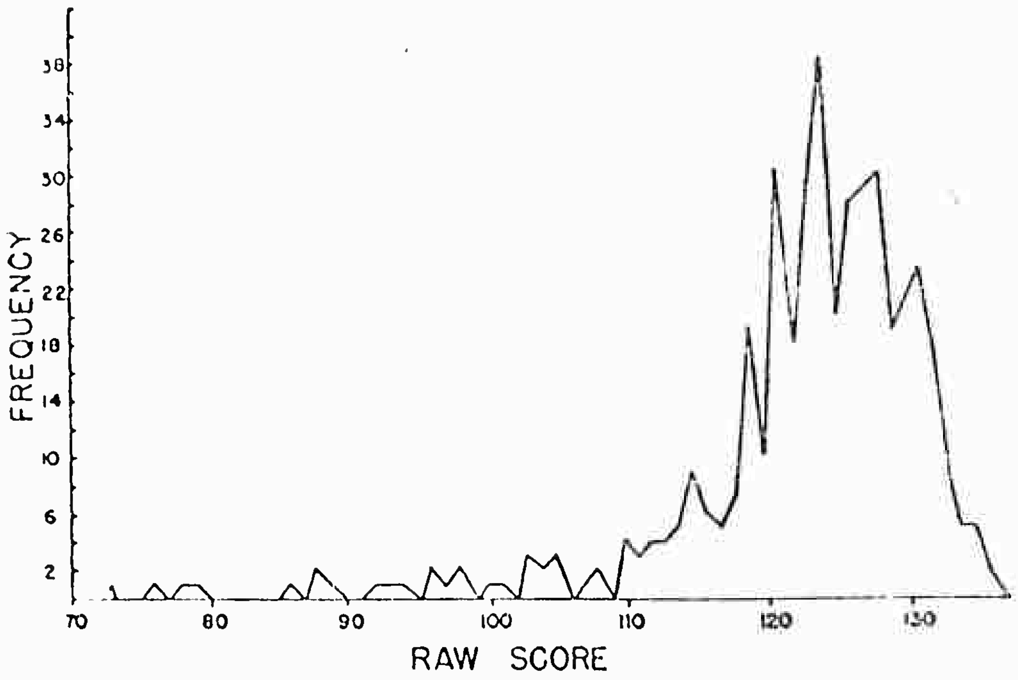
Bench Check Revisions

It has been noted that new forms of the bench check were needed because of equipment differences, changes in curriculum and operating procedure, and the increased emphasis on precision. In all, 11 forms of the Final Bench Set Check were published. Three of these (P6a-A, P6a-B, and P6a-C) applied to the AN/APS-15 equipment at Langley Field, two (P6b-A and P6b-B) applied to the AN/APS-15A at Langley Field, two (P6c-A and P6c-B) applied to the AN/APQ-13 as taught at Boca Raton, two (P6c-C and V P6c-B) applied to the AN/APQ-13 at Victorville, one (W P6c-A) applied to the AN/APQ-13 at Williams Field, and one (W P6d-B) applied to the AN/APQ-7 at Williams Field.[4]

Bench Check Statistical Findings

Figure 6.1 presents a typical distribution of scores on the bench check. The total possible score on this form was 136. The mean of

[4] Cpl. Douglas W. Bray, Sgt. Gerald S. Blum, S/Sgt. Richard T. Mitchell, Sgt. Albert H. Hastorf, and Sgt. Hyman Heller collaborated in developing the first Bench Set Check (P6a-A). Eight of the 11 checks were constructed by Cpl. Bray and Sgt. Blum. The remaining two were developed by Sgt. Stanley [illegible] and Sgt. Nathaniel L. Gage.



N = 428

MEAN = 122.74

SD = 9.32

SKEW = − 2.00

SD_{SK} = .45

$C.R_{SK}$ = 4.44

FIGURE 6.1.—Frequency distribution of raw scores for a standardized final bench check (P6a-B)—administered at Langley Field.

122.7 shows that the check was easy for most students. Note the restricted range and the significant skewness of the distribution.[5]

Since the reliability of bench check scores depended in large measure on standardized administration, there was a desire to examine check data statistically in an attempt to determine whether such standardization had been achieved. It was anticipated that the technique of analysis of variance would be employed to test the null hypothesis that there was no significant difference between mean check scores given by different examiners. For this purpose, data on accredited examiners who had administered six or more bench checks were accumulated for both Langley Field and Boca Raton. These data are presented in tables 6.1 and 6.2.

A chi-square test for homogeneity of variance revealed that the analysis of variance technique could not be used on either set of data.

---

[5] For this and subsequent distributions skewness was computed from the following formula:

$\frac{P_{90}+P_{10}}{2}-P_{50}$, where P_{90} is the 90th centile, P_{10} is the 10th centile and P_{50} is the 50th centile. Standard deviation of the measure of skewness was computed as follows:

$$S.D._{sk}=\frac{0.5185}{\sqrt{N}}, \text{ where } D=P_{90}-P_{10}.$$

Kelley, T. L. *Statistical method*, 1923, p. 77.

TABLE 6.1.—*Examiner means and standard deviations for a standardized bench set trainer performance check (P6a–B) administered at Langley Field*

Examiner	Number of checks administered	Mean raw score [1]	SD	Examiner	Number of checks administered	Mean raw score [1]	SD
A	6	128.17	2.97	I	13	123.00	5.28
B	22	121.86	6.55	J	7	122.29	9.24
C	43	124.30	4.21	K	53	121.26	10.41
D	45	123.71	7.23	L	42	120.69	11.55
E	13	123.69	22.11	M	22	118.77	7.18
F	43	123.26	10.58	N	10	114.10	13.04
G	13	123.15	9.74				
H	38	123.14	10.17	Total	368	122.49	9.80

[1] The highest possible raw score on this check was 136.

TABLE 6.2.—*Examiner means and standard deviations for a standardized bench set trainer performance check (P6c–B) administered at Boca Raton*

Examiner	Number of checks administered	Mean raw score [1]	SD	Examiner	Number of checks administered	Mean raw score [1]	SD
A	7	78.43	0.49	K	8	73.25	1.71
B	11	78.00	.74	L	7	72.43	5.12
C	17	77.00	2.85	M	13	72.08	4.63
D	13	75.54	2.44	N	20	71.85	6.28
E	12	74.92	2.78	O	7	71.14	4.16
F	9	74.33	2.45	P	14	70.29	6.25
G	15	74.27	3.70	Q	7	70.29	4.62
H	15	73.80	3.23	R	8	70.25	3.63
I	10	73.40	5.41				
J	11	73.27	2.70	Total	204	73.68	4.02

[1] The highest possible raw score on this performance check was 79.

The examiner means in both tables show considerable variation, probably more than would be expected by chance. The variations among the examiner sigmas is even more striking. In fact, the tables show not only that examiners varied in the average score given, but also that some examiners gave scores in a much more restricted range than others. An inspection of the tables, particularly table 6.2, suggests a negative correlation between means and sigmas. This may be due, at least in part, to the closeness of most of the scores to the highest possible score on the check, leaving little room for upward variation.

In attempting to interpret the examiner data, the assumption is made that students checked by an examiner are a random sample of the population of students. While it is not known that any factors produced a biased sample of students for any examiner, neither can it be stated that any precautions against a bias were taken. With this qualification, it seems likely that the examiner means and sigmas do not have the consistency which one would desire from standardized administration.

As pointed out earlier, high scores on the bench checks led to a greater emphasis on the inclusion of precision items. Another remedy proposed, but not actually applied, was that the amount of time the student required to go through the check be considered in his grade. To examine this possibility, time required to complete the check was recorded on the scoring sheets for several classes at Victorville Army Air Field. A coefficient of correlation computed between these time scores and the raw error scores on a sample of 209 cases was 0.45. This correlation is low enough to suggest that a combination of check and time scores would produce a more discriminating total grade.

There were difficulties, however, in timing the bench check. Equipment differences influenced the speed with which the student could perform the required steps. For instance, if it were more difficult to pick up a target on one set than on another, equally proficient students might earn different time scores. Also, the examiner might influence student speed since he was required to stop the student and correct him. Before time scores can be used to grade students, both equipment and examiner variables will require considerable control.

STANDARDIZED SUPERSONIC TRAINER PERFORMANCE CHECKS

The Supersonic Trainer and the Student's Task

The supersonic trainer is a ground school device which simulates air-borne navigation and bombing with radar equipment.[6] The trainer consists of a standard radar set, artificial terrain submerged in water, glass map, and control panel. The antenna of the trainer transmits and receives wave impulses above the frequency of sound, rather than radio waves as in standard radar equipment. This antenna is moved over the artificial terrain by a system of electrically driven pulleys. The effect is almost identical to the movement of an air-borne radar antenna over the ground. The terrain of the trainer is made of glass, sand, and carborundum so that returns from it, when seen on the radar scope, duplicate essentially returns from water, land, and cities. The movement of the antenna is determined on the basis of the true air speed, heading, and wind which are regulated from a control panel. It is thus possible to present on the ground, radar navigation and bombing missions which are strikingly similar to air-borne missions.

A feature of the apparatus of particular value to the development of standardized supersonic performance checks is the glass map with marking pen attachment. This map is directly above the artificial terrain and represents every feature of the terrain. A marking pen follows the course of the antenna, corresponding to the track of an aircraft, making a continuous record on paper stretched beneath the map. Thus a complete record of the mission is made, including course corrections and bomb drops.

The instructional purpose of the supersonic trainer was to teach the use of radar equipment for navigation and bombing. In successive training periods, the task became progressively more complex,

[6] For a fuller description of the equipment, see chapter 4.

developing from simple navigation to coordinated bombing missions. The student learned scoped interpretation, fix taking, log procedure, ETA computation, precision turn procedure, drift correction, and the technique of bomb release.

Three performance checks were constructed for the supersonic trainer. One was an intermediate navigation check given early in the course after basic navigation techniques had been taught. The second was an intermediate bombing check given after instruction in coordinated bombing procedures. Third was a final check to evaluate the student's performance on a combination navigation and coordinated bombing mission at the completion of supersonic training. Each of the three checks was planned to require 40 to 50 minutes per student.

Supersonic Check Construction

The selection of critical items to cover the tasks of the supersonic missions was much more difficult than had been the selection of items for the bench set check. On the supersonic trainer, it was impossible to evaluate every aspect of behavior during an entire mission. Measurement had to be limited to a sample of the possible performance. This sample was chosen in conferences with instructional personnel. For instance, it was agreed by training personnel that the proper way to start a navigation mission was to take a fix soon after departure. Since there was complete accord on this, since students were taught to do it, and since it was susceptible to objective measurement, the appropriate item was constructed. In this same manner, critical items were chosen to cover the remainder of the mission.

The Typical Supersonic Check

Discussion of the typical supersonic check will be confined to the final check since this includes both the navigation and bombing items. The typical final check mission consisted of one navigation leg and two coordinated bomb runs.[7] The student's task was to navigate to the first initial point, make a precision turn onto the bomb run heading, and bomb the first target. He was then to give a new heading from this target and make a bomb run on the second target. On the navigation leg the student was checked on 17 items which evaluated such performance as his wind determination procedure, the accuracy of the velocity and direction of his computed wind, the accuracy of his course correction, his log procedure, and the accuracy of his ETA and fixes. He was checked once on each bomb run for each of eight items which evaluated the accuracy of his drift corrections, the accuracy of his announced sighting angles, and the accuracy of his bomb hit.

The introductory material to this check contained the conventional remarks to the instructor, the statement of training prerequisites, and

[7] For a description of coordinated bombing, see chapter 4.

the equipment needed by student and examiner. The characteristics of the mission were stated and, more important, the actual mission to be flown was defined in detail. Necessary data for the mission were reproduced on a separate sheet from which the student prepared his flight plan. In a later form of the check, the flight plan itself was mimeographed on a special student log. This saved time and effected further standardization. The student was given altitude, true air speed, starting point, initial points, targets, and the true courses and distances between these points. Since the determination of the wind was perhaps the most critical item, the student's mission plan did not include wind velocity or direction. Wind information which was to be set into the trainer was given to the examiner in the directions for administration. The student briefing section, which the student read for himself, informed him of certain special requirements of the mission.

The definition of the standard mission in terms of specific routes and targets was essential to check standardization. A given item presented an equivalent task to all students only when the raw material of the item task was the same for each. Since different areas varied in the number of usable radar returns presented, it was necessary that the terrain variable be kept constant by having all students cover the same or strictly comparable routes. Similarly, equivalence in length of mission legs was important since speed of performance was a factor in allowing opportunity to accomplish all the required items.

A refinement of the last form of supersonic check was the inclusion of a working plan for the examiner which outlined the easiest and most efficient way to administer the check. Another refinement inaugurated the sub-item, which was expected to improve check administration. The sub-item was used in any item in which more than one condition had to be satisfied for the student to earn a check. For such items separate check spaces were provided on the scoring sheet. This breakdown had two objectives. It eliminated the need for the examiner to keep several conditions in mind and it prevented him from giving credit because the student got "most of the item" right. Clerical workers who totaled the check marks were instructed to count one check for an item only if all of the sub-items were checked.

Types of Supersonic Check Items

The items which constituted the supersonic checks were of several types. One major distinction is between items which evaluated correct performance procedure and those which measured precision of performance. The procedure items, which comprised about one third of the items on late forms of the final check, were of two kinds. One was a simple observation of student behavior such as:

Interphone procedure.—Check only if student gives time of fix and range(s) and/or bearing(s) of target to navigator immediately after recording them in his log.

The other was a simple examination of the student's log:

Log procedure: Ground speed and ETA.—Check only if student at time of course correction (see item 6)

(*a*) logs ground speed.

(*b*) logs ETA to turning point.

The project developed several devices to aid in evaluating precision on the trainer. One of these was the method of marking fixes on the glass map so that they could be compared accurately with the corresponding plotted fix on the student's chart. This was accomplished by means of the bomb release switch which, on a direct bomb run, caused the marking pen to fall away from the glass map and thus break the ink line. On a bomb run the pen did not mark again until the trainer stopped after the time of fall of the bomb had run out. However, if the reset switch on the control panel was pushed before the time of fall ran out, the pen would continue marking at once, and the trainer would not stop. Therefore, the student was briefed to push the bomb release switch at the instant he took a fix which he planned to plot. He was to call "fix" simultaneously to warn the examiner to push the reset switch about 10 seconds later. Thus every fix was marked by a break in the ink track line on the glass map. On models of the trainer not equipped with a bomb release switch, the student simply called "fix" and the examiner produced the break in the ink track line by use of the recording pen switch on the control panel.

Another aid to precision measurement developed by the project was the supersonic plotter. The first and most simple plotter was a wide, plexiglass straight-edge on which several scales were etched.[8] One scale along the edge was calibrated in glass-map nautical miles so that distances would be read directly on the glass map itself. Another scale consisted of a length-wise center line with various parallel lines so that, by placing the center line over any course, distance of the track away from that course could be read directly. Two scales of concentric nautical mile circles allowed for reading distance away from any point over which the center of the circle was placed.

The latest plotter consisted of a square plexiglass base, etched with a protractor scale, and bearing two rotatable plexiglass arms.[9] With this device, fixes could be transferred from the student's chart to the glass map without computation or additional mechanical aids. Angular differences between desired course and actual track could also be determined easily. Other scales etched on the square base facilitated the measurement of ETA and bomb hit accuracy.

---

[8] This plotter was designed by Sgt. Mangan and constructed by T/Sgt. George N. Bollinger.

[9] This plotter was designed by Cpl. Kelley and constructed by T/Sgt. Bollinger.

To make the measurement of precision at all feasible it was necessary to take account of certain inaccuracies in the trainer. For example, the wind which the trainer produced did not always correspond exactly to the wind set in the control panel. Consequently, since the student was graded on the accuracy of his computaton of wind direction and force, a better standard than the wind set in the trainer was needed. A more accurate standard was provided by having the examiner determine ground speed and track by measurement on the glass map. By combining these data with the true air speed and heading given to the student in the mission plan, the examiner could determine the wind which the trainer was actually producing.

Precision items accounted for about two thirds of the items on the latest forms of the final supersonic check. Such items may be grouped into several types. One type is a simple comparison of data logged by the student with comparable data set into the trainer. A second type involved an observation of behavior plus a simple computation. For example, the accuracy of the student's multiple drift correction was determined by having the examiner note all the course corrections given by the student after the multiple drift correction and add them algebraically to see if they totaled 4° or less. This item illustrates also indirect measurement of performance in that measurement was applied to subsequent behavior causally related to the behavior for which evaluation is desired. In this instance the accuracy of the multiple drift correction, which is the point of concern, could be inferred from the subsequent correction necessary to bring the aircraft over the target. A third type of evaluation of precision was accomplished by measurements on the glass map. In evaluating the bombing deflection er: ., for example, the examiner laid the plotter along the inked track at "Bombs Away" and determined whether the track or its extension came within one nautical mile of the briefed aiming point. A fourth type of precision item compared the inked track record on the glass map with the student's chart work. In such items the accuracy of the student's fixes was determined by transferring the fix data from the student's chart to the glass map by means of the protractor and determining whether his fix was accurate within the stated tolerance. A final type required the examiner to judge the quality of the student's performance, and, if necessary, to compare it with his own. In evaluating set operation, for instance, the examiner was to read range and azimuth while the student was taking a fix and judge whether the scope returns were defined well enough to make accurate readings possible. If he thought they were not, he was to adjust the radar set himself to determine whether better definition was possible. In later revisions of the supersonic checks, there were no items which depended upon this degree of examiner judgment.

For most precision items, such as the accuracy of wind computation, it was necessary to establish acceptable limits or tolerances. The tolerances for the earliest form of the final supersonic check were based upon a study of the mean errors of a trial group of 40 radar students taking a preliminary version of the check. Later forms included tolerances based on more precise study of item difficulty levels.

Supersonic Check Administration

Difficulties in maintaining standard administration of the supersonic checks were numerous. Variability in equipment accounted for many of these. As noted above, data set into the trainer did not always produce the expected result. Most of this difficulty was compensated for, however, by having the examiner compute the wind from the trainer in the manner already described. The recording pen sometimes failed to mark or ran out of ink during the mission, making items to be measured from the glass map difficult or impossible to evaluate. On other occasions, the pen marked adequately but failed to drop so as to mark fixes and ETA's. The examiner was forced to use his fountain pen, if he had one, and lost valuable time in marking by hand. Some trainers tended to stick on cardinal headings which confused the student and made the track line an inaccurate index of the student's intention. The most persistent trainer difficulty was the misalignment of the glass map with the artificial terrain. If the difference between the two was great, it was clearly unfair to take the glass map record as indicative of the student's performance. Many attempts were made to devise a method for checking exactly the alignment of map and terrain, but no solution had been obtained by the time work on the checks was discontinued. Only gross differences were readily recognized and corrected.

Examiner variability accounted also for lack of standardization. As on the bench check, the halo effect was in evidence. In addition, inadequate student briefing occasionally resulted in the student proceeding with the mission without sufficient knowledge of what was required of him. Also, some examiners failed to check each item as it was performed, and scoring by memory, was, of course, unreliable. Other examiners estimated error instead of carrying out the more exact measurements required by the administrative directions. Finally, some examiners seemed to identify with the student and could not avoid hints and leading questions which helped the student correct what they regarded as "foolish" errors.

Supersonic Check Revisions

Four forms of the Final Supersonic Check were published. One (P7a–A) was used at Langley Field and Boca Raton for the AN/APS–15, AN/APS–15A, and AN/APQ–13.[10] Another

[10] Supersonic Check P7a–A was constructed by Sgt. William J. Mangan and Sgt. Sheldon H. Nerby under the supervision of Capt. Horace R. Van Saun.

(L P7a–B), used at Langley Field, was also applicable to each of these sets.[11] One (V P7a–A) was used to check proficiency with the AN/APQ-13 equipment in the Western Flying Training Command while the fourth (W P7b–A) was used at Williams Field to evaluate performance with the AN/APQ-7.[12] The two most recent forms of the final supersonic check incorporated numerous improvements and additions. Appendix B includes one of these, the last in use at Langley Field. The other, prepared for use with AN/APQ-13 in the Western Flying Training Command, is different in certain respects. In addition to requiring the checking of sub-items, as described earlier, it reduces instructions to the examiner to the minimum necessary for checking typical performance. Evaluation of a typical performance was described in a section of the booklet called Unusual Situations. This variation in format represented an adjustment to a recurrent problem encountered in the preparation of material for check administrators who were unsophisticated in the requirements of standardized measurement. It was necessary constantly to compromise between the advantages of exhaustive detail in administrative instructions and the tendency observed in examiners to be confused and antagonized by such detail.

Six forms of the Radar Navigation Intermediate Supersonic Check were produced. P7e–A and P7e–B were used in conjunction with the AN/APS-15 and AN/APS-15A at Langley Field, and the AN/APQ-13 at Boca Raton. Form L P7e–C, applicable to these same sets, was used at Langley Field. Form V P7e–A (AN/APQ-13) was developed for the Western Flying Training Command. Form W P7e–A was a temporary form used at Williams Field with the AN/APQ-13 set. Form W P7f–A measured navigation skill on the AN/APQ-7 equipment.[13]

Three forms of the Radar Bombing Intermediate Supersonic Check were developed. Form P7c–A, used for AN/APS-15 and AN/APS-15A and AN/APQ-13, was administered at Langley Field and Boca Raton. Form L P7c–B, usable on all three sets, was given at Langley Field. Form V P7c–A was used with the AN/APQ-13 in the Western Flying Training Command.[14]

Supersonic Check Statistical Findings

Figures 6.2, 6.3, and 6.4 present diagrams of typical distributions of scores on the navigation, bombing, and final check, respectively.

---

[11] Sgt. Mangan developed this check.

[12] Lt. George S. Klein and Cpl. Bray with the assistance of Sgt. Nerby constructed form V P7a–A, while form W P7b–A was developed by Cpl. Bray.

[13] Sgt. Mangan and Sgt. Nerby were responsible for the development of three of the six Radar Navigation Intermediate Supersonic Checks. The other three checks were constructed by Lt. Klein, Cpl. Bray, and Cpl. Harold H. Kelley.

[14] Chief contributors to the development of the Radar Bombing Intermediate Supersonic Checks were Sgt. Mangan, Sgt. Nerby, Cpl. Bray and Lt. Klein.

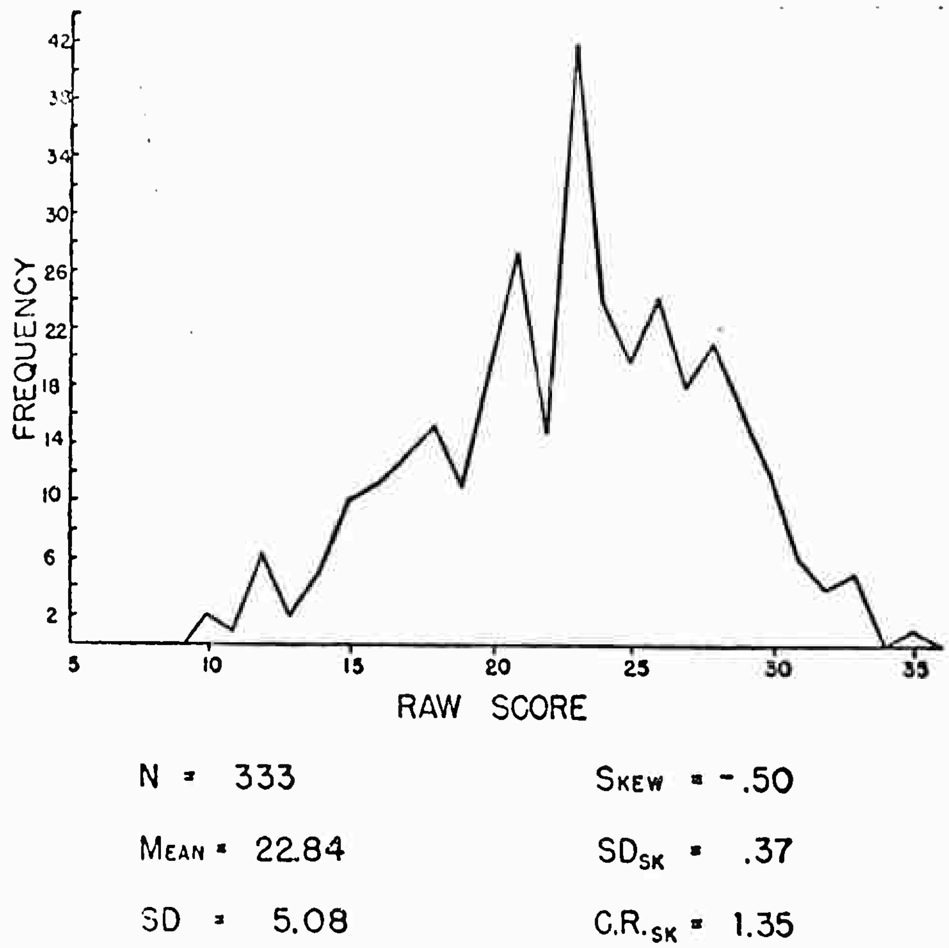
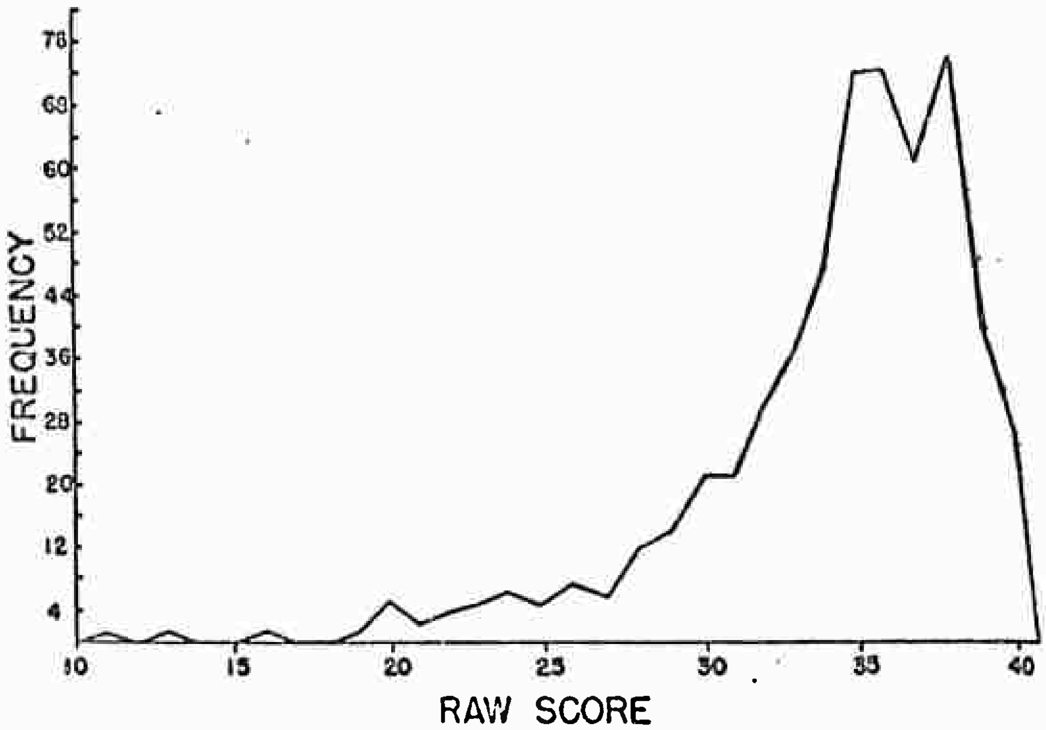


FIGURE 6.2—Frequency distribution of raw scores for a standardized radar navigation intermediate supersonic check (P7c–CX)—administered at Langley Field.

The navigation check had a possible maximum raw score of 37. The mean score of 22.8 indicates that the check was sufficiently difficult. The distribution is not significantly skewed. The bombing check had a total maximum raw score of 40. The mean score of 34.2 and the fact that many students made the highest possible score shows that this check was too easy. The distribution is significantly skewed and has a narrow range. The final check, combining both navigation and bombing items, had a possible maximum raw score of 54 and a mean score of 43.0. This distribution is also significantly skewed.

The product moment correlation between check scores on the first and second legs of the navigation check was 0.361 (N=189). The correlation between scores on the first and third legs as against the second and fourth legs of the bombing check was 0.466 (N=190). The first and second bomb runs of the final check correlated 0.390 (N=516). These coefficients, while of interest, are not presented as indices of reliability. The only acceptable measure of reliability on a check of this sort is a test-retest coefficient. The modified split half



N = 576

Mean = 34.18

SD = 4.81

Skew = −1.50

SD_{SK} = .24

$C.R._{SK}$ = 6.25

FIGURE 6.3.—Frequency distribution of raw scores for a standardized radar bombing intermediate supersonic check (P7c–A)—administered at Langley Field.

method which yielded the correlations given above is not acceptable because the same examiner administered both parts of the check and testing conditions were the same for both halves.

A test-retest reliability study, the only one attempted, resulted in a coefficient of 0.49 (N=79). This is significant at the 1-percent level, the probable error of a zero correlation based on 79 cases being 0.076. The experiment was unstandardized, in part, because of VJ-day, which was announced near the completion of the check administrations. It is feared that this modified the usual student and examiner behavior. Previous cases show a first administration mean of 32.19 and a recheck mean of 33.65 (N=55). The corresponding data for cases administered after VJ-day were 31.42 and 35.25 (N=24). It was recognized during the experiment that the examining board which administered the checks were not motivated to provide well-standardized administration. Consequently, relatively low reliability was expected.[18]

---

[18] Capt. Gabriel D. Ofiesh supervised this reliability study. He was assisted by S/Sgt. Bernard C. Sullivan and Sgt. Mangan.

Tables 6.3 and 6.4 show examiner means and sigmas for final supersonic checks given at Langley Field and at Boca Raton. Again it was impossible to use the analysis of variance technique since a chi-square test of the Langley data for homogeneity of variance resulted in a value of 93, 32 being needed for significance at the 1-percent level. The same test on the Boca Raton figures yielded a value of 37, 20.09

TABLE 6.3.—*Examiner means and standard deviations for a standardized final supersonic trainer performance check (P7a–A) administered at Langley Field*

Examiner	Number of checks administered	Mean raw score [1]	SD	Examiner	Number of checks administered	Mean raw score [1]	SD
A	14	45.79	4.21	K	13	41.46	6.96
B	30	45.79	4.01	L	18	41.06	4.45
C	40	44.78	3.92	M	8	40.38	7.73
D	23	44.52	3.64	N	19	39.74	6.07
E	13	44.46	4.18	O	6	39.67	5.91
F	21	43.57	4.28	P	15	38.47	6.69
G	20	43.10	5.12	Q	22	36.41	8.69
H	18	42.11	5.24				
I	26	42.04	4.92	Total	319	42.37	5.21
J	43	41.86	4.62				

[1] The highest possible raw score on this performance check was 54.

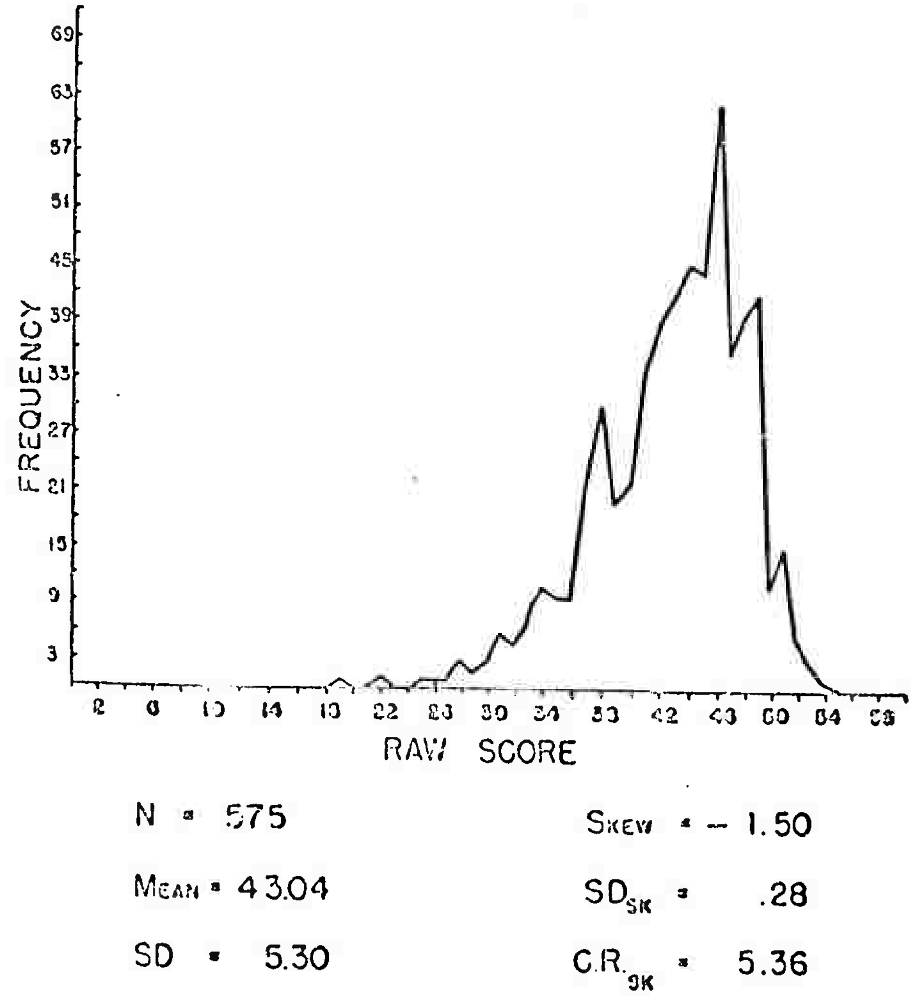


FIGURE 6.4.—Frequency distribution of raw scores for a standardized final supersonic check (P7a–A)—administered at Langley Field.

TABLE 6.4.—*Examiner means and standard deviations for a standardized final supersonic trainer performance check (P7a–A) administered at Boca Raton*

Examiner	Number of checks administered	Mean raw score [1]	SD	Examiner	Number of checks administered	Mean raw score [1]	SD
A	8	42.88	5.04	G	9	35.11	3.07
B	7	41.00	3.38	H	6	31.17	9.19
C	6	40.17	2.10	I	8	30.88	4.87
D	6	37.00	4.50				
E	8	35.88	4.01	Total	66	36.61	4.79
F	8	35.75	4.60				

[1] The highest possible raw score on this performance check was 52.

being needed for significance at the 1-percent level. Both sets of data reveal wide variation in means and sigmas. With the same qualifications as were stated in the similar discussion of the bench set data, it may be concluded that standardized administration was not achieved.

STANDARDIZED AERIAL PERFORMANCE CHECKS

The Airborne Set and the Student's Task

The most heavily weighted and academically important performance checks were those developed to evaluate the student's proficiency as a radar observer under actual flight conditions. Apparatus involved in the administration of these checks was the air-borne radar set and auxiliary flight instruments. Briefly stated, the radar observer's task [10] was to direct the aircraft on the briefed course to the initial point and make an accurate bomb run on the briefed target. To this end, the objective of training was to produce students able to operate the set; to compute accurate winds, course corrections, ETA's and other navigational data; to kill drift, determine absolute altitude, and bomb accurately in collaboration with the bombardier; and to keep a complete and accurate log.

Two aerial performance checks were developed: the intermediate aerial check given mid-way in aerial training after the essentials of radar navigation and bombing had been taught, and the final aerial check given at the completion of the course. The latter check will be discussed first.

Aerial Check Construction

The method of making the original selection of items for the aerial check was similar to that outlined above in the discussion of the supersonic check. Project personnel took numerous flights to make systematic observations of the task of the radar observer, enrolled in the radar course, and worked in conference with training authorities. Then followed the selection of critical aspects of radar observer performance on an aerial mission. Developing a method

[10] For a job analysis of the radar observer, see ch. 4.

of measurement for these critical items was more difficult here than for other performance checks because of the difficulty of finding objective criteria of success.

The Typical Aerial Check

The Aerial checks contained the conventional introductory sections previously described. The introductory material included general comments on performance checks, a listing of the equipment needed by student and examiner, the training prerequisites, detailed instructions for administration, and the characteristics of the standard mission. The aerial checks also had a section containing supplementary instructions for each item; this followed the answer sheet. These instructions described the purpose of the item, how to check the item, and special situations which might arise while checking the item. The list of items was printed on cut-back pages with matching answer sheet.

The description of the standard mission was used to select specific routes over which the check mission was required to be flown. A standard mission was necessary for aerial checks for reasons already stated in the discussion of supersonic trainer performance checks. If anything, a standard mission was needed even more in the aerial situation where other variables such as weather and turbulence were difficult to control.

The final aerial check required for its administration approximately five hours, a complete mission consisting of four or five navigation legs each follwed by a bomb run. The intermediate check required 2½ hours and included either two or three navigation and bombing legs.

Types of Aerial Check Items

The items comprising the aerial check can be divided, as in the checks previously discussed, into the broad categories of those which measure procedure and those which measure precision of performance. Procedure items, which accounted for about one-third of final check items, were of two types. One of these types called for a simple observation of student behavior. For example, in evaluating performance of the preoperational check, the examiner merely noted whether the students adjusted certain designated controls in the proper directions. The second type required only an examination of the student's log or chart. For example, in determining whether the student had taken fixes with sufficient frequency, the examiner was required to inspect the log and chart and from them determine whether more than 15 minutes had elapsed between successive fixes.

The precision items amounting to about two-thirds of the check, were of four types. One type involved an observation of behavior plus a simple computation. Such an item was the evaluation of the accuracy of the student's course correction. In part, the task of the examiner was to note and add algebraically the changes in heading which the student called to the pilot after making the course correction and before reaching the turning point near the initial point. A second type of precision item was based on an evaluation of objective results but depended on the accuracy of the examiner's observation. For example, the further condition for receiving credit for an accurate course correction, (see sample item above) was that the aircraft pass within 2 miles of the correct turning point near the initial point. This required the examiner to compute the correct turning point in relation to the radar return representing the initial point and determine by measurement on the scope how close the aircraft came to this point. A third type of item used for evaluating precision of performance required a comparison of the student's performance with that of the examiner. The accuracy of the multiple drift correction, for example, was evaluated by having the examiner determine the correction simultaneously with the student. He was to give credit only if the student's correction was within 5° of his own estimate. The fourth type of precision item was based on the examiner's judgment. For example, in evaluating the student's skill in operating the set so as to define adequately the aiming point on the scope, the examiner's task was to judge whether the definition was the best possible within the limits of the set. Only a single item of this nature was included in the check.

The intermediate check differed slightly from the final check at two of the radar schools. Besides being a shorter check, as previously stated, the first few items were oral questions concerning trouble shooting and the location of certain units of the equipment. In addition, the bomb runs were by direct rather than by coordinated procedure. In the other schools the intermediate check was identical to the final check except that it was given on two instead of four legs.

Aerial Check Administration

Standardization of aerial check administration was most difficult. The equipment, including the aircraft, accounted for some of the variation. Various malfunctions were likely to develop during check administration. Since these were frequently not serious enough to warrant returning to the base, the mission was completed. In such cases, it was impossible to determine precisely how much the malfunction interfered with the student's performance. One such equipment difficulty was that of varying range of the radar set. One set might show returns from 80 miles away, another might "get" only 50 miles.

Weather was another condition of check administration which could not be standardized. Bad weather on the briefed course involved flying around it; thunderheads interfered with radar returns, and more than usual turbulence affected performance.

Pilot variation was greater than might be expected. Some pilots were conscientious and followed interphone instructions from the radar observer quickly and accurately. Others were not as cooperative, a fact which markedly hindered the student's work. Delay by the pilot in making a course correction on the bomb run, for example, might spoil what otherwise would have been a good performance.

Variation in altitude also resulted in some lack of standardization. Although pilots were briefed for a given check flight altitude, this instruction was not always carefully followed. Differences in altitude changed the difficulty of some items. For example, synchronizing with the bombardier on the bomb run is a more leisurely matter at 20,000 feet than at 10,000 feet because the sighting angles follow one another more slowly. Altitude variation, moreover, required some students to work in an oxygen mask while others were not so encumbered.

In an effort to take account of variations which practical circumstances made it impossible to control administratively, an addition was made to the scoring procedure. If the student failed to meet the standards prescribed for an item, regular practice called for the examiner to enter a zero in the scoring space regardless of whether or not the failure was for reasons beyond the student's control. However, if the examiner believed that the student's performance on a failed item was as satisfactory as it could be under adverse circumstances, he was now instructed to enter an "S" with the zero. Obviously, this attempted compensation for nonstandard conditions injected relatively uncontrolled examiner judgment into a check designed to eliminate it. To offset this weakness, it was provided that all "S" scores would be reviewed by the chief examiner. To accumulate experience for standardizing his judgment, the chief examiner was directed to determine exactly the conditions leading to each "S" score and keep a record of which of these conditions warranted the giving of credit for the item. To aid in this process, it was arranged that each examiner was to keep a mission log on which he noted the reason for each "S" score.

As on the other performance checks, the examiners themselves were responsible for departures from standard administration. The halo effect, hints, and leading questions were in evidence here as elsewhere. Failure to mark items as they were performed resulted in further inaccuracies. One reason for poor administrations of aerial checks was the difficulty which the examiners had in obtaining the data in terms of which student performance was checked. Many of these data had to be taken from the scope itself. This meant that both student and examiner had to use the same scope, frequently at the same time. This difficulty was partially solved by blacking out the radar observer's compartment so that the hood on the scope could be removed. Under these conditions student and examiner could watch the returns simultaneously although the examiner had to move frequently to avoid errors due to parallax. In cases where the compartment was not blacked out and the hood had to be used, the examiner had the uncomfortable task of leaning over the student's shoulder and taking readings as best he could. When the discomfort of heavy flight clothing and oxygen mask is added, the exacting and strenuous nature of the job of an examiner on a 5-hour-check ride can be appreciated.

Aerial Check Revisions.

Only two forms of the Intermediate Aerial Performance Check were necessary. One (P8b–A) was used to evaluate aerial performance with the AN/APS–15 and AN/APS–15A at Langley Field and the AN/APQ–13 at Boca Raton. The other (V P8b–A) was used with the AN/APQ–13 in the Western Flying Training Command.

Four forms of the Final Aerial Performance Check were prepared. Two (P8a–A and P8a–B) were applicable to the AN/APS–15 and AN/APS–15A at Langley Field and the AN/APQ–13 at Boca Raton. Another (V P8a–A) was used with AN/APQ–13 in the Western Flying Training Command. The fourth (W P8c–A) was used to evaluate performance with the AN/APQ–7 set at Williams Field.[17]

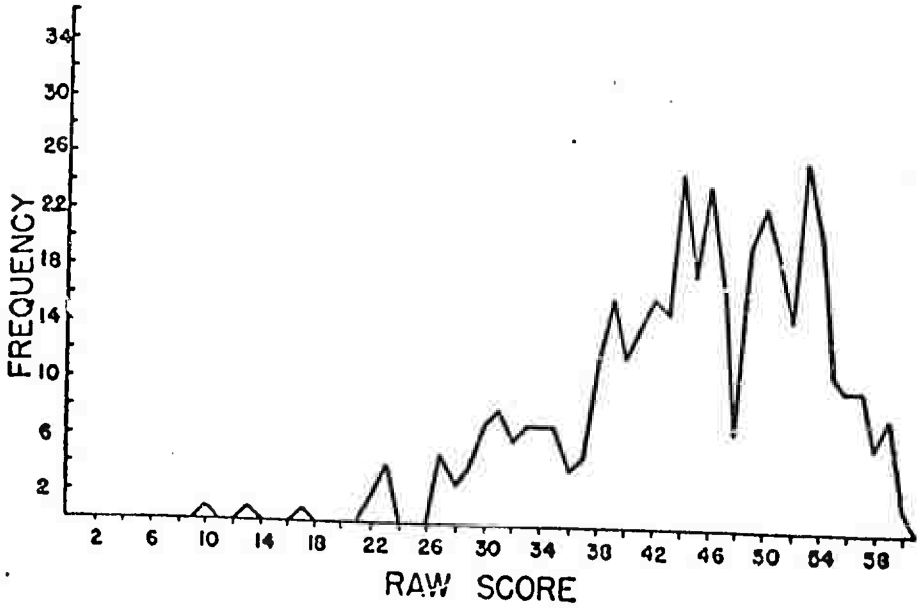
Aerial Check Statistical Findings

Figures 6.5 and 6.6 present typical distributions of scores from the intermediate and final aerial checks. The intermediate check shows a mean raw score of 44.9 out of a maximum possible score of 62. The distribution of check scores is skewed significantly. The final check yielded a mean of 59.4 out of a maximum possible score of 76. Skew for this distribution is not statistically significant.

Tables 6.5 and 6.6 present examiner means and sigmas for final aerial performance checks administered at Langley Field and Boca Raton. Here again it was impossible to use the analysis of variance technique. A chi-square test for homogeneity of variance of the Langley data resulted in a critical ratio against homogeneity significant at the 1 percent level. For the Boca Raton data a chi-square equalled 69 with only 27.69 needed for significance at the 1 percent level. Both tables reveal large differences in examiner means and a surprising variation in standard deviations.

A correlation between the score made on the first, third, and fifth legs of the final aerial check and the second and fourth legs yielded a coefficient of 0.466 (N=90). For reasons already given in discussing the supersonic checks, care should be taken not to interpret such figures as measures of reliability.

[17] Early developmental work on Aerial Performance Checks was performed by Capt. Van Saun and S/Sgt. Mitchell with the assistance of Sgt. Blum and Sgt. Gage. Later checks were constructed by Sgt. Philip H. Kriedt, Sgt. Gage, Cpl. Kelley, Cpl. Bray, and Lt. Klein.



N = 422

MEAN = 44.84

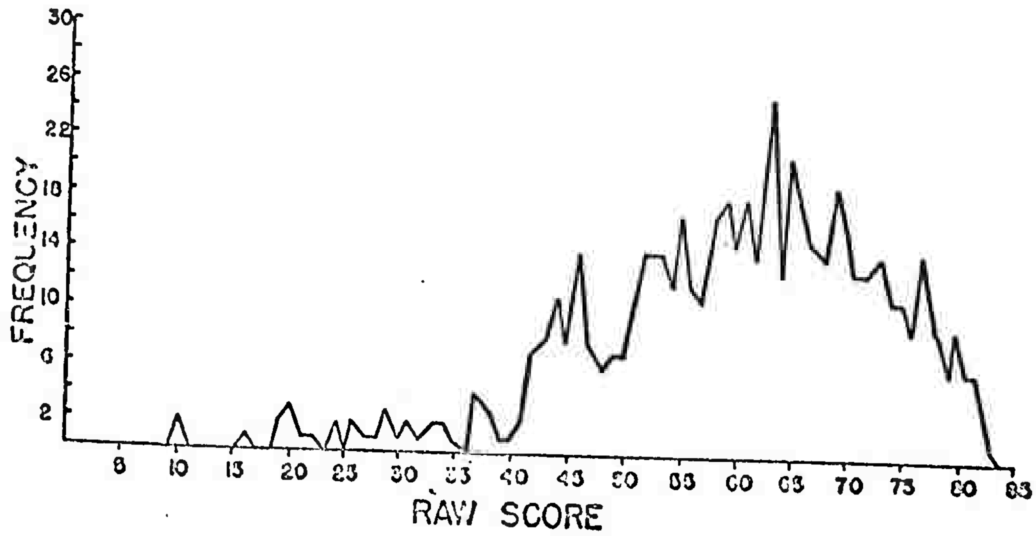
SD = 8.79

SKEW = − 2.50

SD_{SK} = .58

$C.R._{SK}$ = 4.31

FIGURE 6.5.—Frequency distribution of raw scores for a standardized intermediate aerial check (P8b–A)—administered at Langley Field.



N = 543

MEAN = 59.41

SD = 14.10

SKEW = −1.50

SD_{SK} = .73

$C.R._{SK}$ = 2.05

FIGURE 6.6.—Frequency distribution of raw scores for a standardized final aerial check (P8a-B)—administered at Langley Field.

TABLE 6.5.—*Examiner means and standard deviations for a standardized final aerial performance check (P8a-B) administered at Langley Field*

Examiner	Number of checks administered	Mean raw score[1]	SD	Examiner	Number of checks administered	Mean raw score[1]	SD
A	17	75.12	9.37	U	20	58.05	8.16
B	16	74.19	5.51	V	26	57.92	10.63
C	6	73.33	5.73	W	8	57.63	9.27
D	10	70.10	7.80	X	6	57.50	17.62
E	10	70.10	9.20	Y	20	57.35	11.00
F	11	69.27	5.80	Z	19	56.95	9.29
G	13	68.54	5.53	A	21	55.00	9.74
H	7	68.29	11.28	B	20	51.14	17.45
I	9	67.67	7.20	C	10	51.80	9.67
J	38	67.32	14.53	D	17	51.47	13.09
K	10	67.00	5.01	E	13	50.92	8.11
L	16	66.19	6.20	F	8	50.50	9.37
M	20	64.80	11.30	G	21	49.90	10.99
N	22	64.45	6.90	H	9	49.89	15.02
O	15	61.73	12.35	I	14	49.36	7.83
P	7	61.43	3.54	J	13	49.15	17.53
Q	8	61.00	10.54	K	9	48.00	11.76
R	14	60.21	10.28	L	17	47.47	10.34
S	8	58.88	9.74				
T	13	58.08	14.47	Total	511	60.02	10.74

[1] The highest possible raw score on this performance check was 86.

TABLE 6.6.—*Examiner means and standard deviations for a standardized final aerial performance check (P8a-B) administered at Boca Raton*

Examiner	Number of checks administered	Mean raw score[1]	SD	Examiner	Number of checks administered	Mean raw score[1]	SD
A	9	67.22	11.43	I	6	52.83	15.28
B	10	62.60	12.86	J	8	52.25	12.20
C	12	61.33	11.58	K	12	50.83	11.98
D	13	59.31	15.82	L	12	50.58	9.59
E	11	57.18	10.63	M	10	49.20	10.95
F	12	57.00	11.98	N	8	48.63	8.03
G	11	55.73	10.95				
H	10	54.30	12.63	Total	144	55.84	12.05

[1] The highest possible score on this performance check was 81.

The Student Rating Research Form[18]

An addition to the later form of the aerial check was the Student Rating Research Form. The first part of the form provided space for the examiner to describe variables which he felt influenced the student's performance check score. Spaces were provided for set malfunction, weather conditions, crew, and other variables. The second part consisted of four rating scales. The examiner rated the student on set operation, navigation proficiency, bombing proficiency, and over-all proficiency. The examiner rated the student in relation to all the students to whom he had given performance checks and on the basis of how well he thought the student would have done under "normal conditions."

Sample items:

First part.—Did the crew interfere with the student or give him help when they shouldn't have? Explain.

[18] Cpl. Kelley developed the Student Rating Research Form.

Second part.—1. Set operation: Quality of scope picture the student maintained through the mission.

Lowest 10%	Next to lowest 10%	10%	10%	10% just below average	10% just above average	10%	10%	Next to highest 10%	Highest 10%

The purpose of the Student Rating Research Form was to make an initial investigation into the possibility of using examiner comment and evaluation to counteract some of the uncontrolled variability in the administration of the aerial performance checks. Aerial checking was terminated before this possibility could be tested adequately, but statistical findings on the few cases at hand are of interest. Computations were made only for the over-all proficiency rating. At Langley Field 66 students received both check scores and proficiency ratings on the final aerial check. The coefficient of correlation between these two values, as shown in table 6.7, was 0.565. The same correlation at Vic-

TABLE 6.7.—*Product-moment correlations between raw scores and instructor ratings on aerial performance checks*

Performance check	N	Raw score M	Raw score SD	Rating M	Rating SD	r
Intermediate Aerial Check Form P8b-A-(Langley)	49	21.97	5.81	59.59	16.15	[1] 0.777
Final Aerial Check Form P8a-B-(Langley)	66	57.44	14.71	51.66	20.71	[1] .565
Intermediate Aerial Check Form V P8b-A-(Victorville)	206	42.37	7.01	50.70	15.20	[1] .682
Final Aerial Check Form V P8a-A-(Victorville)	333	77.41	13.86	65.40	16.19	[1] .517

[1] Significant at the 1-percent level of confidence.

torville on 333 cases was 0.517. Similar studies on the intermediate check yielded at Langley Field, 0.777 (N=49), and, at Victorville, 0.682 (N=206). It must be remembered that in this study the instructor made his rating immediately after administering the check, and it is reasonable to assume that his judgment was influenced by the student's success on the check. Consequently, it cannot be said with certainty that objective check scores and instructor's ratings are correlated to the extent indicated by these coefficients.

Of note was the strong tendency of the Victorville examining board to give higher ratings on the final check than on the intermediate check. This result was obtained despite the fact that examiners were briefed to compare each student with other students at a similar stage of training.

Further studies involving these ratings will be found in chapter 8 on the interrelations of proficiency measures.

SUMMARY

The necessity of supplementing written tests of radar proficiency with performance measures led Psychological Research Project

(Radar) to construct six standardized performance checks. These checks were for the purpose of evaluating student proficiency on airborne radar missions, on supersonic trainer missions, and on the classroom radar equipment. Each check was constructed on the basis of a study of the task, conferences with instructional personnel, and trial administrations. A uniform format was provided for all checks and a special arrangement of item pages with an expendable scoring sheet facilitated administration. All checks contained a statement of standard conditions under which the check was to be administered. Standardized administration was sought by the training of specialized examiner boards at each radar school. Several revisions of each check were made to provide measures applicable to different radar equipment and different radar schools.

Performance on classroom radar equipment was measured once, at the completion of bench set training. Proficiency in starting, tuning, and calibrating the set was evaluated. Every step in the operating procedure was included in the check and performance of many steps were required in prescribed sequence. A by-product of early check development was the impetus given for the development of a standard operating procedure for purposes of instruction. The majority of bench check items were aimed at evaluating the procedure of adjusting the various controls rather than the quality of the scope picture resulting from the adjustments.

Performance on the supersonic trainer was checked at three points. The intermediate navigation check evaluated proficiency early in the trainer course. The intermediate bombing check measured proficiency in bombing after the appropriate instruction had been given. The final check evaluated performances on a complete radar navigation and bombing mission. Supersonic check items dealt both with adherence to standard navigation and bombing procedure and with the accuracy of the student's results. The graphic record provided by the trainer served as a useful source of performance data. As essential feature of the supersonic checks was the specific mission route for each check; this assured that check items presented equivalent tasks to all students.

Aerial performance was evaluated twice. The intermediate aerial check measured proficiency at a point half way through air-borne training. The final aerial check was given shortly before graduation. Both checks evaluated performance on typical aerial radar navigation and bombing missions. Aerial check items, like supersonic trainer items, measured conformity to accepted methods of navigating and bombing as well as the quality of the student's results. Aerial evaluation presented difficulty in that there was no permanent objective record of the student's performance and much depended on the ability and conscientiousness of the individual examiner. Specific check routes were established at each school.

Statistical findings as well as observation indicate that standardized administration of the battery of performance checks was only partially achieved. The practical demands of the radar training program did not permit study of check-recheck reliability. The one study made of a small number of cases on the final supersonic check suggested that that check had only moderate reliability.

A study of the relation between check scores and instructor ratings on aerial missions resulted in substantial intercorrelations. The degree to which knowledge of check score influenced instructor judgment is unknown.

CHAPTER SEVEN

The Measurement of Performance[1]

A by-product of the radar project's experience in constructing performance checks and applying them to the diverse aspects of radar observer training was the development of certain tentative generalizations about the measurement of performance. While there was no opportunity under wartime conditions to verify such generalizations experimentally, they are presented in this chapter because of their potential value to similar measurement efforts in the future.

The discussion will be concerned first with questions related to the reliability of performance measurement. As an aid to this analysis of reliability, however, a prior discussion is undertaken of the nature or structure of performance check items. Here is outlined a scheme in terms of which subsequent discussions of reliability are oriented. Finally, the topic of validity is treated briefly. Two types of validity are distinguished and the various types of criteria against which validation may be carried out are considered.

THE STRUCTURE OF PERFORMANCE CHECK ITEMS

Background for an Analysis of Sources of Unreliability

As already pointed out in chapter 6, practical considerations made it impossible to conduct statistical studies of the reliability of performance checks. It was generally believed, however, that their reliability was not high. In order to explore systematically the possibilities for improving reliability, it was concluded that some type of analytical framework was necessary. The break-down of performance measurement which constitutes this framework is presented in this section of the chapter.

General Structure of Items

Broadly, any performance check item may be thought of in terms of two components—the student performance and the examiner evaluation of that performance. These components are easily discerned in a performance check item in which the examiner merely decides whether the student behavior which he observes fits the description of

[1] Written by Cpl. Douglas W. Bray.

the correct behavior as given in the item. Such items are common in the bench set performance checks developed for the radar observer program. For example, one item asks whether or not the st dent turns a certain tuning knob counterclockwise. The student performance in this case is the overt act of turning the knob; the examiner evaluation of the performance consists merely of a decision as to whether the correct movement took place. However, the two components are not so obvious in other performance check items. For example, an item from the supersonic trainer performance check for radar students asks whether an inked track line drawn by the trainer lies within 5 degrees of where it should lie. Here student behavior is not observed directly as it was in the previous example, but is reflected through a mechanical device. Likewise, the examiner's evaluation is not a simple decision but a precise measurement in reference to an external criterion.

Student Performance

A review of performance check items reveals that the student performance to be evaluated may be either observed directly or inferred from evidence gained through mechanical devices. These two sources of performance data will be discussed separately below.

Performance directly observed.—Performance may be directly observed in the student's behavior (action or speech) or in his productions (log entries, map plots, etc.). Directly observed student action is illustrated by an item from the aerial performance check for radar students which asks whether the student turns the receiver gain control co nterclockwise. Directly observed student speech is taken as evid ce of performance in an item from the same check which asks whether the student calls absolute altitude to the bombardier over the interphone. Student productions are used as performance data in still another item which asks the time elapsing between certain entries in the student's log and chart.

Performance evidenced through mechanical devices.—Performance, when it is not observed directly, may be evidenced either through the equipment used by the student, such as the radar set, or by special recording devices such as the glass map with marking pen attachment on the supersonic radar trainer. The use of operating equipment—in this case the radar set—as a source of performance data is illustrated by the aerial check item which calls for evaluation of the quality of the scope picture obtained by the student. Here the student's performance is evidenced not by his behavior in turning the tuning controls but by the effect of his behavior on the radar scope presentation. In another item, a recording device—the marking pen on the supersonic trainer—provides performance data. In this case an inked track line is the record evaluated because it represents the results of the student's navigational work.

In summary, the following are the main sources of performance data: The student's action or speech, the student's productions, the operating equipment, and recording devices. The student provides directly observed performance data, while the equipment and recording devices provide indirect performance data.

Evaluation of Student Performance

The second component of the performance check item is the examiner evaluation of the student performance. The function of the examiner, broadly defined, is to determine whether the student performance satisfies, within tolerance limits, the standards for acceptable performance set forth in the check item. A major difference between items with respect to their evaluation, is between those which call for evaluation with reference to a verbal description of satisfactory performance, and those which specify that evaluation be made with reference to certain external criteria. For example, a radar aerial check item measuring systematic habits of log keeping instructs the examiner to give credit if the student makes entries in specified columns at a given time. The evaluation is made merely by comparing the nature of the student's log entries with those called for in the verbal description within the item. Another item from the radar aerial check, on the other hand, asks whether the student has kept the aircraft within 5 miles of the briefed course. To evaluate this item the examiner must first determine the exact course of the aircraft from the scope picture (the student performance) and then relate this to an external criterion, the briefed course. Since wide divergencies exist between evaluation with reference to verbal description and evaluation with reference to external criteria, the two will be discussed separately.

Evaluation in reference to verbal descriptions.—When evaluation of student performance is made in reference to a verbal description, the evaluation consists of a decision by the examiner who determines whether the performance fits the description of the correct performance. The conditions under which such decisions are made may range from those so unambiguous that almost all examiners would evaluate a given performance in the same way, to those so difficult to interpret that any agreement between examiners would most probably be due to chance. At one extreme is an item which asks whether the student turned a certain control in a specified direction. It can certainly be expected that almost all examiners would score this item in the same way. However, an item which asks whether the student secured the best possible definition of scope returns within the limitations of a particular radar set involves a much more difficult decision, and it is probable that examiners would frequently disagree.

Evaluation in reference to external criteria.—It has been indicated that the examiner's evaluation of performance is sometimes related not to a verbal description but to external criteria. A consideration of items of the latter type shows that such criterion data are provided primarily from two sources: the equipment, and recording devices. For example, in making a bomb run with the radar equipment, a "multiple" drift correction is made at the start of the run. If this correction is made perfectly, or if subsequent corrections within the tolerance limits are made soon enough, the aircraft will pass directly over the target. The target constitutes the external criterion with reference to which the accuracy of the course correction is judged. In an aerial performance check item, the radar set serves as the source of information as to the location of the target. In a supersonic check, evidence as to the location of the target comes from the glass map over which the recording pen traces a line representing the aircraft's course. Another illustration is found in the item requiring that the aircraft must be flown within a specified distance of a prearranged route. This route serves as a criterion to which the student's actual route is compared. In the air the route is found by reference to returns on the scope of the radar set. On the supersonic trainer it may be easily determined by laying a straightedge on the glass map between the starting point and the destination.

Examiner functions.—From both the discussion of the sources of evidence regarding student performance and the discussion of evaluation in reference to external criteria, it is clear that the examiner is called upon to perform a variety of operations. The end result of these operations is to put either the data reflecting student performance or that representing the criterion in such form that comparison between them is possible. These operations or examiner functions are of three kinds: participation, computation, and measurement.

Some performance check items require the examiner to use the same equipment and data as the student and to perform the same task as the student in order to provide a standard against which the student's results can be evaluated. In this discussion such examiner activity will be called "participation." For example, a method of evaluating the accuracy of the student's computation of the "multiple" drift correction in the radar aerial performance check is to have the examiner determine the necessary correction by observing the scope simultaneously with the student. If the heading correction which the student interphones to the pilot is within 5 degrees of the heading correction found necessary by the examiner, credit is given on the item. A complete analysis of this item would be as follows: the evidence of student performance is taken directly from his behavior, i. e., his speech when interphoning the drift correction to the pilot; the evaluation by the examiner does not involve an external criterion, and the examiner function is participation.

In the supersonic check item, discussed earlier, which evaluated the accuracy of the multiple drift correction in terms of whether the aircraft passed over the target, an additional task of the examiner was to note heading changes following the initial correction. The algebraic sum of the right and left corrections was determined and evaluated in terms of a tolerance given in the item. Such activity will be called "computation" in the following discussion.

Other performance is evaluated by the use of various measuring devices. In evaluating the accuracy of a student's computation of course correction on a supersonic trainer check, for example, the examiner uses a special protractor. With this he can measure the angular difference between the direction of the inked track and the true direction to the destination. This examiner function will be called "measurement." A complete analysis of this item would be as follows: student performance is evidenced indirectly through a recording device, the inked track line; evaluation is in terms of an external criterion taken from a recording device, the etched glass map on which the inked line is marked; the examiner function is measurement.

When the radar equipment provides the external criterion and measurement is called for, the metrical devices intrinsic to the equipment are often used. Thus, in an aerial radar performance check, distance off course is determined by using the range marks on the radar scope. These marks constitute a convenient distance scale.

It will be found that in many instances participation includes computation and measurement. Many items in the radar observer checks call for both computation and measurement. In some items computation is difficult to distinguish from measurement. The three functions, however, are generally clearly discriminable.

Discrete and Interdependent Items

Although any performance check may be analyzed in the foregoing manner, there are other considerations which may considerably alter the form of the check. One important factor is that the items constituting the performance check may be either discrete or interdependent. If a check is composed of discrete items, each item places a separate task before the student. His performance on an item will not be dependent upon his performance on preceding items. In a performance check of flying ability, for example, the student may be first checked on an Immelman turn, then on a loop, and so on. In checks which are continuous or running evaluations of performance in progress, however, the items are interdependent. A poor performance on one item may cause poor performance on another for which it is, in part, a preparation. For example, in the aerial check for radar observers, the student's computation of wind affects the accuracy of his subsequent course correction.

All of the performance checks developed to measure radar observer proficiency were of the interdependent type. It would have been theoretically possible, of course, to construct checks made up of discrete items. On the supersonic trainer, for example, a known wind, air speed, and heading might have been set into the equipment and the student instructed to determine a "multiple" drift correction on a particular target. The trainer could then have been placed at another point and the student required to take a fix. Here again a separate evaluation could be made. There were several reasons, however, for the use in the radar program of checks consisting of interdependent items. In the first place, checks of performance in progress were indicated because they were more feasible to administer. The manipulation of a bomber in accordance with the standardized conditions of many separate tasks presented impressive practical difficulties. Second, such checks had the advantage of applicability to the normal training program since check missions did not have to differ greatly from usual instructional missions. Furthermore, interdependent items were much less expensive in time since one aspect of performance followed another without interruption of performance or change in the testing conditions. In addition, checks consisting of interdependent items were preferred by training authorities who were strongly inclined towards evaluating total performance in progress. The performance checks developed by the Navigator and the Bombardier Projects had created a precedent for interdependent items. Finally, such an item type was suggested by the nature of the radar observer's task, the components of which are themselves interdependent. In the air, the student's determination of such things as wind, course correction, and fixes are all complexly intertwined. It is quite possible that a check of the total performance evaluates the integration of various skills and so provides a better estimate of complete proficiency than would a separate evaluation of each skill.

Although use of interdependent items appeared necessary in the radar program, there are, of course, many situations in which discrete items are indicated. In pilot training, for example, such items were more in keeping with the nature of the task, were easy to administer in the normal training program, and required little, if any, extra expenditure of time. Where possible, of course, discrete items should be used since they have a definite advantage over interdependent items in terms of reliability.

Direct and Indirect Measurement

Performance check items may also be direct or indirect in their evaluation of performance. A direct measurement is one in which performance itself is evaluated. An indirect measurement is one in which performance causally related to, but distinct from, the performance to be evaluated serves as the object of measurement. Illustration will make this difference clear: On one of the supersonic checks, the accuracy of the student's "multiple" drift correction is judged, in part, in terms of the amount of subsequent drift correction found necessary. Thus the performance of the initial drift correction is never checked directly at all. Later behavior provides the data. Another somewhat different example is the evaluation of accuracy of wind determination on an aerial mission. The only possible way of evaluating this performance directly would have been for the examiner to compute a wind independently. Since this was found to be impracticable, a direct check of wind accuracy was not used. However, because many other aspects of aerial performance, such as course corrections, ETA's, and precision turns, which were checked directly, depend on the accuracy of wind computation, a heavily weighted evaluation of wind computation was indirectly achieved. It will be clear that there is a close connection, on the one hand, between discrete performance check items and direct measurement and, on the other, between interdependent performance check items and indirect measurement. In any check consisting of interdependent items there is the possibility of indirect evaluation—in fact, it is present intrinsically even if each item task is also evaluated directly. Checks consisting of discrete items naturally rest upon direct evaluation.

RELIABILITY OF PERFORMANCE CHECK SCORES

A variety of factors influence the reliability of scores from performance measures. In the following discussion of these factors they are separated into three broad categories: those which determine the reliability with which the student's performance reflects his proficiency, those which determine the reliability of the performance data themselves, and those which determine the reliability with which student performance is evaluated. Discussion of the latter two groups of influences follows the analysis made in the preceding section.

Reliability With Which Performance Reflects Proficiency

Since a performance check is a device for measuring student proficiency, it is essential that proficiency be truly represented by the behavior measured. The sources of attenuation of performance as an index of proficiency may be discussed under three general headings, intrastudent variability, variations in equipment, and variations in testing conditions.

Closely related to the variability of performance as an indicator of proficiency are those variations which take place from time to time within the student. If the student is not motivated to do his best work on the performance check or if conditions such as fatigue and illness interfere with his performance, the score he earns will not be a good estimate of his proficiency. Poor student motivation may at times be due to the attitude of the examiner. Some examiners inspire a desire to perform well while others make the student feel that the check is just one more bothersome requirement.

In the evaluation of any performance involving the use of equipment, it is clearly necessary that the equipment upon which the task is to be performed must operate consistently from one student to the next. In radar, for example, a student could not be expected to achieve good results if the set were calibrated improperly. It should be noted, however, that the importance of standard equipment is related to the kind of behavior to be checked. Check items evaluating the precision of results are highly sensitive to even minor variations of the equipment. Items which ask only the order in which the student performs certain operations, on the other hand, are relatively immune to any but gross equipment difficulties. In the interdependent items found in checks of performance in progress, however, even such procedure items may be affected by equipment variation. For example, one of the radar aerial check items asks whether the student recorded an estimated time of arrival previous to a specified point. Poorly operating equipment might so handicap a student that he would have no ETA to record within the time allowed.

The other major interference with reliable performance comes from administering the check under varying testing conditions. Contributing to variation in testing conditions are such things as faulty examiner behavior, differences in the objective task, and inequalities in physical conditions. The examiner may be responsible for poor performance by failing to acquaint the student fully with the performance expected of him and the content and conditions of the task. Again, the examiner may be at fault in changing the objective difficulty of the task. On the radar bench set, for example, the examiner is required to preset the controls in prescribed positions before administering the check. Failure to do this makes the check of a different level of difficulty than when conditions are established as prescribed. The difficulty of the task may vary, also, for other reasons. In aerial radar performance, for instance, it is known that terrain determines to some extent the difficulty of navigational and bombing problems. For this reason, specific check mission routes were insisted upon. Finally, there is the variety of physical conditions under which a student may be required to work. To again illustrate from the aerial radar performance check, variations in weather, the turbulence of the air, the adequacy and cooperation of the crew, the altitude and the temperature, all may have a considerable effect upon student performance. It is obvious that every effort must be made to make the objective task and the environment in which it is performed comparable for all students and to assure equivalent instruction of the student by all examiners.

Reliability of Student Performance Data

As pointed out in discussing the structure of performance check items, student performance may be observed directly in the student's behavior and productions or indirectly from the equipment of the task or from recording devices. Performance data are most subject to error when they are observed indirectly. As shown above, variation in the equipment and recording devices will cause the student to perform in a manner which does not truly represent his proficiency. However, even though the student's performance satisfactorily reflects his skill, it is still possible for the equipment or recording devices to present an inaccurate picture of his accomplishment. In administering the supersonic trainer performance check, for example, the glass map on which the student's performance was recorded was occasionally found to be out of alignment with the artificial terrain. This resulted in a discrepancy between the actual position of the aircraft, as represented to the student on the scope, and the record of that position on the glass map. Under these conditions the student might have actually directed the aircraft exactly over the center of a town while the record would show that he had missed the town completely.

Another somewhat different possibility for error in the data representing student performance, has to do with the nature of the equipment or recording device being used. Some instruments are such that accurate measurement or observation is possible, while others require interpolation, estimation, etc.

Reliability of the Evaluation of Student Performance

In the break-down of the structure of performance check items, it was seen that in some cases, evaluation of student performance took place with reference to a verbal description within the item while in others it was carried out with reference to an external criterion. In the earlier discussion, it was also noted that evaluations made in reference to verbal descriptions may vary from those so simple that all examiners may be expected to agree to those so difficult that any agreement would probably be due to chance. A review of various performance check items indicates that an important factor in the difficulty of such evaluations is the degree of completeness of the description of the correct performance contained in the check item. This, in turn, is related to the difficulty in objectifying the description. For example, an item which asks merely whether or not the student turns a knob counterclockwise contains a precisely formulated and complete statement of the expected performance. On the other hand, an item which asks whether a student did something accurately, as well as possible, or satisfactorily, leaves the decision as to the correct performance largely to the experience and subjective standards of the examiner. Evaluations in reference to verbal descriptions can be made more reliable, in part, by defining acceptable performance as precisely and completely as possible, in part, by selecting as examiners highly experienced personnel, and, in part, by training examiners thoroughly in an attempt to standardize their concept of the correct performance for the appropriate items.

When evaluation takes place in reference to an external criterion, the question of reliability hinges in part upon the nature of the criterion utilized. If, as in the case of a target or a route on the glass map of the supersonic trainer, the criterion may be precisely identified, then measurement in relation to it may be expected to be reliable. If, on the other hand, the criterion can not be precisely established, some unreliability of measurement must result.

From the point of view of the examiner functions of participation, computation, and measurement, as described earlier, reliability is influenced in several ways. Participation appears most likely to result in unreliability since it assumes a high degree of examiner care and skill in the performance being checked. The reliability of items relying on computation may be improved by constructing such items so that the numerical operations involved are the simplest possible. The function of measurement may be made more dependable by providing metrical devices to facilitate the examiner's task.

The Measurement of Reliability

Measurement of the reliability of the performance check scores presents several difficulties of definition and methodology. In this discussion, the coefficient of reliability will be understood to represent the extent of agreement between scores which students receive on a performance check taken under routine administrative conditions and scores which they would have received had they taken the check at a different time with examiner, equipment, recording device, and conditions of administration, varied as they are under routine administrative conditions. An important provision in this definition is that the reliability of the performance check score is not independent either of intrastudent variability or of variations in examiner, equipment, etc. In the routine use of performance checks, such independence is not found. Conventional methods of determining reliability are discussed below in terms of this definition.

There are two principal internal methods of computing reliability: the part-part or split-half method, and the odd-even method. The part-part method correlates the score earned on one half of the check with that earned on the other half, while the odd-even method correlates the score earned on odd items with that earned on even items.

Neither of these methods is suitable for determining the reliability of a performance check since they rely on a single administration of the measure. Within a single administration, several important sources of variation between administrations are held relatively constant. These are intrastudent variability in performance, variation between examiners, variation in testing conditions, variation in testing equipment, etc.

An alternate method of determining performance check reliability is by the test-retest method. This may be accomplished either by a second administration of the same check or by using an alternate form of the measure for the recheck. When either of these methods is used, learning and memory will act to increase scores on the recheck over the check. This, of course, would not attenuate the coefficient of reliability if it were not for the fact that amount of improvement will vary from student to student. A point of note in connection with use of the test-retest method is that no attempt should be made to standardize student motivation and training, examiner, equipment, recording devices, or conditions of administration more completely than would be the case in the normal testing routine.

The construction of an equivalent test form presents numerous practical difficulties in the case of checks of performance in progress consisting of interdependent items. In such instances, all critical, measurable aspects of the task are presumably included in the original form. While it may not be possible to create a second check of alternate equivalent tasks, it is in some instances possible to vary the content of the task. In aerial radar performance checks, for example, changing the route of the mission would result in a recheck similar to, but not equivalent to, the original check.

The coefficient of reliability described above expresses the reliability of the performance check under routine administration. It is occasionally useful to discover just how reliable the check is when administered by different examiners but with all other conditions held constant. To distinguish this concept of reliability from the more general one, it will be referred to as the coefficient of objectivity. The term objectivity was chosen because differences between evaluations by different examiners are presumably due to subjective differences in their use of the performance check.

To determine the coefficient of objectivity of a performance check, it is necessary to eliminate extra-examiner variability. In many situations the only practicable way to do this is to have two examiners use the check independently to evaluate the same performance.[2] A correlation between these two sets of figures will produce the desired coefficient.

---

[2] This technique was used by Psychological Research Project (Pilot) and was called "observer reliability."

In the case where the two sets of examiners agree perfectly, the coefficient of objectivity will be 1.00 and a lower coefficient of reliability will mean that all unreliability comes from variation in testing conditions, student performance, equipment, recording devices, etc. In the case where the coefficients of objectivity and reliability are equal, it will mean that such variables do not influence check scores and that all unreliability is due to lack of examiner agreement (lack of objectivity of the measure). Such information reveals whether effort expended to produce standardized performance checking should be directed toward improving the instrument and its administration or toward eliminating extra-check variability. When it is found that the coefficient of objectivity is too low to be acceptable, tetrachoric correlations for the scoring of each item by the two sets of examiners will point out those items which are most lacking in objectivity.

VALIDITY OF THE PERFORMANCE CHECK

In reference to proficiency measures, two different concepts of validity may be distinguished. According to the usual concept, a measure is valid if scores from it predict future attainment. For convenience of discussion this will be referred to as predictive validity. In radar observer basic training, for example, the predictive validity of performance checks would be determined from the relationship between performance check scores and one of several other criteria of proficiency. These criteria might represent proficiency in combat or in advanced training, or they might even be other measures from basic training. The second concept of validity holds a measure to be valid if it furnishes a comprehensive test of achievement for a given area of instructional material. Validity of this type will be referred to as curriculum validity. The curriculum validity of performance checks would be judged in terms of the extent to which they tested the skills taught in the sections of the training course they were intended to cover.

The degree of relationship between predictive validity and curriculum validity may, of course, be of any magnitude. If, for example, a course of instruction is poor preparation for future proficiency, measures possessing high curriculum validity may be poor predictors of later performance. On the other hand, if proficiency in training is highly related to later proficiency, measures having high curriculum validity will have high predictive validity. The performance checks developed for radar observer training were pointed toward curriculum validity, since their purpose was to provide a complete evaluation of student proficiency during the basic training course. It would have been highly desirable, of course, to have been able to determine their predictive validity.

Curriculum Validity

The initial work on a proficiency measure must be pointed toward curriculum validity. Even if suitable criteria of predictive validity are obtainable, routine use of the measures can seldom be postponed until complete validation studies can be made. The use of curriculum validity is, however, not wholly based on negative reasoning. Curriculum validity is usually achieved in cooperation with training authorities. From the point of view of the practical training situation, there can be little quarrel with the validity of a performance check if it measures those abilities in which the student must be proficient before he is considered ready to progress to a later stage of training. The most effective use of expert opinion in establishing curriculum validity is achieved by obtaining expert criticism of proficieny measures developed by psychologically trained personnel. This method is facilitated if such psychological personnel have achieved moderate proficiency in the performance to be evaluated.

Predictive Validity

Problems associated with the question of the predictive validity of a proficiency measure relate primarily to the criteria against which validation should be carried out. These criteria fall into several categories: measures of final proficiency, measures of proficiency in later training, and alternate measures of proficiency within the same stage of training.

Final proficiency as a criterion.—Had it been practicable, it would have been of considerable value to determine for proficiency measures their predictive validity against a combat criterion. However, the difficulties of collecting precise proficiency data on radar observers engaged in aerial combat were great and such efforts as were made bore little fruit. The value of criteria which might have been obtained, such as bombing accuracy, would have been seriously attenuated by such variable influences as type aircraft, effectiveness of crew, and equipment.

Proficiency in later training as a criterion.—Indices of subsequent proficiency may also be obtained from measures of ability in advanced stages of training. In the preparation for radar observers for combat a second period of training followed graduation from the basic training schools. This training took place after the combat crew had been assembled and was intended to organize the several members of the crew into an efficient team. Had a suitable measure of proficiency at this level been available, it would have provided a useful criterion for the validation of proficiency measures in basic training. However, experience with such data show them to have serious weaknesses. The difficulties of securing reliable measurement of individual proficiency in crew training are at least as great as they are at the individual training level. To the variable influences of equipment and of the physical conditions of aerial performance is added the systematic error of measuring individual performance in a team activity. The performance of the radar observer is a function not only of his own proficiency but also of that of the other crew members.

Unless there is assurance that proficiency measures from a subsequent stage of training are free of such faults, they are of little practical value as criteria. The development of such assurance frequently necessitates ambitious research at the advanced training level. Moreover, there is almost invariably a long waiting period before such criterion data mature. In the practical situation it may be impossible to postpone the administrative use of proficiency measures while awaiting the results of future measurement of proficiency.

MEASURES OF PROFICIENCY WITHIN THE SAME STAGE OF TRAINING AS CRITERIA

1. *Scores made by good and poor students as a criterion.*—A measure of the predictive validity of the performance check at the training level may be obtained by administering the check to students of different proficiency.[8] The students, in such a case, either may be determined by some independent means to be of different proficiency or may be assumed to differ in skill by reason of having undergone different amounts of training. In the former instance, the independent evaluation of proficiency will have been based on previous performance measures such as instructor ratings. The use of extremely good and extremely poor groups of students increases the likelihood of getting two groups that are actually different in proficiency.

Essentially similar to the use of check scores made by students of varying proficiency is the use of the graduation-elimination criterion for validating the performance check. It is important, of course, that the check score being validated not enter into the determination of graduation-elimination, since this, of itself, would produce a sizeable correlation. Where the grading system is not acceptably reliable, the use of graduation-elimination is of little value as a criterion. For practical usefulness, it is also necessary that a significant number of eliminations be made. In the radar program, for example, the need for radar observers did not permit the elimination of more than a few students, and gradution-elimination was therefore of little value as a criterion.

2. *End result of performance as a criterion.*—Since it is frequently impracticable to use measures of subsequent proficiency as criteria against which to determine validity, more readily available sources of criterion data should be considered. One such source is the measurement of the end result of the performance being measured. In radar

[8] Such a technique was used in Psychological Research Project (Pilot).

observer training the primary objective of the student's performance was to bomb a target. His circular error, the distance away from the target that the bomb actually hit, would appear to be a valuable criterion. However, to be of value, a criterion must not only be an obvious consequence of the performer's skill; it must also be reliable. The usefulness of circular error was expected to be limited because of the previously reported findings of the Bombardier Project concerning the influence of variables other than student ability on bombing accuracy. These variables included equipment, the aircraft, the weather, and other crew members. In addition, it must be remembered that the circular error resulting from any navigation leg-and-bomb run is only a single measure of proficiency. It is known that the larger the number of measurements obtained, the more reliable tends to be the evaluation of performance. The gathering of data for many bomb drops would presumably have offset variable effects but the number of drops needed was expected to greatly exceed the number possible in the regular training program. The use of the end results of performance as validation data is dependent upon the possibility of making such results meet standards of satisfactory reliability.[4]

3. *Intercorrelations of proficiency measures as a criterion.*—The validity of performance measures has occasionally been inferred from their intercorrelations. It is clear, however, that if none of the proficiency measures which are intercorrelated is known to have predictive validity, their intercorrelations, of themselves, can tell us nothing of predictive validity. Furthermore, even if one of the measures is of known predictive validity, its correlation with other measures may tell little of their validity because the other measures may correlate merely with that portion of the valid measure which is not predictive of the criterion.

Neither is this method particularly helpful in determining the curriculum validity of performance checks. Even if one check is known to be a complete and reliable measure of achievement in the course of training, the fact that another check of a specific section of the course does not correlate highly with it does not indicate conclusively that the latter check is invalid as a measure of the task it evaluates. It may well be that the more specific check is devoted to evaluating the learning of a task in which individual differences disappear through overlearning by the time the measure of final proficiency is taken. For example, a check of achievement on the bench set might be a satisfactory measure of achievement in that specific training situation and yet correlate poorly with the final aerial check because early differences in proficiency in elementary set operation are a minor factor in later radar navigation and bombing.

---

[4] For a report on the reliability of circular error in radar bombing and its correlation with performance checks, see ch. 9.

SUMMARY

Experience with the construction and use of performance checks in the radar observer-training program is summarized in this chapter. Suggestions are made regarding possible sources of unreliability of measurement, methods of determining reliability, and the evaluation of validity.

As an aid to the discussion of reliability, the performance check is analyzed from the point of view that it comprises two primary components: student performance, and the evaluation of that performance by the examiner. In a further analysis it is noted that student performance may be observed directly in student behavior and in student productions and indirectly through operating equipment and recording devices. Evaluation of student peformance by the examiner may be made either in reference to a verbal description of satisfactory performance or in reference to external criteria. In radar observer performance checks there are two sources of external criteria: the operating equipment and recording devices. In the course of his evaluation the examiner is called upon to perform a variety of operations, the end result of which is to put either the data reflecting student performance or that representing the criterion in such form that comparsion between them is possible. These operations or examiner functions are of three kinds: participation, computation, and measurement. In systematically exploring the performance check from the point of view of possible sources of unreliability, the above analysis is followed.

In a discussion of the measurement of performance check reliability, it is pointed out that the test-retest method provides the only entirely suitable results. Internal measures of reliability neglect important sources of variation between successive check administrations. As a rough guide to the sources of unreliability, a coefficient of objectivity is distinguished from the coefficient of reliability. The former coefficient describes the reliability of the check when administered by different examiners but with all other conditions held constant.

When this coefficient is low, it is taken to mean that the check is insufficiently objective for consistent application by two examiners. This suggests attention to the mechanics of the check rather than to increased standardization of testing conditions, equipment, or other variables.

Two concepts of validity are distinguished. A measure is said to have predictive validity if scores from it predict future attainment. It is said to have curriculum validity if judged by experts to test adequately the skills taught in that section of the training course it is intended to cover. Criteria against which predictive validity may be determined fall into three categories: measures of final or combat proficiency, measures of proficiency in later stages of training, and alternate measures of proficiency within the same stage of training.

CHAPTER EIGHT

Interrelations of Proficiency Measures[1]

In this chapter the interrelations of the printed tests and performance checks discussed in preceding chapters are examined. There are generally considered to be four basic radar observer skills, namely, radar navigation, radar bombing, set operation and scope interpretation. These skills are not discrete; there is considerable overlapping and interaction among them. For each skill except scope interpretation, printed tests and standardized performance checks have been developed. These tests and checks constituted a battery of 11 proficiency measures administered routinely at several radar observer schools.

As a basis for analyzing the interrelations of these measures, their intercorrelations were determined. The coefficients obtained will be interpreted from several points of view. First examined are the correlations between tests and performance checks which measure similar skills. The question is raised as to whether or not such correlations are high enough to warrant the substitution of administratively economical printed tests for administratively expensive performance checks.

Next considered are certain questions in connection with the correlation between pairs of performance checks and pairs of printed tests. Among these are the degree of relationships between measures of the same skill as compared to measures of different skills, and the magnitude of the intercorrelations yielded by different measures of the same basic skill. Where these correlations involve performance checks, they are compared to parallel correlations between instructor ratings of proficiency.

Correlations between measures of one skill with those of another are next examined. Evidence is sought as to the magnitude of relationship between proficiency in different skills and the question is raised as to whether certain skills are more highly correlated than others.

---

[1] Written by Sgt. Philip H. Kriedt with the assistance of S/Sgt. Roland E. Johnston and S/Sgt. Harold F. Kunsman.

Finally, where correlations are available for parallel measures used elsewhere in the AAF Aviation Psychology Program, those are introduced for comparative purposes.

THE VARIABLES AND THE SAMPLE

The correlation coefficients reported in this chapter are, with a few exceptions, based on the scores of 190 students who were graduated from Langley Field. The students were members of classes 45–19 through 45–33. All were rated bombardiers with no combat experience. Class 45–19 graduated on 12 May 1945; class 45–33, on 18 August 1945.

This sample was selected in order to best meet the following requirements:

1. A group of students was desired who had taken all of the important proficiency measures. This would make it possible to base all correlations on the same sample and thus render comparisons of correlation coefficients more meaningful. For most of the correlations, the N is 190 or very close to it. In a few instances, however, the N drops to as low as 160. Because of the emergency need to graduate as many radar observers as possible, students occasionally missed taking some of the proficiency measures. For those tests and checks considered most important, however, the N is over 185.
2. Students were desired who had taken the performance checks during a period when they were believed to be most reliably administered. In choosing this period it was assumed that scores given by relatively experienced examiners would be more reliable than scores given by the less experienced examiners who administered proficiency measures early in the training program.
3. Students were desired from a period during which the fewest changes were made in the forms of the proficiency measures. For the sample selected, each printed test was administered in only one form; however, of the six performance checks, four were given in two slightly different forms.

Variables

The five printed tests referred to in this chapter are described briefly below. The development of these tests is described in chapter 5.

Final Test I for AN/APS-15, AN/APS-15A, and AN/APQ-13, form B (P1a-B).—This test is divided into four sections. Section A is designed to measure scope interpretation and navigation skills through the technique of "flying" a simulated mission. Students are provided with an expendable Mercator map of the area in which the navigation and bombing is to take place, 10 scope photographs at various points on the route, and readings which would be obtained from various navigational instruments. With this information the student uses navigational tools to take the computations required at various points. Section B is designed to measure skill in using the E-6B computer to solve radar navigation problems. Section C is designed to measure skill in determining the position of the aircraft from simulated scope photographs and in plotting this position on a map. Section D is designed to measure skill in the navigational technique of wind determination by airplot. Throughout the chapter this test is referred to as Final Test I.

Final Test II for AN/APS-15, form B (P1b-B).—This test is divided into two sections. Section A contains items testing technical information about radar bombing and radar navigation procedures. Emphasis is placed on radar bombing. Section B contains items on technical aspects of tuning, use of the beacon, and other phases of set operation. This test was used with Final Test I as a comprehensive examination. Throughout this chapter it is referred to as Final Test II.

Radar bombing intermediate test for AN/APS-15, form B (P1e-B).—This test contains items on the theory and technique of bombing, including the use of equipment controls before and during the bomb run. Throughout this chapter it will be referred to as the radar bombing intermediate test.

Radar navigation intermediate test for AN/APS-15, form B (P1h-B).—This test is divided into three parts. Part I consists of items dealing with the operation of radar equipment and the techniques involved in navigating from radar returns. Part II consists of problems dealing with the use of the E-6B computer in basic and radar navigation. Part III contains problems in the determination of wind by air plot. Throughout the chapter this test will be referred to as the radar navigation intermediate test.

Set operation intermediate test for AN/APS-15, form A (P1k-A) and AN/APS-15A, form A (P1l-A).—This test is divided into three sections. Section A deals with the operation of AN/APS-15 equipment in searching for targets and beacons; with the location, function, and effects of malfunction of various components of the equipment; and with the operation of the set in bombing. Section B deals with the SCR-718A altimeter, a radar device used to determine absolute altitude, and with the adjustment of the AN/APS-15 set in terms of readings from the altimeter. Section C differs from sections A and B in that it covers material specific to the AN/APS-15A set. Throughout the chapter this test is referred to as the set operation intermediate test.

The development of the performance checks to be discussed in this chapter is described in chapter 6. Brief descriptions of these checks are given below.

Final bench set performance check for AN/APS-15, form B (P6a-B).—Each item in this check defines concisely a distinct step in the procedure used in the preoperational check, in starting and timing, in range and altitude calibration, and in the turn-off. The student is told by the examiner to go through a given procedure and call out each step as it is performed. If an item is omitted or performed incorrectly, the examiner corrects the student immediately. Throughout the chapter this check is referred to as the final bench set check.

Final supersonic trainer performance check for AN/APS-15, AN/APS-15A, and AN/APQ13, form A (P7a-A) and Langley form B (L P7a-B).—Two forms of this check were administered to different students in the sample studied. The forms are essentially the same; each is a combination of the intermediate checks for navigation and bombing described below. Where the scores available were for form B, such scores were transformed to scores comparable to those for form A (P7a-A). Throughout the chapter these checks are referred to as the final supersonic check.

Radar bombing supersonic intermediate check for AN/APS-15, AN/APS-15A, and AN/APQ-13, form A (P7c-A), and Langley form B (L P7c-B).—Two forms of this check are listed for the same reason as stated below. Scores from form B were transformed to scores comparable to those for form A. Form A contains items covering both direct and coordinated bombing, while Langley form B tests coordinated bombing only. Differences between the two forms are thought to be minor. Throughout this chapter these checks are referred to as the bombing supersonic intermediate check.

Radar navigation supersonic intermediate check for AN/APS-15, AN/APS-15A, and AN/APQ-13, form CX, and Langley form C (L P7c-C).—Both forms of this check are designed to measure skill in the navigation procedures required to conduct a simulated radar navigation mission successfully on the supersonic trainer. The scores on form C were transformed to scores comparable to those for form CX. Throughout this chapter these checks are referred to as the navigation supersonic intermediate check.

Final aerial performance check for AN/APS-15, AN/APS-15A and AN/APQ-13, form B (P8a-B).—This check is designed to evaluate aerial proficiency in radar navigation, radar bombing, and set operation. This check is administered on a 4-hour mission at the conclusion of a student's aerial training. Throughout this chapter it will be referred to as the final aerial check.

Intermediate aerial performance check for AN/APS-15, AN/APS-15A, and AN/APQ-13, form A (P8b-A).—This check is designed to measure a student's aerial proficiency in radar navigation, radar bombing, and set operation at approximately the midpoint of his aerial training. The navigation and bombing items on this check are the same as those in the Final Aerial Check except that coordinated bombing procedure is not checked and items dealing with trouble-shooting, starting, and tuning are included. Prior to 1 June 1945 this check was administered to only one student on a single flight during which that student was checked on three legs. On 1 June 1945 and thereafter the procedure was modified so that two students could be checked on a single flight. Scores from the more recent type of administration have been transformed to scores comparable to those for three legs. Throughout the chapter this check is referred to as the intermediate aerial check.

INTERCORRELATIONS OF TOTAL SCORES FOR 12 PROFICIENCY MEASURES

Table 8.1 presents a matrix of intercorrelations for 12 proficiency measures. Variables 1 through 6 are the five printed tests described above, plus a score obtained by totaling Final Tests I and II. Variables 7 through 12 are the performance checks. All of the coefficients were computed by the Pearson product-moment method.[2] The table contains essential data which are referred to and interpreted, for several different purposes, throughout the remainder of the chapter. It is inserted here to give a general background for the specific problems and interrelations which will be discussed.

[2] The majority of the statistical computations were made by T/Sgt. Hyman Schnierer, S/Sgt. William J. Woywood, Cpl. Owen R. Munger, Cpl. Irving Fudeman, and Cpl. James R. Holt.

TABLE 8.1.—*Intercorrelations, means, and standard deviations of total raw scores for 12 proficiency measures*

	1	2	3	4	5	6	7	8	9	10	11	12
1	......	......	......	[1] 0.39	[1] 0.41	[1] 0.28	0.11	[1] 0.31	[2] 0.17	[2] 0.15	[2] 0.17	[1] 0.22
2	......	......	[1] 0.25	[1] .28	[1] .42	.10	.09	[1] .25	.08	[2] .17	[2] .17	.13
3	......	[1] 0.25	......	[1] .30	[1] .28	[1] .42	.08	.13	[2] .16	.01	.00	[1] .22
4	[1] 0.39	[1] .28	[1] .30	......	[1] .36	[1] .30	.09	.06	.07	[2] .17	[1] .24	.03
5	[1] .41	[1] .42	[1] .28	[1] .36	......	[2] .11	.11	.05	.13	.13	−.01	.10
6	[1] .28	[1] .10	[1] .42	[1] .30	[2] .14	......	.05	.11	.09	.06	.13	[1] .34
7	.11	.09	.08	.09	.11	.05	......	.12	.11	.07	.10	.06
8	[1] .31	[1] .25	.13	.06	.05	.11	.12	......	[2] .18	.05	.11	[1] .32
9	[2] .17	.08	[2] .16	.07	.13	.09	.11	[2] .18	......	[2] .18	[1] .19	.07
10	[2] .15	[2] .17	.04	[2] .17	.13	.06	.07	.05	[2] .18	......	[2] .17	.13
11	[2] .17	[2] .17	.06	[1] .24	−.01	.13	.10	.11	[1] .19	[2] .17	......	.14
12	[1] .22	.13	[1] .22	.03	.10	[1] .34	.04	[1] .32	.07	.13	.11	......

	Mean raw score	SD	N
Printed tests:			
1. Final Tests I and II	129.26	16.61	190
2. Final Test I	59.03	11.56	190
3. Final Test II	69.33	9.21	190
4. Radar Bombing Intermediate Test	41.03	4.98	190
5. Radar Navigation Intermediate Test	42.29	7.55	190
6. Set Operation Intermediate Test	72.80	8.41	190
Performance Checks:			
7. Final Aerial Check	60.79	11.60	190
8. Intermediate Aerial Check	45.[illegible]8	8.55	189
9. Final Supersonic Check	43.37	5.70	190
10. Radar Bombing Intermediate Supersonic Check	31.07	4.69	190
11. Radar Navigation Intermediate Supersonic Check	22.75	5.31	190
12. Final Bench Set Check	122.78	9.51	173

[1] Significant at the 1-percent level of confidence.
[2] Significant at the 5-percent level of confidence.

Of the 61 coefficients, 19 are significantly different from zero at the 1-percent confidence level. An additional 11 are significant at the 5-percent level. One coefficient is negative but is not significant.

In interpreting table 8.1 two factors which have operated in unknown strength to reduce the magnitude of the obtained coefficients should be kept in mind. First, the 11 proficiency measures were administered at various times in the course, ranging from the third week to the tenth week. It is probable that students progressed at varying rates of learning during the periods between tests or checks. Members of the Radar Project enrolled in the course reported that their proficiency improved considerably after a single period of instruction from a good instructor, and improved slightly or not at all over long periods of mediocre instruction. Second, as pointed out in chapters 5, 6, and 7, the reliability of the variables correlated may in a number of cases be quite low. It was not always possible to train test or check administrators as well as was desired. Examiners as well as the students were often not well-motivated, nor were administrative conditions always well standardized. However, as already noted, it was impractical to compute meaningful reliability coefficients and, in the discussions to follow, it has been necessary, consequently, to disregard almost entirely attenuation from this source.

CORRELATIONS BETWEEN PRINTED TESTS AND PERFORMANCE CHECKS MEASURING SIMILAR SKILLS

A comparison of test scores with scores on performance checks which are intended to measure similar skills gives an indication of the extent to which verbal knowledge of basic skills is related to actual performance. If the two are closely related, a printed test which can be administered to a large group of students at one time may be substituted for a performance check which must be administered individually and with apparatus.

Some light may be shed upon the first question by the answer to a second. How do the correlations tests and checks compare in magnitude to correlations among tests and to those among checks? If the correlations among tests and those among checks are uniformly high, a relatively unambiguous answer to questions involving the relations of tests to checks would be possible. If, on the other hand, the correlations of either test with test or check with check are low, less clear interpretation will be possible.

A third question is concerned with the effect of the time of administration upon the correlations of tests with checks having similar content. Are the correlations between tests and checks having similar content and administered at the same time in the course higher than the correlations between equally similar tests and checks administered at varying times in the course?

Answers to these questions will be sought in the correlations of the total scores of tests and performance checks, in the correlations of test scores and performance check part scores, and in the correlations of more inclusive measures, the proficiency stanines.

Test Scores and Performance Check Total Scores

The correlations between scores on printed tests and scores on performance checks are presented in table 8.2.

TABLE 8.2.—*Relationships of printed test total scores to performance check total raw scores*[1]

	7	8	9	10	11	12
1. Final Test I and II	0.11	[2].31	[3].17	[3]0.15	[3]0.17	[2]0.22
2. Final Test I	.09	[2].25	.08	[3].17	[3].17	.13
3. Final Test II	.08	.13	[3].16	.04	.06	[2].22
4. Radar Bombing Intermediate Test	.09	.06	.07	[3].17	[2].24	.03
5. Radar Navigation Intermediate Test	.11	.05	.13	.13	−.01	.10
6. Set Operation Intermediate Test	.05	.11	.09	.06	.13	[2].34
7. Final Aerial Check						
8. Intermediate Aerial Check						
9. Final Supersonic Check						
10. Radar Bombing Intermediate Supersonic Check						
11. Radar Navigation Intermediate Supersonic Check						
12. Final Bench Set Check						

[1] Means, standard deviations, and N's for these variables appear in table 8.1.
[2] Significant at the 1-percent level of confidence.
[3] Significant at the 5-percent level of confidence.

Are the correlations between tests and checks having apparently similar content sufficiently high to justify discarding a check and employing a test exclusively? In partial answer to this question several of the relationships in table 8.2 may be noted.

Final Test I, a printed test of radar navigation skills, is related significantly, at the 5-percent confidence level or better, to three of the performance checks. Final Test I correlates significantly at the 1-percent level of confidence with the intermediate aerial check and at the 5-percent level of confidence with the two intermediate supersonic checks.

Final Test II, a printed test of knowledge of navigation, bombing, and especially set operation, has significant relationships with two of the three checks which are not significantly related to Final Test I. These are the final supersonic check and the final bench set check, which is essentially a performance check of set operation.

Final Test I and Final Test II in combination are related significantly to all performance checks except the final aerial check. In general, the combination of Final Test I and Final Test II does not cause a higher relationship with any particular performance check than is obtained with either Final Test I or Final Test II separately.

The set operation intermediate test correlates significantly only with the final bench set check. This is the only instance in which similarity of content appears to be very important in determining the relationship of a printed test to a performance check. In table 8.2

no higher relationship is found between two bombing measures or between two navigation measures than between a bombing measure and a navigation measure.

Since the highest relationship between a printed test and a performance check measuring similar skills is represented by the coefficient 0.31, between the set operation intermediate test and the final bench set check, it is clear that without further information it would not be feasible to discontinue the performance checks in favor of the more easily administered printed tests. As already indicated it is not possible to estimate the effect upon the correlations discussed, of the low reliability which may characterize the performance checks.

How do the correlations between tests and checks compare in magnitude to the correlations among tests and to those among checks? In table 8.1 significant correlations are found more frequently between pairs of printed tests than between printed tests and performance checks. Twelve, or 92 percent of the 13 intercorrelations among printed tests are significant at the 5-percent level or better, whereas 13, or 36 percent, of the 36 correlations between printed tests and checks are significant at the 5-percent level. Significant correlations are found with equal frequency between printed tests and performance checks and between pairs of performance checks. In Table 8.1, 13 or 36 percent, of the 36 correlations between printed tests and checks are significant, as compared to 5 significant coefficients, or 33 percent, among the 15 correlations between checks.

Are the correlations between tests and checks having similar content and administered at approximately the same period in the course higher than the correlations between tests and checks administered at varying times? There is no indication in table 8.2 that this is the case. The combined score for Final Test I and Final Test II, given in the tenth week, correlates more highly with the intermediate aerial check given in the sixth week than with the final aerial check given the ninth or tenth week, and just about as highly with bombing supersonic intermediate check (fifth week) or the navigation supersonic intermediate check (third week) as with the final supersonic check (seventh week).

Test Scores and Performance Check Part Scores

The final aerial check and final supersonic check both contain items on radar navigation and radar bombing. The inclusion of two skills in a single performance check might attenuate any correlation of that check with printed tests designed to measure only one of the two skills. Therefore, the navigation items and the bombing items of the two checks were separately scored and the part scores were correlated with printed tests considered to be homogeneous tests of navigation or of bombing. In the case of the final supersonic checks, the navigation items and the bombing items of each of the two forms were correlated with each of the printed tests, and, using the Fisher z-function transformation, the coefficient between form A of the check and a given printed test was averaged with that between form B and the test. The correlation coefficients and other relevant data are found in table 8.3.

TABLE 8.3.—*Relationships of printed test total raw scores to performance check part scores*[1]

	Final aerial check		Final supersonic check		Mean raw score	SD	N
	Navigation items	Bombing items	Navigation items[2]	Bombing items[2]			
Final Test I	0.04	0.08	0.06	0.10	60.57	11.35	162
Radar Navigation Intermediate Test	.06	.10	−.07	.13	42.43	7.77	162
Radar Bombing Intermediate Test	.07	.05	.02	.11	42.02	6.15	162

	Final aerial check: Navigation items	Final aerial check: Bombing items	Form A	Form B	Form A	Form B
Mean raw score	35.26	22.00	23.21	11.47	21.10	19.47
SD	8.00	6.07	2.68	2.96	2.99	3.33
N	162	162	99	82	99	82

[1] None of these coefficients is significant at either the 1-percent or 5-percent level of confidence.
[2] Coefficients in these columns are averages of coefficients for form A and form B of the final supersonic check. The Fisher Z correlation function was employed in computing these averages.

None of the coefficients in table 8.3 is significant at the 5-percent level of confidence. When bombing or navigation test scores are correlated with similar bombing or navigation part scores for the final aerial check and the final supersonic check, the coefficients obtained are somewhat lower than those obtained with the total scores for these checks. The latter coefficients appear in table 8.2. This decrement in relationship may be due to the lower reliability associated with the smaller number of items in the part scores as compared with the total scores.

Proficiency Stanines

Three composite proficiency measures were computed, one based upon printed classroom tests, one upon trainer performance checks, and the third upon aerial performance checks. Each has been converted to stanine form. The classroom stanine is based on Final Test I and Final Test II, which together are weighted three, and Radar Navigation Intermediate Test, which is weighted one, the Radar Bombing Intermediate Test, weighted one, and the Set Operation Intermediate Test, weighted one. The trainer stanine is based on the Final Supersonic Check, weighted three, the Navigation Supersonic Intermediate Check, weighted one, the Bombing Supersonic Intermediate Check, weighted one, and the Final Bench Set Check, weighted one. The aerial stanine is based on the Final Aerial Check weighted two, and the Intermediate Aerial Check weighted one. In computing each of the three stanines, the individual scores were first converted to standard scores, then weighted and averaged.

The intercorrelations of these three proficiency stanines are presented in table 8.4. The correlations of the classroom stanine with the

TABLE 8.4.—*Intercorrelations of proficiency stanines*

	1	2	3	Mean	SD	N
1. Classroom stanine	....	[1] 0.26	[1] 0.20	4.8	1.38	175
2. Trainer stanine	[1] 0.26	....	.11	5.15	1.26	175
3. Aerial stanine	[1] 0.20	.11	....	5.02	1.44	175

[1] Significant at the 1 percent level of confidence.

trainer and aerial stanines are significant at the 1 percent level. The relationship of the classroom stanine to the trainer stanine may be somewhat higher than it is to the aerial stanine although the difference is not statistically significant. The correlation of the trainer stanine with the aerial stanine is not significant. In interpreting these observed relationships the possibility of low reliability for the trainer and aerial stanines must be kept in mind.

CORRELATIONS BETWEEN PAIRS OF TESTS AND BETWEEN PAIRS OF PERFORMANCE CHECKS MEASURING SIMILAR SKILLS

This section, like the one preceding it, is concerned with the interrelations of proficiency scores measuring similar skills. The analysis presented is based upon correlations between pairs of printed tests or pairs of performance checks. All total scores or part scores have been used which clearly measure one of three basic skills of the radar observer: radar navigation, radar bombing, or set operation. Correlations between measures of different skills have been included in the tables to allow comparison with correlations between measures of the same skill.

Comparisons will be made with the objective of answering specific questions such as the following. Do measures of similar skills correlate higher than measures of dissimilar functions? Do printed tests, trainer checks or aerial checks yield the highest interrelations? Of the three basic skills studied, which yields the highest intercorrelations? Does the time in the course at which a test or check was administered affect its correlation with other scores?

CORRELATIONS BETWEEN PERFORMANCE CHECKS MEASURING SIMILAR SKILLS

The correlations in table 8.5 suggest that so far as the supersonic trainer is concerned measures of similar functions do not correlate

TABLE 8.5.—*Correlations of radar navigation and radar bombing intermediate supersonic checks with navigation and bombing items of the final supersonic check*

	Final supersonic check				SD	N
	Naviga-tion items [2]	Bombing items [2]	Total score	Mean raw score		
Radar Navigation Intermediate Supersonic Check	0.01	[1] 0.18	[1] 0.19	22.75	5.31	100
Radar Bombing Intermediate Supersonic Check	[1] 0.19	[1] 0.16	[1] 0.18	34.97	4.69	100

	Form		Form		
	A	B	A	B	
Mean raw score	23.21	11.47	21.10	19.47	43.37
SD	2.68	2.06	2.99	3.33	5.70
N	99	82	99	82	100

[1] Significant at the 5 percent level of confidence.
[2] Coefficients in these columns are averages of coefficients for form A and form B of the Final Supersonic Check. The Fisher z transformation was employed in computing these averages.

more highly than measures of dissimilar functions. The Navigation Supersonic Intermediate Check correlates significantly with the bombing items and with the total score of the Final Supersonic Check; it does not correlate significantly with the navigation items. On the other hand, the Bombing Supersonic Intermediate Check is almost equally related to the navigation part score, the bombing part score, and the total score of the Final Supersonic Check.

Correlations between part and total scores of supersonic trainer performance checks and part scores of the Final Aerial Check are presented in table 8.6. Since only one coefficient in the table is signifi-

TABLE 8.6.—*Correlations of supersonic check raw scores with navigation and bombing items of the final aerial check*

Final aerial check	Radar naviga-tion inter-mediate super-sonic check	Final super-sonic check naviga-tion items [2]	Radar bomb-ing inter-mediate super-sonic check	Final super-sonic check bomb-ing items [2]	Mean raw score	SD	N
Navigation items	0.00	0.12	0.08	0.01	35.20	8.00	100
Bombing items	.01	−.04	[1] .23	.08	22.00	6.66	100

	Form			Form		
		A	B		A	B
Mean raw score	22.81	23.25	11.42	35.11	21.08	19.34
SD	5.12	2.65	2.80	4.40	2.69	3.52
N	100	99	55	100	95	55

[1] Significant at the 1-percent level of confidence.
[2] Coefficients in these columns are averages of coefficients for form A and form B of the Final Supersonic Check. The Fisher Z correlation function was employed in computing these averages.

cant at the 5-percent level no significant comparisons are possible. The correlation of 0.23 between the Bombing Supersonic Intermediate Check and bombing items in the Final Aerial Check may be taken as slight evidence that bombing performance on the supersonic trainer is more closely related to aerial bombing performance than navigation performance on the trainer is related to aerial navigation performance.

Table 8.7 summarizes the correlations of total scores on the supersonic trainer performance checks with total scores on the aerial performance checks. The correlation of 0.18 between the Final Supersonic Check and the Intermediate Aerial Check is the only correlation significant at the 5-percent level. These two performance checks were administered within a week of each other and are both combined measures of navigation and bombing.

TABLE 8.7.—*Intercorrelations of supersonic check raw scores with total raw scores on aerial checks*

	Final aerial check	Intermediate aerial check	Mean raw score	SD	N
Radar navigation intermediate supersonic check	0.10	0.11	22.75	5.31	100
Radar bombing intermediate supersonic check	.07	.05	34.97	4.69	100
Final supersonic check	.11	[1] .18	43.37	5.70	100
Mean raw score	60.79	45.38			
SD	11.60	8.55			
N	100	159			

[1] Significant at the 5-percent level of confidence.

Correlations Between Tests Measuring Similar Skills

Table 8.8 presents the intercorrelations of total scores of all printed tests. Of the 13 correlations in this table, 11 are significant at the 1-

TABLE 8.8.—*Intercorrelations, means, and standard deviations of total raw scores for printed proficiency tests*

	1	2	3	4	5	6
1				[1] 0.39	[1] 0.41	[1] 0.28
2			[1] 0.25	[1] .28	[1] .42	.10
3		[1] 0.25		[1] .30	[1] .28	[1] .42
4	[1] 0.39	[1] .28	[1] .30		[1] .36	[1] .30
5	[1] .41	[1] .42	[1] .28	[1] .36		[2] .14
6	[1] .28	.10	[1] .42	[1] .30	[2] .14	

	Mean raw score	SD	N
1. Final Tests I and II	129.26	16.61	100
2. Final Test I	59.93	11.56	100
3. Final Test II	69.29	9.34	100
4. Radar Bombing: Intermediate Test	41.03	4.08	100
5. Radar Navigation: Intermediate Test	42.29	7.65	100
6. Set Operation: Intermediate Test	72.80	8.41	.

[1] Significant at the 1-percent level of confidence. [2] Significant at the 5-percent level of confidence.

percent level and another at the 5-percent level of confidence. Final Test I correlates more highly with the other navigation measure, the Radar Navigation Intermediate Test, than with any other test. The latter test, however, correlates about as well with the dissimilar Radar Bombing Intermediate Test as it does with the similar Final Test I. The Set Operation Intermediate Test correlates more highly with Final Test II, which is heavily weighted with set operation items, than with any dissimilar test. On the whole, table 8.8 gives only slight evidence for the expected result, that printed tests measuring the same skills are more closely interrelated than printed tests measuring different skills.

Since Final Test II measures all three basic skills, part scores for this test have been correlated with total scores of tests considered to be relatively homogeneous measures. The correlations are presented in table 8.9. Differences in the obtained coefficients, while in the expected direction, are not of great magnitude.

TABLE 8.9.—*Correlations between intermediate tests and part scores of Final Test II*

	Navigation items	Bombing items	Navigation and bombing items	Set operation items	Mean raw score	SD	N
Radar Navigation Intermediate Test	[1] 0.39	[1] 0.31			42.29	7.65	190
Radar Bombing Intermediate Test	[1] .30	[1] .38			41.93	4.93	190
Set Operation Intermediate Test			[1] 0.38	[1] 0.31	72.80	8.41	190
Mean raw score	16.10	22.11	40.36	28.58			
SD	2.75	3.20	6.71	5.44			
N	190	190	190	190			

[1] Significant at the 1-percent level of confidence.

Certain additional data regarding the relationships among measures of similar skills are available which are not presented in tabular form. Correlations were computed between parts of the Radar Navigation Intermediate Tests and corresponding parts of Final Tests I and II. Scores of 177 Langley students on the airplot and E–6B sections of the Radar Navigation Intermediate Test correlated 0.30 with scores on the airplot and E–6B computer sections of Final Test I. Scores of 183 Langley students on Final Test II navigation items correlated 0.42 with their scores on similar items in the first part of the Radar Navigation Intermediate Test. While these results suggest that printed measures of navigational knowledge are more closely related than measures of navigational performance, it is equally likely that they only reflect differences in reliability.

In summary, measures of the same skill, as skill is here defined, seem very slightly if at all more closely related than measures of different skills. Printed tests furnish the highest intercorrelations whether the skills correlated are similar or dissimilar. Of the three basic skills, there is some evidence in the correlations between performance checks that the bombing measures are more closely related than either navigation or set operation measures, although there are few data involving the set operation measures. The Bombing Supersonic Intermediate Check is more closely related to the bombing items in the Final Supersonic Check than the Navigation Supersonic Intermediate Check is related to the navigation items in the Final Supersonic Check. Also, the Bombing Supersonic Intermediate Check is more closely related to the bombing items in the Final Aerial Check than the Navigation Supersonic Intermediate Check is related to navigation items in the Final Aerial Check. With regard to the influence of time of administration the evidence from tables 8.5 through 8.9 indicates that the period of the course during which the proficiency measures were administered has no consistent effect on the magnitude of their relationship.

CORRELATIONS BETWEEN RATINGS, PERFORMANCE CHECKS, AND FINAL TESTS

It was noted in the preceding section that the correlations between scores on the various performance checks were uniformly low. This finding suggested the desirability of correlating instructors' ratings of performance on these same checks on the hypothesis that such correlations might be higher. In formulating this hypothesis it was reasoned that ratings might not only take account of proficiency demonstrated during the check ride, but also allow for the variable conditions under which the student had to work. If ratings were independent of such variations while performance check scores were influenced by them, correlations of the former would be expected to reveal more faithfully the magnitude of any true relationship.

The ratings correlated were made by examiners at the completion of their administration of performance checks. Because the examiners were accustomed to thinking in terms of centile grades, the scale used was divided into 10 equal divisions, each supposedly containing 10 percent of the students. Instructions called for making the rating in relation to all students to whom the examiner had given performance checks and on the basis of how well he thought the student would have done under normal conditions.

Product-moment correlations were computed between ratings on performance checks and between such ratings and the final proficiency tests. Of the correlations computed, only two are significant at the 5-percent level. Ratings on the Final Aerial Check for 74 Langley students correlate 0.22 with ratings on the Final Supersonic Check. The correlation between scores on the corresponding performance checks for 190 Langley students is presented in table 8.1; the coefficient, 0.11, is not significant. The correlation between ratings on the Final Aerial Check and the combined scores on Final Tests I and II for 71 Langley students is 0.23. As indicated in table 5.1, the Final Aerial Check correlated 0.11 with the scores of 190 Langley students on the combined final proficiency tests. The latter coefficient is not significant. Additional correlations were computed between ratings for 75 Victorville students on the Intermediate Aerial Check and the Final Aerial Check (-0.17), between ratings for 66 Langley students on the Intermediate Aerial Check and the Final Supersonic Check (0.12); and between ratings on Final Aerial Check and scores on Final Tests I and II for 320 Victorville students (0.06). None of these reached the 5 percent level of significance.

CORRELATIONS BETWEEN NAVIGATION, SET OPERATION, AND BOMBING SKILLS

The curriculum assigned approximately equal weight to radar navigation, radar bombing, and radar set operation. In the classroom, 40 hours of lecture time are devoted to radar navigation, 40 hours to radar bombing, and 26 hours to set operation, although instruction in the latter is also included in both radar navigation and bombing. Set operation instruction on the trainer consists of 10 hours on the bench sets. Combined navigation and bombing instruction is given during 28 hours of supersonic training and 80 hours of aerial training. In almost all instances, however, the three categories overlap. Navigation and bombing skills may hardly be taught or applied apart from set operation skill. The set must be readjusted frequently to satisfy varying navigational and bombing requirements. Consequently, measures of bombing and navigation skill tend to be influenced by skill at set operation. Also, navigation and bombing overlap. Making heading corrections on the bombing run is, in a sense, merely precision navigation.

From such considerations as these, it is expected that test scores measuring these three different skills will be interrelated in some degree. The comparisons to be made will fall in five different categories: Part of a test against another part of the same test, part of a performance check against another part of the same check, test against test, performance check against performance check, and test against performance check. Within these categories it should be possible to make comparisons between correlations based upon pairs of measures with roughly equivalent reliabilities.

Part Against Part of the Same Test

Section A of Final Test II consists of 33 bombing and 22 navigation items. For the Langley sample of 190 students, the correlation is 0.50 between scores from those two groups of items. When odd and even scores from this same section are correlated, the uncorrected reliability coefficient is 0.48. This suggests that navigation items in this test are about as closely related to bombing items as they are to other navigation items or as bombing items are related to other bombing items.

Part Against Part of the Same Performance Check

For the same sample of 100 Langley students, Final Aerial Check navigation items correlate 0.50 with bombing items. In the Final Supersonic Check, navigation items correlate 0.35 with the bombing items. As was done above, these correlations may be compared to others in which one mixed bombing-navigation score is correlated with another. For the Final Aerial Check, for example, the correlation is 0.47 between legs 1, 3, and 5, and legs 2 and 4. No parallel correlation is available for the Final Supersonic Check.

The first and second legs of the Navigation Supersonic Intermediate Check correlate 0.36, and the first and second legs of the Bombing Supersonic Intermediate Check correlate 0.47.

It is suspected that halo effect operated to increase the magnitude of the correlations reported in the above paragraphs, since the scores correlated are assigned by the same instructor in a single testing period. The greater halo effect for the aerial check than for the more standardized supersonic check is in the expected direction.

Test Against Test

Correlations between radar navigation and radar bombing tests are presented in table 8.10. It is apparent that skill in navigation is positively correlated with skill in bombing, at least as measured by printed tests.

TABLE 8.10.—*Correlations between tests measuring radar navigation and those measuring radar bombing*

Radar navigation measures	Radar bombing measures		Mean raw score	SD	N
	Radar bombing intermediate test	Final test II bombing items			
Final Test I	[1] 0.28	....	59.93	11.56	100
Radar Navigation Intermediate Test	[1] .38	[1] 0.31	42.29	7.55	100
Final Test II, navigation items	[1] .30	[1] .50	16.10	2.75	100
Mean raw score	41.03	22.17	....	....	....
SD	4.98	3.19	....	....	....
N	100	100	....	....	....

[1] Significant at the 1-percent level of confidence.

Three correlations are available relating set operation skill to radar navigation and bombing skills. The Set Operation Intermediate Test correlates 0.10 with Final Test I, a navigation measure, and 0.14 with the Navigation Intermediate Test. The latter figure is significant at the 5-percent level. The Set Operation Intermediate Test and Radar Bombing Intermediate Test correlate 0.30, significant at the 1-percent level.

In summary, the results for printed tests indicate that bombing skill is significantly related to both navigation and set operation skills. Correlations between tests of those skills range from 0.28 to 0.50 or higher. The relationship of navigation to set operation is low and of questionable statistical significance.

Performance Check Against Performance Check

Table 8.11 presents the correlations between performance checks measuring radar navigation and those measuring radar bombing. Three of the seven correlations are significant at the 5-percent level. All are between navigation and bombing items from supersonic checks.

TABLE 8.11.—*Correlations between performance checks measuring radar navigation and those measuring radar bombing*

Radar navigation measures	Radar bombing measures				Mean raw score	SD	N
	Radar bombing, intermediate supersonic check	Final supersonic check, bombing items [2]	Final aerial check, bombing items				
Radar navigation intermediate supersonic check	[1] 0.17	[1] 0.18	.00	....	22.81	5.12	102
Final supersonic check	[1] .19	....	-.01	Form A	21.21	2.68	99
Navigation items [2]	....	....	....	Form B	11.47	2.96	82
Final aerial check, navigation items	.06	.04	....	....	35.20	8.00	102

		Form A	Form B	
Mean raw score	35.19	21.10	19.47	22.00
SD	4.31	2.09	3.33	6.07
N	102	99	82	102

[1] Significant at the 5-percent level of confidence.
[2] Coefficients are averages of those from form A and form B. The Fisher *z* transformation was employed in computing these averages.

From other tables two correlations are available between set operation skill as measured by performance checks and radar navigation and bombing. The Final Bench Set Check correlates 0.14 with the Navigation Supersonic Intermediate Check and 0.13 with the Bombing Intermediate Check, as indicated in table 8.1. Neither of these coefficients is significant at the 5-percent level.

Test Against Performance Check

In table 8.12, none of the correlations between navigation and set operation skills or between bombing and set operation are significant at the 5-percent level. Of the nine correlations between navigation and bombing measures one is significant at the 1-percent level, and another at the 5-percent level. The Radar Bombing Intermediate Test correlates slightly higher with the Navigation Supersonic Intermediate Check ($r=0.24$) than with the Bombing Supersonic Intermediate Check ($r.=.17$, table 8.1). The Radar Navigation Intermediate Test correlates slightly higher with the Bombing Supersonic Intermediate Check ($r=0.13$) than with the Navigation Supersonic Intermediate Check ($r=-0.01$, table 8.1). This comparison seems again to confirm the view that bombing skill overlaps considerably with navigation skill.

TABLE 8.12.—*Correlations between printed tests and performance checks measuring radar bombing, radar navigation, and set operation*

Printed test measures of radar navigation	Check measures of radar bombing		
	Radar bombing intermediate supersonic check	Final supersonic check bombing items [1]	Final aerial check bombing items
Final Test I	[2] 0.17	0.10	0.03
Radar Navigation Intermediate Test	.13	.13	.10

Printed test measures of radar bombing	Check measures of radar navigation		
	Radar navigation intermediate supersonic check	Final supersonic check navigation items [1]	Final aerial check navigation items
Radar bombing intermediate test	[3] 0.24	0.02	0.07

Set operation measures	Navigation measures			Bombing measures	
	Final Test I	Radar navigation intermediate test	Radar navigation intermediate supersonic check	Radar bombing intermediate test	Radar bombing intermediate supersonic check
Final bench set check	0.13	0.10	....	0.03	....
Set operation intermediate test	....	....	0.13	....	0.06

	Mean raw score	SD	N
Final test I	60.93	11.56	190
Radar navigation intermediate test	42.21	7.55	190
Radar bombing intermediate supersonic check	31.97	4.60	190
Final supersonic check, form A, bombing items	21.10	2.99	99
Final supersonic check, form B, bombing items	19.47	3.33	82
Final aerial check, bombing items	22.00	6.07	163
Radar bombing intermediate test	41.93	4.98	190
Radar navigation intermediate supersonic check	22.75	5.31	190
Final supersonic check form A, navigation items	23.21	2.68	99
Final supersonic check form B, navigation items	11.47	2.96	82
Final aerial check, navigation items	35.26	8.00	162
Final bench set check	122.78	9.51	173
Set operation intermediate test	72.80	8.41	190

[1] Coefficients in these columns are averages of those for form A and form B.
[2] Significant at the 5-percent level of confidence.
[3] Significant at the 1-percent level of confidence.

Insofar as the printed tests and performance checks are valid measures of the three basic skills examined in this section of the chapter, it appears that navigation and bombing skills are the most closely related, bombing and set operation skills are somewhat less closely related, and navigation and set operation skills are least closely related.

CORRELATIONS BETWEEN PART SCORES OF FINAL TEST I

Final Test I, a measure of radar navigation skills, consists of four sections. Section A is a simulated navigation and bombing mission involving the use of simulated scope photos, a radar Mercator chart of Germany, an E–6B computer, dividers, and Weems plotter. Sections B, C, and D are measures of the three primary skills involved in section A. Section B requires the student to plot aircraft position from 20 simulated scope photos. Section C consists of 20 E–6B computer problems. Section D consists of 20 airplot problems. It was expected that intercorrelations between sections B, C, and D, which were included in the test in order to provide added length and reliability, would be fairly high. It was expected also that correlations between section A and each of the other sections would be high, with the correlation between section A and the sum of sections B, C, and D being the highest of all. These correlations are presented in table 8.13. All are significant at the 1 percent level of confidence. The relationship of the mission to its component parts is somewhat lower than the relationship of the component parts to each other. The relationship of the mission to the sum of its component parts is, surprisingly, lower than the relationship of the component parts to each other.

TABLE 8.13.—*Intercorrelations of four sections of Final Test I (P1a–B) based on Langley Field classes 45–19 through 45–33*

	1	2	3	4	5
1. Section A (mission)	....	[1] 0.24	[1] 0.25	[1] 0.28	[1] 0.32
2. Section B (scope photo)	[1] 0.24	....	[1] .31	[1] .37	....
3. Section C (E–6B)	[1] .25	[1] .31	....	[1] .43	....
4. Section D (airplot)	[1] .28	[1] .37	[1] .43	....	....
5. Sections B, C, D (total)	[1] .32	....	....	....	....
Mean raw score	27.23	10.25	13.60	8.75	32.41
SD	6.10	3.54	3.31	2.95	7.09
N	190	189	189	189	189

[1] Significant at the 1-percent level of confidence.

There is some reason to believe, however, that those results are peculiar to the sample upon which intercorrelations in this chapter are based. Two similar studies were made previously, one of 236 students in classes 44–12 through 44–16 at Langley Field, and another of 309 students in classes 45–8 through 45–14 at Langley Field. These results are presented in tables 8.14 and 8.15.

Comparison of tables 8.13, 8.14, and 8.15 shows that the interrelationships between sections B, C, and D remain approximately the same in each of the three studies. In Tables 8.14 and 8.15, however, the simulated mission, section A, correlates as high with each of its component parts, B, C, and D, as do these component parts with each other. Moreover, in both these tables the highest correlation is that of section A with the sum of section B, C, and D. The results are very similar to what was originally anticipated.

TABLE 8.14.—*Intercorrelations of the four sections of Final Test I (P1a–B) based on Langley Field classes 44–12 through 44–16*

	1	2	3	4	5
1. Section A (mission)	....	[1] 0.49	[1] 0.45	[1] 0.47	[1] 0.59
2. Section B (scope photo)	[1] 0.49	....	[1] .40	[1] .37	....
3. Section C (E-6B)	[1] .45	[1] .40	....	[1] .49	....
4. Section D (airplot)	[1] .47	[1] .37	[1] .49	....	....
5. Sections B, C, and D (total)	[1] .59	....	....	....	....
Mean raw score	[2] 15.06	9.41	13.09	8.14	30.64
SD	4.06	3.29	3.67	2.41	7.39
N	236	236	236	236	236

[1] Significant at the 1-percent level of confidence.
[2] The mission section consisted of 25 items at this time.

TABLE 8.15.—*Intercorrelations of the four sections of Final Test I (P1a–B) based on Langley Field classes 45–8 through 45–11*

	1	2	3	4	5
1. Section A (mission)	....	[1] 0.45	[1] 0.51	[1] 0.44	[1] 0.61
2. Section B (scope photo)	[1] 0.45	....	[1] .35	[1] .37	....
3. Section C (E-6B)	[1] .51	[1] .35	....	[1] .49	....
4. Section D (airplot)	[1] .44	[1] .37	[1] .49	....	....
5. Sections B, C, and D (total)	[1] .61	....	....	....	....
Mean raw score	31.77	10.38	14.49	9.71	34.58
SD	6.49	3.53	3.40	3.02	7.70
N	309	309	309	309	309

[1] Significant at the 1-percent level of confidence.

A possible explanation for the differences in results in the three tables lies in differences in the samples. The sample in table 8.13 consists almost entirely of bombardiers, whereas the samples in table 8.14 and table 8.15 consists of approximately two-thirds navigators and one-third bombardiers. Navigators, by reason of their greater experience with navigation on aerial and trainer missions, will have had considerably more practice than bombardiers in the planning and organizational requirements of the simulated mission type of problem. For them it is possible that section A constitutes a different test than for bombardiers. If, for the navigators, the organizational features of navigation have become routine, it would be likely that their performance on the simulated mission would be determined in good part by their skill at operations tested in sections B, C, and D. The correlations of these sections with section A would be correspondingly large.

If, for the bombardiers of table 8.13, on the other hand, the simulated mission presented new and unfamiliar problems, the same relationships might be lower.

RELATED REPORTS FROM THE AVIATION PSYCHOLOGY PROGRAM

Since the training of the radar observer is similar to that of navigators and bombardiers, it is of interest to compare intercorrelations of proficiency measures reported here with those reported elsewhere by Psychological Research Project (Bombardier) and Psychological Research Project (Navigator). Correlations from other projects are available for printed tests with aerial performance checks, and printed tests with other printed tests.

Printed Tests Against Aerial Performance Checks

Two types of printed tests have been correlated with aerial performance checks. In one type each item is a separate and independent measure of some knowledge or skill; the other type attempts to simulate an aerial mission and to measure a series of related tasks. In the latter type, the items of the test are interdependent. The Bombardier Project has correlated the first type of printed test with an aerial check. The Navigator and Radar Projects report correlations for both types of tests.

The Bombardier Proficiency Test (form B) is a 3-hour examination providing a comprehensive coverage of basic bombing subjects. The Bombardier Project has developed standardized phase checks, which are aerial performance checks. For 107 cadets, the Bombardier Proficiency Test correlates 0.19 with Phase Check—Form 1. For 177 instructors from all bombing schools the test correlates 0.33 with Phase Check—Form 3.

TABLE 8.16.—*Correlations of printed tests measuring navigation proficiency with the navigation flight mission* [1]

	Printed tests		Mean raw score	SD	N
	Navigation proficiency test form A	Ground mission [2]			
Flight mission	0.09	[3] 0.35	3.14	0.22	261
Mean raw score	60.20	52.54			
SD	10.00	4.83			
N	261	231			

[1] AAF Aviation Psychology Program, Research Report No. 10, Psychological Research on Navigator Training, table 5.3, p. 49.
[2] The ground mission is a printed test which simulates an aerial mission.
[3] Significant at the 1 percent level of confidence.

The Navigator Proficiency Test (form A) consists of eight parts dealing with various aspects of navigation, including technical vocabulary, theory, and procedure. The flight mission developed by the Navigator Project is a standardized check of aerial performance; the ground mission is a printed test simulating an aerial mission. Correlations of the flight mission with the proficiency test and the ground mission are presented in table 8.16. The ground mission is correlated significantly with the flight mission while the proficiency test is not. Comparable correlations between radar printed tests and performance checks have been discussed earlier in this chapter. They are summarized for comparison in table 8.17.

TABLE 8.17.—*Correlations of printed tests measuring radar navigation and radar bombing with aerial performance checks*

	Final test I[1]	Final test II	Radar bombing intermediate test	Radar navigation intermediate test	Set operating intermediate test	Mean raw score	SD	N
Final aerial check	0.09	0.08	0.09	0.11	0.05	60.79	11.60	190
Intermediate aerial check	[2].25	.13	.06	.05	.11	45.38	8.55	189
Mean raw score	59.93	69.29	41.93	42.29	72.80			
SD	11.56	9.34	4.98	7.55	8.41			
N	190	190	190	190	190			

[1] Final test I is a simulated aerial mission type of test.
[2] Significant at the 1-percent level of confidence.

Correlations Between Printed Tests

The Navigator Project reports intercorrelations between the Navigator Proficiency Test (form A), the ground mission, and a written examination containing items of information about basic navigation. The sample consisted of 261 navigation students. The Navigator Proficiency Test (form A) and the written examination correlate 0.68. The Navigator Proficiency Test (form A) and the ground mission correlate 0.29. The ground mission and the written examination correlate 0.41.[3]

The Bombardier Project reports intercorrelations of the Bombardier Proficiency Test (form B) described in the preceding section, the Radar Navigation Test P–5B, the Bombardier Proficiency Test (form C), First Booklet, and the Bombardier Proficiency Test (form CN). The Radar Navigation Test P–5B is a printed test of basic navigation designed to select students for radar observer training. Bombardier Proficiency Test (form C), First Booklet, is a test dealing with basic bombing subjects and weather. Bombardier Proficiency Test (form CN) is a basic navigation test.

For a sample of 182 student bombardiers, the Bombardier Proficiency Test (form B) and the Radar Navigation Test P–5B correlate 0.51. The Bombardier Proficiency Test (form B) and Bombardier Pro-

[3] See Carter, L. F., ed. *Psychological research on navigator training.* AAF aviation psychology program research reports, no. 10. Washington: Government Printing Office, 1947.

ficiency Test (form CN) also correlate 0.51. The Bombardier Proficiency Test (form B) and Bombardier Proficiency Test (form C), First Booklet, correlate 0.57. The sample for this coefficient was 219 students.

Table 8.8 of this report gives intercorrelations of printed tests used by the Radar Project. For all three projects, the correlations between printed test scores are generally higher than the correlations between printed test scores and aerial performance scores. Both the Bombardier Project and the Radar Project obtained correlations between bombing tests and navigation tests which are only slightly lower than the correlations between two bombing tests or two navigation tests. None of the projects found the correlation between printed tests and aerial performance checks to be high.

SUMMARY

The chapter presents and interprets correlations between scores on printed tests and performance checks. Since the reliabilities of the measure reported are unknown, the analyses made are subject to qualification and are regarded as tentative only.

For purposes of analysis, the radar observer's task is broken down into three basic skills: radar navigation, bombing, and set operation. A fourth skill, oscilloscope interpretation, is not considered in this chapter because no proficiency tests or checks were developed for its measurement. One of the questions raised concerned the practicability of substituting tests for performance checks which are less easily administered and which require apparatus. No support of this possibility may be gained from the obtained correlations. Most coefficients are not significantly different from zero. The highest correlation obtained between a test and check measuring similar skills is 0.34 between the Set Operation Intermediate Test and the Final Bench Set Check. It is recognized that this lack of relationship might be a consequence of the unreliability of the measures involved.

Printed tests are closely related to performance checks as performance checks are to each other. Higher than either are the intercorrelations among printed tests. This difference holds regardless of the similarity or dissimilarity of the skills correlated. Printed tests which simulate radar observer skills are somewhat more closely related to performance checks than are printed tests of technical information. There is no indication that printed tests correlate more highly with checks given at the same period in the course than with those given at a different period.

Tests and performance checks measuring similar skills are found to be related about as strongly as are measures of unlike skills. Bombing scores are more closely interrelated than are navigation measures.

The relation of bombing and navigation scores to set operation scores is not investigated thoroughly because there are not enough set operation measures. The highest intercorrelations for measures of similar skills are generally found between two printed tests, but printed tests measuring unlike skills are almost as closely related as printed tests measuring similar skills.

There is no evidence that instructor ratings of student proficiency are more highly correlated from one performance check to the next than are the corresponding performance check scores. It had been hoped that these ratings would gain consistency by taking account of the influence of poorly standarized checking conditions.

Evidence seems to show that there is a considerable overlap between the three basic skills of the radar observer. Insofar as printed tests and performance checks are valid measures of particular skills, navigation and bombing skills appear to be most closely related, bombing and set operation skills somewhat less so, and navigation and set operation skills least related.

Intercorrelations of approximately 0.40 are found for three sections of Final Test I, each of which includes relatively homogeneous navigation computations. Correlations of each of these three sections with a simulated mission section are almost equally high. Because the operations in the three subtests constitute a major part of the simulated mission, it was expected that the correlation between the mission and the sum of the other three sections would be highest of all. The anticipated result was found for two samples of navigators but not for a sample of bombardiers. A hypothesis is offered for this discrepancy.

A comparison of proficiency measure intercorrelations reported by the Psychological Research Project (Bombardier) and the Psychological Research Project (Navigator) indicates a general uniformity of results. All three projects obtained their highest intercorrelations between printed tests. None found the correlations between printed tests and aerial performance checks to be high.

CHAPTER NINE

The Circular Error in Radar Observer Training[1]

INTRODUCTION

Circular error is a measure of bombing proficiency. From the standpoint of being a sample of the task required in combat, it is the most direct evidence of accuracy in the fundamental operation toward which radar training is directed. In this sense, it has more inherent validity than any other proficiency measure. Consequently, it constitutes an obvious criterion against which to validate selection tests.

However, the use of the circular error as a proficiency measure or validation criterion is contingent upon its reliability. If a student's circular error scores are unrelated to each other or if his standing in his group in regard to circular error changes from one measurement to the next, one must be skeptical of its usefulness for either of the above purposes. Therefore, the reliability of circular error is the central consideration of the present chapter. However, a brief review of recording and scoring procedures precedes the discussion of reliability since an understanding of these procedures is necessary to an understanding of circular error in radar training. The reliability section discusses the pertinent theoretical considerations in determining reliability, and the results of several reliability studies. Succeeding sections are concerned with other salient characteristics of the circular error in training. One section will discuss the learning curve of radar bombing, and another, the relationships between circular error and other proficiency measures. A final section deals with an analysis of a systematic bombing error commonly found in radar observer training, a range error in which bombs tend to fall beyond the target.

RECORDING AND MEASURING BOMBING HITS

During radar observer training, as students practice direct and coordinated bombing procedures,[2] their bombing accuracy for each bomb run is determined from records obtained by one of three meth-

[1] Written by Lt. George S. Klein.

[2] For a description of these procedures, see chapter 4.

ods: photographs of actual bomb drops, photographs of simulated bomb drops, and direct electronic recording by gun-laying radar equipment.

Photographs of Actual Bomb Drops

Most students drop actual 100 pound bombs containing a small powder charge on an artificial target during one or two training missions. A terrain photograph is taken at the moment of each impact. Since in this type of practice the impact point is directly discernible, determining the bombing error from the photograph is relatively simple. The scale of each photograph is a function of the altitude at which it is taken and accurate measures of error require, therefore, a precise record of the altitude at the moment of impact.

In general, radar bombing accuracy is believed to vary with bombing altitude, and circular errors are corrected for this factor. The usual practice is to convert all circular errors to a standard altitude of 12,000 feet. The intended effect of this correction is to render a circular error found at one altitude comparable to one obtained at any other altitude.

Photographs of Simulated Bomb Drops

The most common type of practice in radar training is simulated bombing, usually referred to as "camera bombing." The student performs all the operations of bombing either direct or coordinated but does not release actual bombs. A series of at least three vertical photos of the terrain is taken from the aircraft at predetermined points along the final portion of the bomb run. The exact time of each picture is obtained either from the time settings on the camera intervalometer, a device permitting automatic photos at prescribed points, or from a photographic recording of the time as each picture is snapped. The impact point in camera bombing is a hypothetical point which must be deduced from the photographic data. This is usually scored in terms of circular error and converted to the standard 12,000 feet altitude.

An important factor upon which hinges the practical use of the photo scoring techniques is the precision with which the hypothetical impact points can be located. Accuracy checks on this computation were made by dropping actual bombs on an artificial target and comparing the bomb impact points with photo-determined impact points of the same releases. It was found that at 10,000 feet, under the particular conditions of this comparison, impact points determined photographically differ from actual impact points by less than 100 feet.

Scoring is often carried out with the aid of bomb scoring mosaics. In this method the optical centers of the several photographs taken on the bomb run are plotted on a composite of several vertical photographs of the target area. The photographs are of known scale and altitude. From the plotted data on the mosaic, objective measures may be made of track, ground speed, true heading, and altitude. A line connecting the three plotted points on the mosaic gives the track. Ground speed is determined by measuring the distance on the mosaic between two of the plotted points and relating it to the time interval between them. True heading is found by relating the longitudinal axis of the photos to true north of the mosaic. Finally, photo altitude may be computed by relating the focal length of the camera lens used to the distance measured on the last photograph between any two points a known ground distance apart. With the aid of the above measures, and knowing the actual time of fall of the imaginary bomb, the impact point can be located on the mosaic.

The use of the mosaic allows the scoring of larger bombing errors than do the photographs alone. In instances where the error is very great, the target may not appear on a photo. In such cases the mosaic is frequently large enough to allow the plotting of impact point in an area of known relation to the target. Errors so extreme as to be outside the mosaic area are not scoreable even by this method.

If the same target area is not frequently used for camera bombing, the preparation of scoring mosaics is not justified. Fairly satisfactory scoring may be accomplished without the use of mosaics. The procedure is essentially similar to the mosaic method except that the photo data are not plotted on a mosaic. The photos taken during the bomb run are put together to form a photo strip. A line connecting the photo centers represents the track. The point of impact, as in the mosaic method, may be located with the aid of true heading, track, ground speed, and altitude which are obtained from the photo strip. As mentioned above, if the bombing miss is so great that the target does not appear on the photographs, the circular error cannot be accurately scored.

The photo scoring technique in camera bombing presents numerous problems in its everyday application. It rarely provides a continuous record of a student's bombing performance in training and, under some conditions, its accuracy is limited. Weather imposes a severe handicap in the use of the technique. The presence of an undercast prevents terrain photography. Moreover, accurate photography requires that the camera be leveled for each picture and oriented with respect to the longitudinal axis of the aircraft. Since camera equipment is not gyrostabilized, undue turbulence will frequently upset the vertical alignment. In addition, deficiencies in camera equipment or in the obtained photographs often result in indistinct reference points, making difficult the plotting of impact points. Inaccuracy may also occur from the fact that the photo recording is not wholly independent of student mistakes in using the recording equipment. For example, the "bombs away" picture is indicated automatically by the bombsight.

If an incorrect disc speed has been set into the bombsight, error will result which is unrelated to the student's proficiency. A major limitation of the photo scoring technique already noted is that frequently the bombing miss is so great that the target is not included in the picture.

Recording by Gun-Laying Radar (SCR-584).

This method was a late development in the radar training program and never achieved wide use. It employs a ground based radar device,

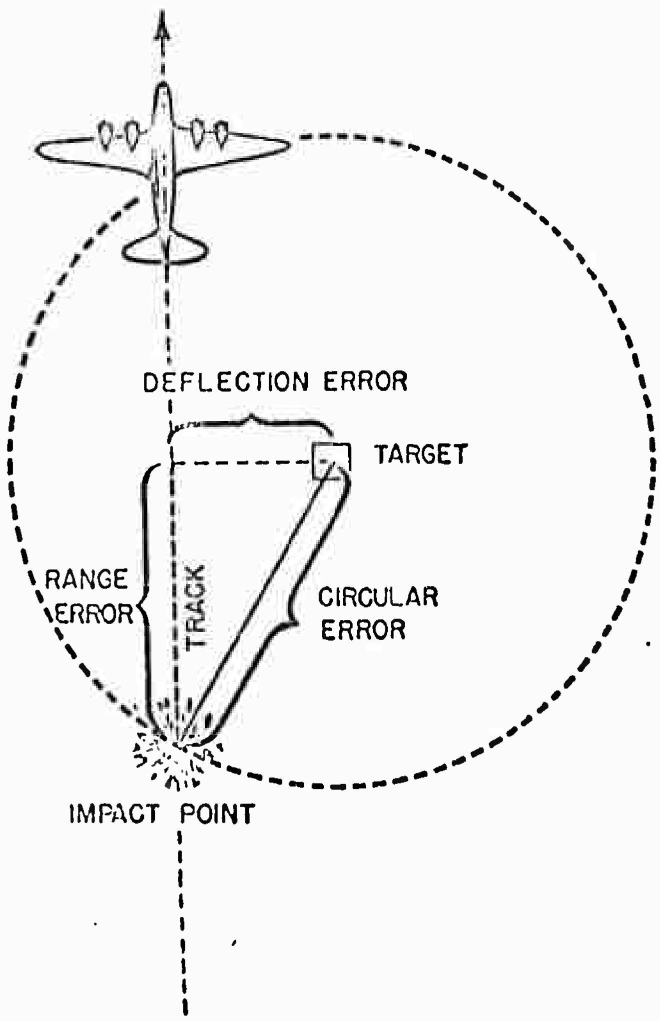


FIGURE 9.1.—Principal scores in radar bombing

developed originally for gunlaying which is placed in the vicinity of the target and tracks the aircraft on its approach to the target. A synchronized recording pen provides a paper record of the aircraft's track. The aircraft and ground radar operator are in radio communication. Through signals transmitted to the ground radar operator, the release and impact points can be plotted. A major advantage of the technique is that it overcomes the weather difficulties encountered with photography. The possibilities and limitations of this method are further examined on page 190 in connection with the discussion of creating a more reliable circular error measure.

Bombing Proficiency Scores

The procedures described above yield three principal types of scores for a bomb drop. These are shown in figure 9.1. Circular error may be defined as the distance measured along the radius of a circle whose center is the target and whose circumference passes through the bomb impact point. The deflection error of the impact point is the distance along a perpendicular dropped from the target, right or left, to the track line. The range error is the distance, over or short, from the bomb impact point to the intersection of this perpendicular line and the track line. In radar bombing training, the latter two measures were not ordinarily used since a more inclusive summary of bombing accuracy is provided by circular error.

Radar bombing errors may be analyzed into other components in addition to circular error, deflection error, and range error. The photo and gun-laying radar methods permit objective measurement of true heading, track, ground speed, and altitude, all of which are also computed by the student on each bomb run. Scores for these factors would add to the variables measured for each bomb run. Further study of these components is warranted as part of the search for reliable measures of bombing proficiency.

DESCRIPTION OF SAMPLES USED IN CIRCULAR ERROR STUDIES

Bombing records were selected for analysis largely from two radar training schools, Boca Raton and Victorville. These schools were among the first established in the radar observer training program and possessed well-established bomb scoring departments. As a result, they had available the largest samples of pre-VJ-day bombing data. Records of a third school, Langley Field, were used for a single study, the comparison of proficiency measures and average circular error.

The bombing records from the schools differ in completeness, number, and scoring method used. All bombing at Boca Raton and Victorville was done with the AN/APQ-13 set and was scored by the mosaic method, whereas at Langley Field bombing was performed on AN/APS-15 or 15A and was scored by the nonmosaic method. Records of actual bomb drops were available only for Boca Raton. Because of the differences existing among the schools, the principles guiding the selection of data from each varied somewhat. Some special characteristics of the data for each school may be noted.

Boca Raton

All students were rated bombardiers. When the present study began, data were available for five pre-VJ-day classes which were graduated prior to VJ-day: 555, 615, 625, 635, 645. These classes represented actual bombing records for 115 students and camera bombing records for 143 students. Records were selected for analysis which had a total of four or more scored photographed drops for a student. In all, 141 actual bombing records and 112 camera records met this requirement. The minimum of four scored drops was chosen because it represented the best compromise between the desirability of having as many drops as possible per student, and the undesirability of reducing the sample greatly by setting the minimum too high. A total of 110 student records had a minimum of four scored drops for both actual and camera bombing. In the actual bombing missions at Boca Raton, virtually all runs were made on one target. The data for a second target was discarded for the present studies.

Victorville

In this school the students were rated navigators. The camera bombing records from Victorville were more complete than those of either of the other schools and provided the largest single sample studied. Individual records were available for students of 19 classes, 45-16 through 45-35 (except 45-24), all of whom had completed their training missions before VJ-day.

Only those records which had a total of 10 or more scored drops were selected as the sample used in the reliability analyses. This selection provided a total of 400 cases for the odd-even drop reliability study. Ten drops closely approximates the mean number of scored runs made by students at this school. Use of a higher standard for minimum number of drops would have reduced the sample to a much smaller number.

Langley Field

The students at this school were rated bombardiers. Camera bombing records were available for seven classes, 45-30 through 45-36, totalling 179 cases. Of these, only 4 classes, comprising 111 students had completed all bombing training before VJ-day. The bombing records could not be studied for reliability. The recorded data for individual bomb drops were very meagre and no odd-even drop reliability coefficient could be computed. Moreover, the number of missions recorded for any one student were too few to permit any study of odd-even mission reliability.

The limitations of these data stem in part from the weather conditions existing at Langley and in part from the use of the nonmosaic method of scoring. Weather was a constantly disturbing factor, resulting in frequent changes in planned missions and interfering with the photo recording of drops. Because the nonmosaic method was used, extreme bombing misses were not scored. As was seen earlier, when the target falls outside the field of view of the photographs, the nonmosaic method cannot estimate the circular error of the bomb-hit with acceptable accuracy. Bombing errors of this magnitude occurred frequently but were not included in the average circular error computed by the school for each student. For these reasons the Langley data proved useful only in the study of the relationship between average circular error and proficiency measures. The post VJ-day classes were retained in order to provide a fair-sized sample.

THE RELIABILITY OF AVERAGE CIRCULAR ERROR

A perfectly reliable measuring device will place individuals in the same rank order for a number of measurements separated in time. In practice, measurement with perfect reliability is rarely encountered. It is important to a discussion of reliability coefficients determined by different methods to note the nature of the factors which may influence reliability. These factors fall into two broad categories: variation of conditions under which measurement occurs and variations within the individuals measured. If students are measured under different conditions from one testing to another, such variations will, in random fashion, penalize them (raise their circular error scores) at one time, and help them (lower their circular error scores) at another. Since such results reflect variations in testing conditions, as well as individual differences in bombing accuracy, the results of a single mission can yield only limited generalization about true differences in bombing ability. Likewise, individuals do not always perform with the same efficiency from one testing to another due to variations in their physical condition, motivation, attitudes, learning rate, etc. As in the case of variations in testing conditions, intraindividual variations will randomly raise or lower a student's score. A reliability coefficient, to be comprehensive, must reflect the influence of both kinds of factors. It must summarize the effects of all conditions that limit the extent to which generalizations may be made as to true differences in bombing ability.

A variety of influences of the two types mentioned operate in the radar training situation. Some of these are shown in figure 9.2 in terms of their effect in altering a student's rank from his true rank within the group. By "true rank" is meant the rank order determined by a very large (theoretically infinite) number of bomb drops obtained for an individual under normally varying testing conditions. At the top of figure 9.2 is an hypothetical distribution of true ranks for a very large number of radar observer students, each point on the scale representing a true rank.

Bomb releases in radar training are separated in time in two ways: within mission (from drop to drop) and between mission (from day to day). Figure 9.2 shows the varying influence on circular error of mission-to-mission and within mission (drop-to-drop) variability in each of the factors listed. The direction of the influence of each factor is shown by an arrow. The varying length of the arrows is intended to indicate that the variables probably differ with respect to their relative influence upon mission-to-mission variation and drop-to-drop variation. However, the relative importance of each factor as shown in the figure is purely speculative. More detailed consideration of the specific sources of unreliability in radar performance and their relative significance will be deferred to a later discussion (see p. 185). For the

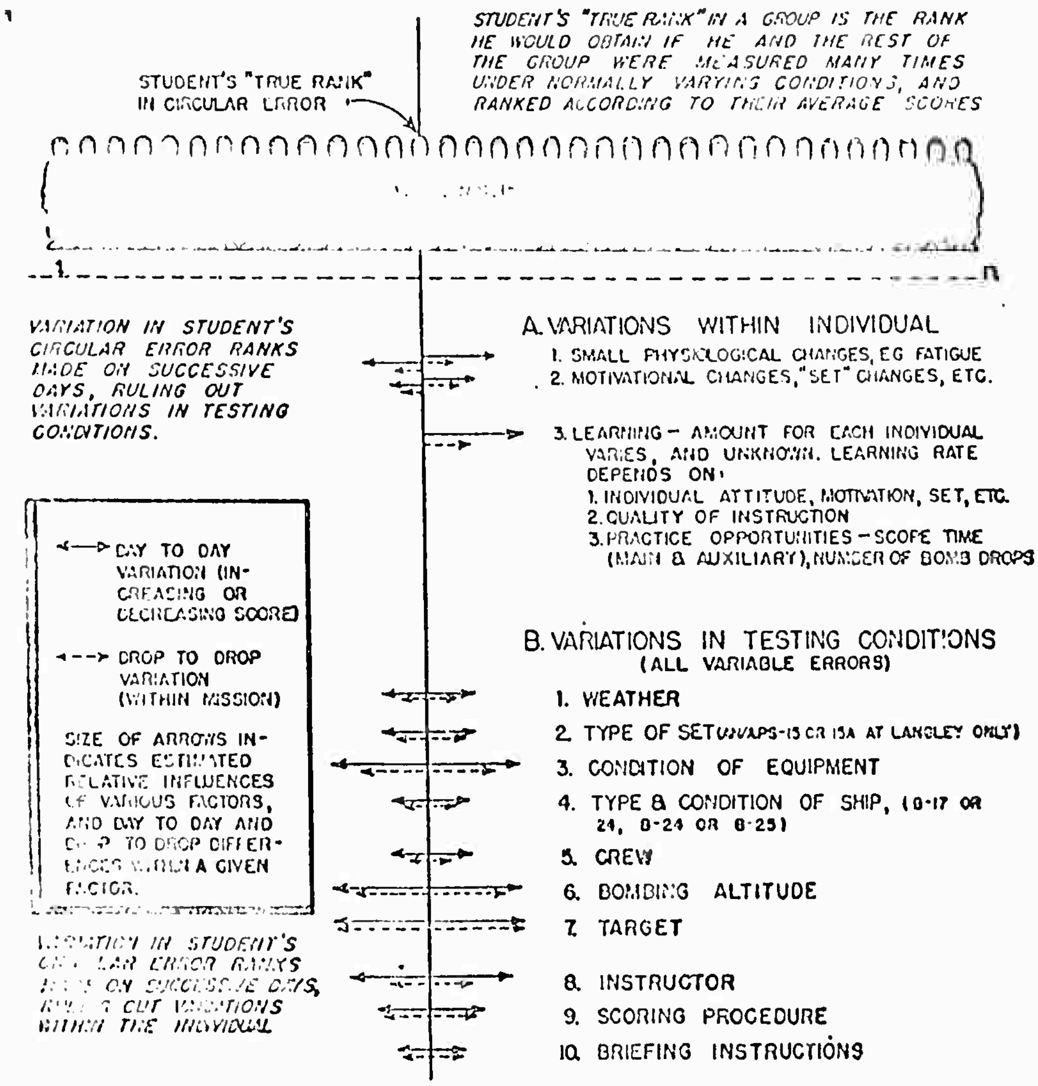


FIGURE 9.2.—Sources of unreliability of individual circular error in radar bombing performance.

present it is pertinent to note only that mission-to-mission fluctuations in the variables listed influence scores to a greater extent than do drop-to-drop fluctuations.

The upper portion of figure 9.2 illustrates how much variation might occur in an individual's rank around his "true rank" if all extraindividual sources of unreliability and all effect of memory and learning were eliminated. In such a case, each successive circular error measure would be obtained under identical conditions and within a learning plateau and would, consequently, be affected almost solely by small physiological and psychological changes.

Evaluation of Methods of Determining Reliability

As pointed out, the merit of a method for use in estimating the reliability of circular error scores is determined by the degree to which it reflects the influence of both mission-to-mission and drop-to-drop variations. A measure of reliability that fails to reflect mission-to-

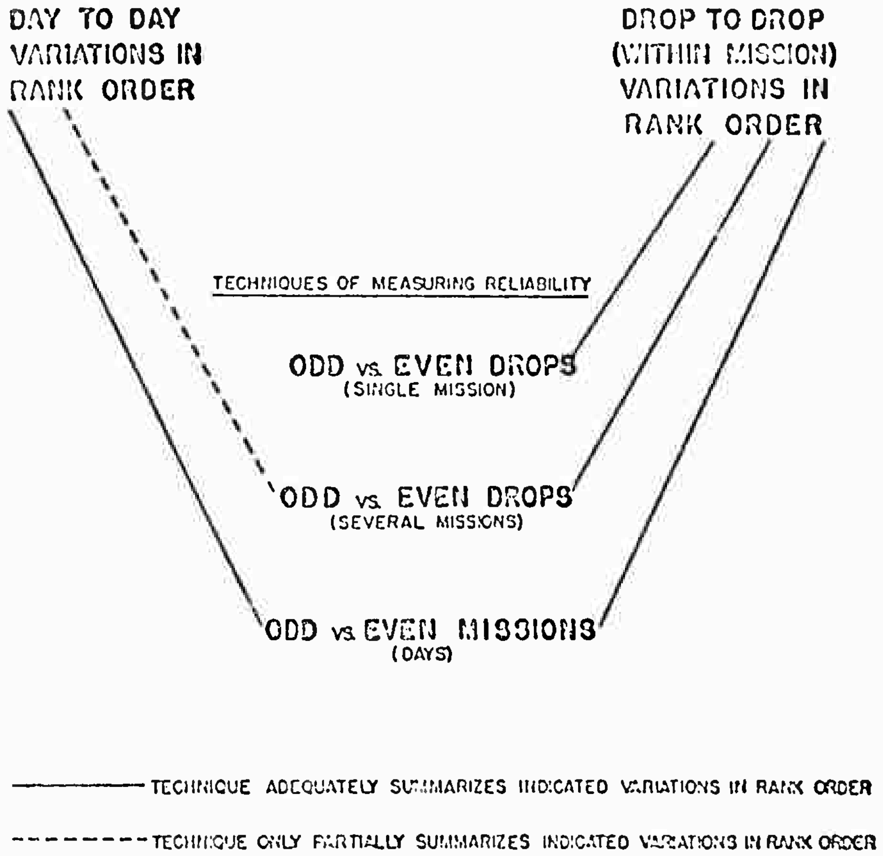


FIGURE 9.3.—Evaluation of reliability methods applied to radar bombing performance. Reliability coefficient should take into account the variability of the ranks of obtained circular error scores caused by all the factors listed in figure 9.2.

mission variations, in particular, seriously exaggerates the extent to which a set of scores warrants generalization as to true bombing ability. Figure 9.3 summarizes a comparison of three principal methods which may be evaluated in this respect.

1. *Odd-even drops on one mission.*—Of the three methods, this one provides the least adequate index of reliability since it fails to measure any mission-to-mission variations. Since it is likely that the effect of within mission variable errors is relatively small, as suggested in figure 9.2, this method gives a spuriously favorable picture of cir-

cular error reliability. An important characteristic of this method is that factors which remain constant over the entire mission are correlated with the scores of all drops. For example, relatively temporary factors such as pilot, crew, set, and altitude, which differ little from drop to drop within a mission, but which would normally vary over several missions, affect scores on both halves of the single mission in the same way. This spuriously raises the reliability coefficient because the odd and even series are affected equally and in the same direction by the correlated influences.

2. *Odd-even drops on several missions.*—The division of drops in this method is such that a part of the drops from each mission are placed in the odd series and a part in the even. As a result, the constant conditions of each mission will influence both the odd and even series in the same direction, thus obscuring the attenuating effect of mission-to-mission variation. In this respect, the objection raised to the method of odd-even drops on one mission applies to this method as well. However, as number of missions represented in the odd and even series of drops increases, the effect of conditions existing for any one mission will influence less and less the size of the odd and even averages. Any interpretation of reliability coefficients obtained by this method must, consequently, consider the number of missions from which the data are taken.

3. *Odd-even missions.*—This method provides the only adequate estimate of reliability of average circular error since it takes into account both mission-to-mission and drop-to-drop variations. Since the student's performance is divided in terms of missions, a correlation of the odd missions against the even is affected by all factors which cause the scores to fluctuate. Since it measures the effects of more sources of variation than do the other two methods, this method may be expected to yield appreciably lower reliability coefficients.

The foregoing comparison permits predictions of the relative sizes of the reliability coefficients likely to result from application of three techniques. In the comparison, of course, an equal number of drops is assumed. Coefficients obtained from odd-even drops on single missions will be higher than those from odd-even drops on several missions and both of these will be higher than coefficients from odd-even missions. Coefficients from odd-even drops on several missions will decrease as a function of the number of missions represented in the series of drops. As explained, the amount of mission-to-mission variability introduced by a large number of missions should tend to reduce the effect of any one extreme mission on both the odd and even series of scores, a factor increasing the reliability coefficient spuriously in single mission studies. Data to test this latter hypothesis were not available.

Statistical Studies of Reliability

The method of odd-even drops on one mission could not be used to compute reliability in the present study because too few drops were made within a single radar training mission. Each of the two remaining methods described above was applied. In the method of odd-even drops on several missions, all drops for each student were divided into an odd-even series and averages for each series were correlated. Varying numbers of missions are represented in the series. In the second method the student's data were divided into an odd and even series of missions and the average circular error of the odd series of missions were correlated with that of the even series. In this procedure the number of drops within each mission was permitted to vary. Because of the nature of the data, these methods were applied somewhat differently at Boca Raton and Victorville. For this reason the results for the two schools will be discussed separately. A summary section will bring together all results and state general conclusions.[3]

Boca Raton.—Although similar bombing procedures were used in actual bombing and camera bombing practice, it seemed advisable to examine the reliability of each form of practice separately. As shown in table 9.1 there is no correlation between the two ($r = -.08$, $N = 118$). The difference in average CE is also considerable (mean actual=1,109.82 feet; mean camera=5,042.87 feet). While, as will be shown later, this absence of correlation may be attributed partly to low reliability, it is probably due to some extent to marked differences under which the two types of bombing were carried out. Actual bomb runs were all on one target and at two altitudes—10,000 and 15,000 feet—while camera bombing was done on 20 targets which varied in complexity and altitudes varying from 8,000 to 20,000 feet. The use of one target simplified the navigation problem in actual bombing. The nature of the targets differed sharply between the two types of practice. In actual bombing a pin-point target was used, surrounded by water and easily recognizable on the scope. In camera bombing, aiming point recognition was a difficult problem since targets were usually located in industrial areas not easily differentiable on the scope. A final consideration of unknown effect is that the position of the actual bombing series in relation to the camera bombing series varied from student to student. For some students, actual bombing was done early in training, while for others, it was done either in the middle or towards the end of training. Comparison between the two types of

[3] The reliability of circular error may differ for direct and coordinated bombing. Student bombing records at both Victorville and Boca Raton did not distinguish bombing runs made by one or the other technique. However, it is known that the number of direct bombing runs attempted and photographed at the two stations were negligible. It may be assumed, therefore, that the reported reliability coefficients for actual and camera bombing reflect primarily the reliability of coordinated bombing procedure.

practice will be difficult to interpret unless the place of actual bombing mission is held constant in the entire series of practice bombing runs.

TABLE 9.1.—*Correlation between actual bombing and camera bombing for 118 Boca Raton students*[1]

Source of average circular error	Mean	SD	r
Actual bombing	1,109.82	590.59	−0.081
Camera bombing	5,012.87	1,773.54	....

[1] Camera records with 3 scored drops were admitted into the sample.

1. *Odd-even drops on several missions.*—The mean number of total scored drops in both types of bombing was considered to be too small to hold constant the number of drops in the odd and even series for each student. As seen in table 9.2, the mean of actual bombing is 8.53, with an S. D. of 3.15, and the mean of camera bombing is 5.34 with an S. D. of 2.31. Only bombing records with four or more scored drops were used in the study. This provided reliability samples of 144 and 112 students for actual bombing and camera bombing respectively. In both sets of data, average CE's were computed for the odd and even series of each student and correlated.

TABLE 9.2—*Description of Boca Raton circular error data*

Variable	Actual bombing data					Camera bombing data				
	N	Mean	Median	SD	Range	N	Mean	Median	SD	Range
Number of drops photographed for each student	145	8.53	8.37	3.15	[1] 2–17	143	5.34	5.21	2.31	[1] 1–13
Number of missions in which each student made scored drop	145	2.20	2.71	.85	1–5	143	3.23	3.23	1.00	1–6
Lowest altitude at which each student made scored drop	145	11,845	10,683	2,217	[2] 10,000 to 15,000	143	10,070	10,378	1,059	5,000 to 15,000
Highest altitude at which each student made scored drop	145	15,093	15,458	1,382	[2] 10,000 to 15,000	143	14,234	15,258	2,274	10,000 to 20,000

[1] Minimum of four drops required for inclusion in odd-even drop reliability study.
[2] All actual bombing was done at recorded altitudes of from 10,000 to 11,000 or from 15,000 to 16,000 feet.

The results are shown in the upper portion of table 9.3. The uncorrected r's are 0.46 for actual bombing and 0.35 for camera bombing, both significant at the 1-percent level. Corrected for length by the Spearman-Brown formula, the r's are 0.63 for actual bombing and 0.52 for camera bombing. Since the numbers of drops in the odd and even series were not equalized, the Spearman-Brown formula is not strictly applicable. In using it here, it is assumed only that the corrected correlations approximate those which would be obtained from the lengthened series somewhat more closely than do the uncorrected correlations.

Table 9.3.—*Reliability results for circular error data, Boca Raton and Victorville*

Type of bombing practice	Variables correlated	N	Classes	Mean, odd	SD, odd	Mean, even	SD, even	r'_{11}	r_{11} [1]
Boca Raton:									
Actual	Mean odd versus mean even drops	141	555, 615, 625, 635, 645	1,174	728	1,073	605	[2] 0.46	0.63
Camera	Mean odd versus mean even drops	112	555, 615, 625, 635, 645	5,283	2,067	5,031	2,160	[2] .35	.52
Camera	Mean odd versus mean even missions	112	555, 615, 625, 635, 645	5,454	2,595	5,078	2,644	[3] .19	.31
Victorville:									
Camera	Mean odd versus mean even drops	134	45-16 through 45-20	4,580	1,623	4,386	1,381	[2] .31	.45
		268	45-21 through 45-35	4,477	1,459	4,019	1,363	[2] .32	.48
Camera	Mean odd versus mean even drops (TNV's scored).	134	45-16 through 45-20	6,423	2,601	5,821	2,255	[2] .39	.56
Camera	Median odd versus median even drops	134	45-16/45-20	4,274	1,767	3,861	1,484	[2] .27	.42
		268	45-21/45-35	3,894	1,534	3,530	1,425	[2] .16	.28
Camera	Mean odd versus mean even missions	372	45-16 through 45-35	4,592	1,821	4,169	1,530	[3] .11	.20

[1] Corrected by the Spearman-Brown formula.
[2] Significant at the 1 percent level.
[3] Significant at the 5 percent level.

If learning occurs and if practice, as indicated by the number of scored drops, is permitted to vary from student to student, the more the practice the smaller the average CE of both the odd and even series of a student. This factor thus acts as a constant error, influencing both series in one direction and affecting, by itself, a positive correlation. As a means of measuring this effect in the present data, practice was partialled out of the odd-even drop coefficients. The results are presented in table 9.4. It is apparent that variation in the number of drops is a negligible influence in both types of bombing. One interpretation of this result is that only a small amount of learning occurred in the sample of bombing practice represented by the scored, photographed drops. All scored drops occurred on an average of two and three missions, as seen in table 9.2.

TABLE 9.4.—*Odd-even drop reliability coefficients with and without bombing practice partialled out, Boca Raton students*

Type of bombing practice	N	r_{12}[1]	r_{13}[1]	r_{23}[1]	$r_{12.3}$[1]
Actual bombing	113	−0.19	−0.08	0.46	0.46
Camera bombing	110	−.09	−.23	.35	.34

[1] 1=odd distribution; 2=even distribution; 3=number of runs (practice).

2. *Odd-even missions.*—Only camera bombing data were used in the study carried out by the method of odd-even missions. Actual bombing spread over too few missions to make possible a similar analysis. Since the average number of bombing missions was too small to permit holding missions constant at four (Mean=3.23, SD=1.00), the number of missions was allowed to vary. However, only records with a minimum of two missions are represented in the sample. This provided a total of 112 cases.

The missions for each student were numbered and divided into an odd and even series. The average CE's of each series of odd missions were then correlated with the average CE's of the even missions. The range of missions for the odd series was one to three missions, and for the even, one to two missions. It is apparent from this that the systematic error due to variation in the number of missions is not likely to be great enough to require partialling it out.

The results are presented in table 9.3. The uncorrected r is 0.19, significant at the 5-percent level. When corrected for length of the series, this figure becomes 0.31. The Spearman-Brown formula is used with the assumption already stated.

Victorville. 1. *Odd-even drops on several missions.*—Only records which had 10 or more scored and photographed bomb drops were selected as the sample for all circular error studies carried on with the Victorville data. This provided a group of 400 cases for the study of odd-even drops on several missions. For this analysis, the sample was divided into two groups, one consisting of classes 45–16 through 45–20 (N = 131), and the other 45–17 through 45–35 (N = 268). The difference between the groups lay in the treatment of runs recorded as TNV, which means target not visible on photo. As was seen earlier, such drops were difficult to score accurately with the photo-scoring method. In the early classes (45–16 through 45–20), TNV drops were not scored by the school. Either a gross estimate of circular error was recorded for these runs, for example, "greater than 5,000," or merely the notation "No target" was recorded. In later classes, 45–17 through 45–35, the school attempted to estimate the circular errors in such instances. In bombing records for these classes, it is not possible to distinguish TNV releases from other bomb drops. TNV drops represent either extremely poor bombing performance or the operation of disturbing influences beyond the student's control. In the former case, an average CE which fails to include such drops would be spuriously low. For this reason, an attempt was made to take account of the TNV drops in the early classes of the sample. Each such drop occurring in the series of 10 was given an arbitrary circular error of 16,000 feet. This value closely approximates the highest circular error found for any single scored photo drop in the Victorville records. The data for the latter group of classes, 45–21 through 45–35, could not be rescored since TNV runs were indistinguishable from others. Thus, the study of odd-even drops for Victorville yielded three coefficients: one for each sample based upon the uncorrected data and one for the first sample based upon the data corrected for drops marked TNV.

For this study, practice was held constant. Only the first 10 drops for each student were used and these were divided into an odd-even series with five drops in each series. The number of missions, on the other hand, varied in both the odd and even series, but did not vary systematically. In other words, the 10 drops represent a varying number of missions from student to student.

Since the data for Victorville were more numerous and complete than for Boca Raton, more detailed analyses were possible. Odd-even drop reliability was computed by two methods: correlatiton of odd-even mean CE's and correlation of odd-even median CE's. Mean CE is average circular error computed as the arithmetic mean of a series of drops; median CE is average circular error computed as the median of the series. The latter method was suggested by the hypothesis that extremely large circular errors include a greater measure of the factors causing unreliability in the testing situation than do smaller errors. Since such extreme scores exert more effect on a mean than on a median, the median circular error, if the hypothesis is true, should reflect to a greater extent the student's true bombing ability. This hypothesis is, of course, somewhat opposed to the hypothesis previously advanced that extreme bombing errors represent very poor performance and should be included in a student's average CE to represent his proficiency adequately.

The results are given in table 9.3. All uncorrected *r*'s between the odd and even means for both early and late classes are fairly high, 0.31, 0.32, and 0.39. All are significant at the 1 percent level. Corrected by the Spearman-Brown prophecy formula, they become 0.48, 0.48, and 0.56. Within the early classes, the odd-even mean *r* when TNV scores were included is higher than when they are not included, the correlation being 0.39 compared with 0.31. It is worth noting again in this connection that the "target not visible" runs of a student were scored in their temporal position in the series of 10 drops. The arbitrary scores assigned them, therefore, occurred in unsystematic fashion in the odd and even series of drops. For this reason, it is unlikely that the rise in correlation was occasioned spuriously by a constant error associated with the TNV scores. This result suggests the possibility that a more accurate method of scoring extremely high circular errors will favorably affect reliability. A later discussion will consider the point further.

The odd-even mean correlations are appreciably higher than the odd-even median correlations which, in the earlier and later samples, are 0.27 and 0.16, uncorrected. It appears that median CE will not be a more reliable score than mean CE. The implication is that extreme scores generally are not mainly due to large variations in external conditions and are no less typical of the student's usual performance than are his lower scores.

2. *Odd-even missions.*—In this study the number of missions was held constant at four, the approximate mean number of missions flown by the group. Of the 400 records with 10 or more drops, records were selected which had the required four missions. A total of 372 cases satisfied this requirement. The data for both early and later Victorville classes were combined for this study. Only scored and photographed drops were included in the average CE for each mission. TNV drops of the early classes were not used.

The first four missions for each student were split into an odd and an even series. Average CE's were obtained for the odd and even missions and the two series were then correlated. Since in training the number of scored releases varied from mission to mission, there is a possibility that the odd and even series were not equated with respect to practice However, this factor was probably unimportant in the present study inasmuch as a negligible relationship ($r = -0.07$, N=400) was found between average CE and number of scored bomb runs (practice) in the Victorville population.

The results are shown in the lower part of table 9.3. The uncorrected odd-even mission reliability is 0.11, significant at the 5 percent level. Corrected to twice the length of the present series by means of the Spearman-Brown formula, the r becomes 0.20.

Discussion of Reliability Results

As predicted in the discussion of the three methods of determining reliability, the odd-even drop reliabilities are consistently higher than the odd-even mission reliabilities. It was found that the reliability coefficient of odd-even drops for actual bombing was highest of all. The explanation suggested is that actual bombing, being concentrated into fewer missions, was less affected by mission-to-mission sources of unreliability than was camera bombing. Such variation as did occur between missions was less for actual bombing than for camera bombing since, for the former, bombing was carried out on 1 rather than on 20 targets and at limited altitudes rather than at widely varying altitudes. Another point is that scoring actual bomb drops was simpler than scoring camera drops where the impact point was hypothetical and computed. This would tend to free the actual bombing data from unreliability attributable to the scoring process.

It is to be noted that even though CE of actual bombing is more reliable, it is not, because of this, necessarily a more valid index of bombing proficiency. Since in training, actual bombs must naturally be dropped upon unrealistic targets, the CE resulting may be less representative of over-all bombing ability than is the average CE of camera bombing. It will be recalled that no target identification problems existed on actual bombing missions since a single pin-point target was bombed. Of interest to this comparison are the lower correlations found between actual bombing CE and certain proficiency measures as compared with the r's between camera CE and the same measures. A later section will discuss this finding.

In view of the numerous influences which normally vary in the radar observer training program, the method of odd-even missions affords the most useful estimate of the reliability of circular error. Reliabilities for odd-even missions are low but the uncorrected coefficients are significant at the 5 percent level. The favorable effect on reliability of scoring extreme bombing errors calls for improvements in the procedure for recording bombing errors and computing average CE.

In another part of this chapter, some of the major sources of unreliability of bombing circular error in radar observer training are specifically discussed and suggestions are made for reducing their influence on circular error.

Comparison of Results With Studies of Visual Bombing

Studies of reliability of average CE in AAF visual bombing are in accord with the general findings of this study. Odd-even drop reliabilities have been reported to be considerably higher than odd-even mission reliabilities, with the coefficients for odd-even drops on a single mission being highest. When such factors as pilot, bombsight, aircraft, and other factors which normally vary in bombardier training are permitted to remain uncontrolled in the data, odd-even drop reliabilities for visual bombing are of the same order of magnitude as those reported here. In general, however, the odd-even mission reliabilities of the present study, corrected r's of 0.20 and 0.31, are somewhat higher than similar coefficients for visual bombing.

In connection with the somewhat higher reliability of radar mission bombing, it should be noted that average CE scores are considerably more variable than those of visual bombing. The standard deviation of the radar scores are typically about 1,500 feet while the standard deviations of visual bombing CE scores are approximately 85 feet. If the larger standard deviation reflects a greater variability among students in radar bombing skill than exists in visual bombing skill, the variance due to uncontrolled influences will be relatively less for the former. If there is a considerable range of true ability within the group, the effect on rank order of extraneous factors will be less than if students differed from each other by small amounts.

Several visual bombing studies have succeeded in isolating and measuring the influence on circular error of certain major variables. Adequate control of these factors results in a marked increase in reliability. In general, thus far, only statistical controls have been attempted and these have had very limited application. Certain variables such as weather do not easily lend themselves to control through statistical adjustment of scores. A program of administrative control of bombing conditions in visual bombing training was initiated in a long-range study by Psychological Research Project (Bombardier). For a discussion of this and other matters related target was bombed. Of interest to this comparison are the lower to the reliability of visual bombing, the reader is referred to *Psychological Research on Bombardier Training*, report No. 9 in the series of AAF aviation psychology program research reports.

Variable Factors in Radar Bombing Missions

The efforts made in visual bombing studies to control and correct for the influence of sources of unreliability suggest a similar study of radar bombing. This section will explore some of the major sources of variation in radar bombing circular error and suggest tentative methods for reducing their influence. These sources have already been listed in figure 9.2. Some tentative ideas concerning their relative influence may be gathered from table 9.5 in which is presented descriptive data tabulated from the bombing records used in the Victorville analyses.

TABLE 9.5.—*Description of Victorville circular error data*[1]

Variable	Classes 45-16 to 45-20					Classes 45-21 to 45-35				
	N	Mean	Median	SD	Range	N	Mean	Median	SD	Range
Highest altitude at which each student made scored drop	......	(2)	(2)	(2)	(2)	268	15,284	15,414	978	12,000 to 18,000
Lowest altitude at which each student made scored drop	......	(2)	(2)	(2)	(2)	268	10,814	10,863	1,593	7,100 to 14,200
Difference between highest and lowest altitude at which each student made scored drops	......	(2)	(2)	(2)	(2)	268	4,514	4,476	1,697	400 to 9,000
Number of missions in which each student made scored drops	133	4.38	4.34	1.08	2-8	268	4.25	4.22	.90	2-7
Number of targets bombed by each student	133	6.59	6.52	1.49	2-10	268	6.62	6.65	1.19	3-10
Student's main scope radar time (to first scored drop)	133	9:39	9:20	4:01	1:30 to 19:00	268	13:03	13.01	4:18	3:00 to 27:00
Student's main scope radar time (to tenth scored drop)	133	22:55	23:1[illegible]	4:36	11:00 to 35:00	268	28:28	28:39	5.49	14:30 to 48:30
Student's main scope radar time (difference between first and tenth scored drops)	133	14:16	14:24	4:05	4:00 to 27:00	268	15:25	15:07	5.18	4:00 to 36:30
Number of bomb runs to mission on which student first made scored drop	116	3.83	2.13	4.74	0-22	268	4.92	4.68	3.27	0-19
Number of bomb runs to mission on which student made tenth scored drop	116	18.87	18.10	6.08	2-39	268	17.85	17.96	5.43	6-35

[1] Only records with a total of 10 or more scored, photographed releases used in study.
[2] Data not recorded.

Variations in bombing conditions. 1. *Variations in targets.*—The nature of targets bombed was varied differently at Langley Field than at Boca Raton and Victorville. In the latter schools missions were planned in terms of increasing difficulty and complexity of targets but at Langley Field weather conditions and terrain resulted in the assignment of routes and targets in almost random order. For all schools the relative difficulty of targets was unknown but probably varied to a considerable extent.

Students varied also with respect to the number of different targets bombed, as shown in table 9.5. For the data analyzed at Victorville, it was possible for a student to release all his bombs on as few as two targets or on as many as 10 targets. The average for all students was seven targets. This fact is of possible significance because of the general assumption that bombing becomes more accurate as familiarity with a target increases.

2. *Variation in number of missions.*—Table 9.5 discloses that the number of missions on which a student dropped his minimum of 10 bombs ranged from 2 to 8. This means that in the odd-even drop reliability study where number of missions was permitted to vary, the records of some students reflect to a greater extent than others the mission-to-mission variability in the factors causing unreliability.

3. *Variation in bombing altitude.*—At Victorville, bombing altitude varied between missions from 7,100 feet to 18,000 feet, as can be seen in table 9.5. At Boca Raton the variation in camera bombing is even greater, the range being 5,000 feet to 20,000 feet as shown in table 9.2. Similar data are unavailable for Langley Field, but it is known that variable weather conditions often necessitated changes in altitude ranging from 8,000 feet to 20,000 feet. Table 9.5 discloses that considerable differences existed among students at Victorville in the range of altitudes at which each student dropped his bombs. Some students dropped all their bombs within a restricted range of altitudes during the training period, while others released theirs at widely varying altitudes.

The amount of fluctuation in circular error scores arising from variations in altitude was not measured. As was seen earlier, radar circular errors were converted to a standard 12,000 feet altitude. If the conversion factors are correct, variation in altitude should not lead to measurement error. This could not be verified since no information was available concerning the criteria employed in the construction of the conversion table which was in use at the radar training schools. In view of the marked variations noted in practice bombing altitudes, this factor becomes of great potential significance. It is clear that the conversion factors used should be determined empirically from the altitude differences found in radar training rather than from those known to exist in visual bombing.

4. *Variation in weather conditions.*—This factor was undoubtedly strongest at Langley Field where weather was characterized by high winds, turbulence, frequent fogs, and rain. Weather frequently changed during a mission and this necessitated alterations in course, altitude and target. A further consequence of this type of weather is that it severely limits the use of the photo method of scoring. For this reason, at Langley Field the number of usable photo drops for each student was quite small.

5. *Variations in radar equipment.*—A highly important influence on scores at all schools resulted from variations in the condition of the equipment from mission to mission and, to some extent, from drop to drop on the same mission. Equipment condition at times ranged from excellent operation to complete malfunction with a single mission.

6. *Variations in crew efficiency.*—This factor varied considerably between missions and to a lesser extent within missions. The amount of error attributable to crew efficiency may be a function of other factors, such as time of day of mission, first or second halves of a mission and the experience and motivation of the crew. Examples of crew operations in which variation will influence CE are pilot performance in drift correction and in maintaining altitude or course,

and bombardier performance with the bombsight. Interphone difficulties among the crew provides a further illustration.

7. *Variations in aircraft.*—Missions at the three schools represented in this study were conducted in either B-17 or B-24 aircraft. The cramped quarters and generally difficult working conditions for the radar observer in the B-24 as compared to the B-17 were a constant source of complaint by both students and instructors. Assignment to one or the other type of aircraft was not always systematic from mission to mission. In other schools where B-24 and B-25 aircraft were used, a somewhat different problem existed. The higher cruising speeds of the smaller bombers increased the difficulty of navigation and, to some extent, shortened the time available to the radar observer on the bomb run.

Variations in the condition of the aircraft also effected bombing scores to some extent. Thus, for example, the condition of the interphone equipment, the operating effectiveness of auxilillary instruments, such as the bombsight and altimeter, and arrangements made to increase the visibility of the scope picture by blacking out the working compartment could all affect circular error to a noticeable extent.

8. *Variation in briefing procedure.*—Briefing procedures at each school, consisted of presentation of scope pictures, descriptions of routes and aiming points, and announcement of weather information. These procedures varied in quality between missions and between briefing officers. The analysis through scope pictures of aiming points and routes differed with respect to accuracy and completeness. Some briefing officers took pains to organize the briefing in detailed fashion while others did not. It is recognized that aiming point identification is a major source of difficulty in radar bombing. Variation in the quality of briefing instructions could affect identification of these points and hence influence bombing scores.

9. *Variation in instructors.*—Students had instructors of varying ability from one mission to the next. Moreover, the length of training missions, usually over 4 or 5 hours, was known to affect instructor motivation considerably, depending particularly on the interval between missions and on whether a mission occurred in the early morning, later afternoon, or at night. A student's circular error score on any given bomb run probably reflects in part the quality of instruction received during the mission. Certain mistakes in procedure can be corrected by alert instruction after a bomb run and this can result in lower circular errors on subsequent drops.

Student variation in rate of learning.—The reliability of circular error varies as a function of the degree to which learning rate varies for different students. If the rate of learning varies markedly the general tendency is for circular errors to decrease with practice but in varying amounts for different students. Variations in the amount learned between bomb releases will reduce the probability of a stable ranking of the student and this is reflected in a low reliability coefficient.

It should be noted that constant differences among individuals with respect to practice or quality of instruction will result in spuriously high reliabilities. Differences in practice opportunities may take the form either of differing amounts of aerial training prior to the first bombing mission, or of differing numbers of practice missions after bombing starts. The former may introduce a constant error by determining a rank order of individuals at the first testing which, if based upon large differences in CE, will tend to persist throughout training. The effect of the latter would be to build up such differences between students as bombing practice proceeds. Either of these differences will cause consistencies in rank order which are not attributable to differences in bombing ability.

The Victorville data provide some evidence concerning differences among students with respect to practice opportunities. Table 9.5 discloses that at the time students release their first scored practice bomb, their previous experience with the main scope varied from 1½ hours to 19 hours of practice for earlier classes, or from 3 to 27 hours for the later classes. By the time a student has made his tenth scored release, his total practice time can range from 11 to 35 hours for the earlier classes, or from 14½ to 48½ hours for the later classes. Between his first and tenth scored and photographed drops, the average Victorville student obtained an average of about 15 hours bombing practice. However, this varied among individual students from 4 to 25 hours.

It should be remembered, of course, that these data do not constitute continuous records of bombing practice. Releases which could not be photographed or scored because of weather or for other reasons were not recorded. Hence, students come to their first and tenth scored drops with somewhat more practice than is actually shown in the table. There is no reason, however, to doubt the existence of considerable variability of the type described. The results shown in table 9.4, on the other hand, indicate that differences in practice, as represented by the number of drops during the training period, had virtually no influence on the reliability coefficients for odd-even drops. It is unlikely, therefore, that any of the correlations in the various reliability studies reported here were raised spuriously by this factor.

As against differences in practice, differences in the rate of learning during training probably always reduce reliability. If the amount learned by different students varies between successive measurements, the rank order of the students from time to time will change. Since no estimate could be made of the differential learning rates of individuals, the amount of variable error of this nature reflected in the reliability coefficients is unknown.

In a relatively complex task such as radar bombing, variations in learning rate are probably greatest during the early stages of learning. As already noted, the circular error measures of this study represent the student's earliest drops. It is possible the reliability of that circular error would be less affected if it were obtained at a more advanced stage of learning for the group, on a terminal plateau where performance is stabilized and further improvement is not anticipated. As will be seen in a later section, evidence suggests that the average student does not reach such a plateau in the curve of his learning until a considerable number of additional hours of practice beyond the wartime training period.

Improving the Reliability of Average Circular Error

Two types of control are possible over the sources of unreliability: statistical and administrative. The usual aim of the former is, through conversion tables, to correct available circular error data for known sources of variability, such as altitude. Also, average CE may be recomputed to give less weight to releases which seem to reflect extraneous influences beyond the student's control; for example, bomb releases of certain missions may be eliminated, medians rather than means may be used, and releases where the photographed target was not visible may be rescored. At best, two difficulties arise in making statistical corrections. If the conditions are not controlled under which bombs are released, the extraneous influences are likely to be complex and difficult to assess, creating a perplexing problem in devising correction formulae. The second problem is that certain factors such as weather are difficult if not impossible to quantify.

Administrative control of the sources of unreliability is concerned with the conditions under which bombs are dropped and the results are recorded. In this approach modifications are made in the training procedure so as to eliminate the influence of known sources of variation from any single release. Certain pertinent modifications may be suggested which might conceivably increase both the reliability of the circular error and its validity as a measure of bombing ability.

One needed improvement is for recording and scoring procedures which will provide a more continuous record of bombing performance and a larger number of scored releases. Reliability increases as a function of the number of measures upon which it is based. By increasing the number of drops the influence of various sources of unreliability on the course of an individual's performance may be more nearly randomized.

Of the various methods of scoring, scope photography and recording by ground radar appear to be the two most promising from the point of view of increasing the proportion of recorded and scored releases. The first method of scope photography overcomes certain limitations of the present terrain-photo method. In this method, the impact point is determined and scored from scope photographs; the scorer merely takes a fix from the best points visible on the impact scope photo and plots this fix on a chart which contains the target. Since the scope presentations are photographed, weather is minimized as a handicapping factor in scoring releases. With this procedure a continuous record of each bomb run up to the impact point is recorded. The fact that the scope camera is under direct control of the radar observer reduces the likelihood of camera malfunctions or improper camera settings. A further desirable feature is that the area covered by scope photographs is large, a characteristic permitting the ready scoring of extreme misses. A derivative benefit of this method is that it would enable training authorities to insist upon a photographic record of every bombing run for each student. Hitherto, on missions where terrain photos of releases could not be obtained, it is likely that the care taken in bombing practice was less than might have been the case if photographic records were possible. Difficulties with this method are that it requires trained experts to score the picture and detailed charting of local points of the target area so that they may be used in fix-taking at ranges limited to 5 or 10 miles. It is possible that scope interpretation will be easier in the newer sets which have vastly improved definition.

The second development in recording devices, the ground radar method described on page 170, also has advantages over the photo-scoring procedures currently used. Like scope photography, it overcomes weather limitations, obtains an accurate record of extreme misses, and plots a continuous record of the entire bomb run. It may also involve less error in determining impact point than the photo-terrain method although this is as yet undetermined. Extensive use of the method over an entire mission route or over several different routes requires considerable equipment inasmuch as a recording device must be located at every target. The constant radio liaison required between the ground installation and aircraft also presents a problem. Another cumbersome feature is the unwieldy length of the record for a single bomb run, a characteristic which raises a considerable physical problem in processing the detailed record sheets for scoring. It should be noted that scores other than circular error, range, and deflection error are obtainable with the photographic and radar recording procedures just described. Photographs may be used to score objectively a student's computation of heading, track, ground speed, altitude, and bombsight sighting angles. Such measures, if used, would add to the number of variables scored for each run.

The descriptions presented in a previous section of the influence of such factors as weather, condition of equipment, type of aircraft, crew, bombing altitude, target, briefing instruction and instructors upon circular error scores imply requirements for control which need not be enumerated. Basic to effective control of any type, however, is the location of basic training schools in geographical areas where weather makes it possible to delimit the conditions under which missions are conducted. A point frequently made during the war was that the location of schools in areas of relatively constant weather would not provide practice under the varying conditions frequently met in combat. However, during the early stages of learning, constant conditions with respect to weather and other factors may facilitate rather than detract from learning. Radar performance under all types of weather conditions should, without question, be a part of training at a more advanced training stage.

THE LEARNING CURVE FOR BOMBING IN RADAR OBSERVER TRAINING

Circular error data are here applied to two questions of practical importance in radar observer training. What is the relation between bombing practice and circular error? What is the optimum length of the radar observer course?

Relation Between Bombing Practice and Circular Error

In approaching this question, the assumption is made that aptitude for radar bombing is not associated with the amount of practice a student receives. In some air-force training programs it was the policy to give additional practice to poor students. Such a policy would tend to associate a large amount of practice with high circular error and would lower the expected negative relationship between circular error and amount of practice. However, such a system of additional aerial practice for poor students was not followed in radar training and there is, consequently, no evidence that the data of the present study were liable to a systematic bias of this sort.

For this study, the index of bombing practice for Langley Field and Victorville was the total number of bombing runs, whether scored or unscored. At these schools the data included a record of all drops of a mission which were unphotographed or unscored for various reasons. At Boca Raton, the record of the number of unphotographed runs was incomplete; therefore, only the number of scored and photographed runs was used as a measure of practice. For all schools, the average CE based on all scored and photographed drops was the measure of bombing proficiency.

The Learning Curve for Radar Bombing

Table 9.6 gives the results for the three schools. All the correlations are low and in the expected direction but only one is significant at the 5-percent level. The single significant correlation suggests that the most consistent improvement with practice occurs in actual bombing. As was seen previously, however, this type of mission was not typical of the bombing situation generally encountered by the radar observer and was probably simpler than the camera missions. The time devoted to actual bombing was probably sufficient to show learning progress in the simpler problems posed. The remaining correlations suggest, on the other hand, that the camera missions, in which industrial targets were usually bombed, presented complex learning tasks with which little progress can be made on the basis of as little practice as is represented by the recorded data.

TABLE 9.6.—*Relationship between bombing practice as indicated by number of drops and average circular error*

		N	Number of drops		Average CE		
			Mean	SD	Mean	SD	
Boca Raton [1]	Actual bombing	144	9.38	3.20	1,172	634	[2] −0.17
	Camera bombing	110	9.19	1.88	5,150	1,724	−.15
Langley	Camera bombing	179	13.98	6.06	2,936	1,080	−.03
Victorville	Camera bombing	400	26.81	7.26	4,200	1,092	−.07

[1] At Boca Raton, total number of scored photographed runs was taken as the index of practice effect. Cases with fewer than four scored photographed runs were not included because they were not used in the reliability study. This, therefore, slightly curtails practice effect in this sample.
[2] Significant at 5-percent level.

The above conclusions are confirmed in learning curves based upon five Boca Raton student classes. In figure 9.4 is plotted average circular error for the group for both complex and simple targets. The curve for camera bombing is based upon 143 students, while that for actual bombing involves 144 students. For each type of practice the average CE of the entire group was found for each consecutive drop. For some students, photographs either were not taken of a particular drop or were not scorable. Consequently, the total number of scores varies for each drop. The number of students entering the average CE is given on the graph for each drop. A minimum of 10 drops was required before an average CE was plotted.

Inspection of figure 9.4 reveals that in bombing complex targets there is a definite difference (approximately 1,000 feet) between accuracy on the first two drops and accuracy on later drops. Beyond this, however, little learning is evident. In the curve for bombing simple targets, learning occurs through the first four drops (approximately 700 feet) but thereafter the curve quickly decelerates.

In figure 9.5 are plotted the average circular errors of the same group of men for consecutive bombing missions. Thus, all the first

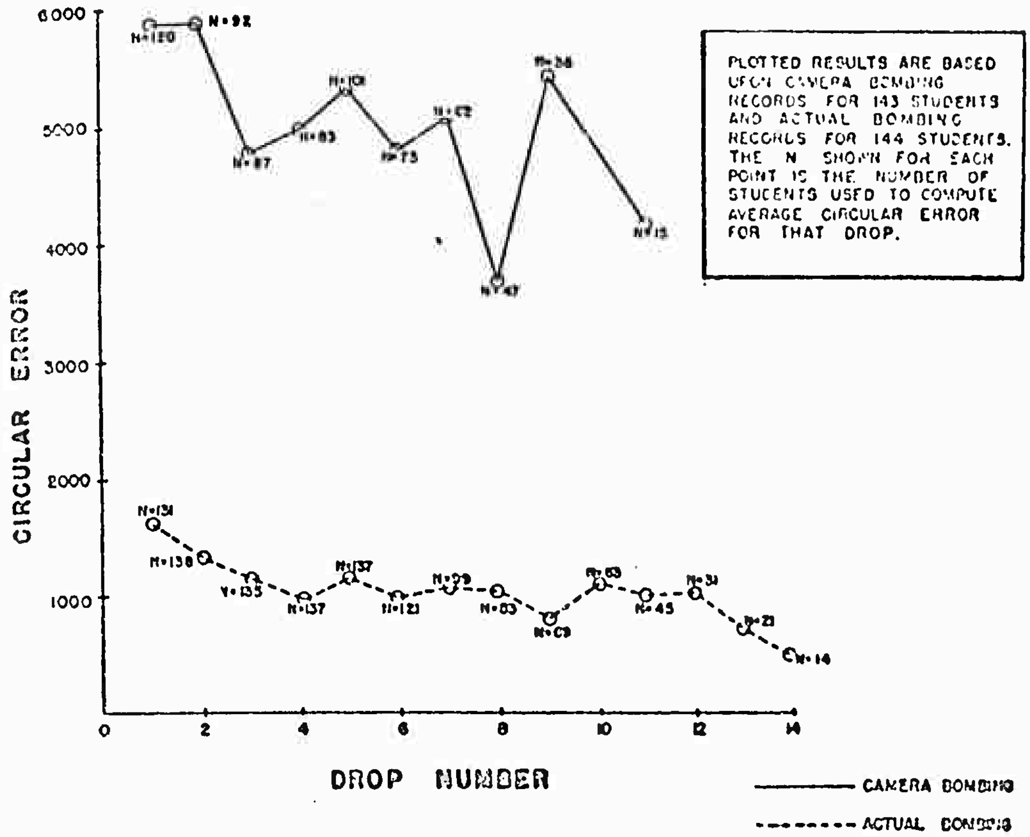


FIGURE 9.4.—Drop-to-drop learning curve for five Boca Raton classes.

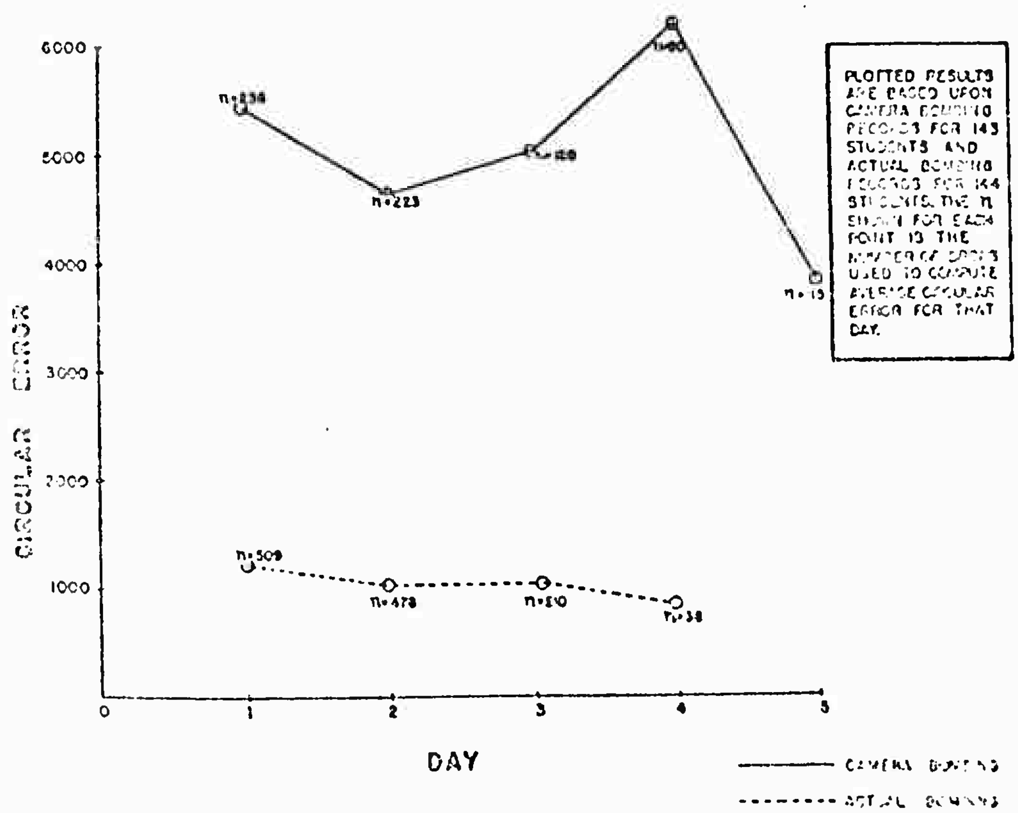


FIGURE 9.5.—Day-to-day learning curve for five Boca Raton classes.

mission runs of all students are combined and an average CE for the day plotted. This was done for all missions for which there was a minimum of 15 scored drops from which to compute an average. It was found that only scattered runs were recorded for a sixth mission. In actual bombing virtually all drops occur in three missions. The number and percentage of drops used to compute average CE for a given day are shown on the graph.

Figure 9.5 discloses very little improvement in the bombing of complex targets over the period plotted. The decrease from the fourth to the fifth mission must be cautiously interpreted because the number of drops comprising the last plotted point is very small. In actual bombing there appears to be slight but continual improvement from the first to the fourth mission.

To summarize the evidence from both sets of curves, the clearest evidence of learning within the period plotted occurs on the first two drops in complex target (camera) bombing and on the first four drops in simple target (actual) bombing. The improvement consists of a decrease in CE of approximately 1,000 feet in camera bombing and about 700 feet in actual bombing. After these early drops no measurable learning is apparent.

However, these conclusions are subject to several qualifications. An equivocating factor in the interpretation of the learning curves is that the circular error scores of the present studies were not corrected for target difficulty. A total of 20 targets of undetermined difficulty were used at Boca Raton. It is unknown whether the difficulty of targets was disregarded in assigning students to routes or whether routes were assigned in such a way that targets became increasingly more difficult during the course. If target difficulty were handled in the latter fashion, its effect in the present data was probably to mask any learning which occurred.

A second qualification relates to the use of only scored, photographed drops. Scored photographed drops provide an unknown portion of the total number of drops made in the entire Boca Raton course and may or may not be representative of all drops. Moreover, the true position of scored drops in the learning process is obscured by the fact that the exact sequence of scored drops in relation to unscored drops is unknown.

The findings presented above suggest that the radar observer training program was of insufficient length to develop a high degree of skill in radar bombing. Additional evidence for this conclusion is presented by an extended training experiment conducted at Victorville by project SC-70, NS-146 of the National Defense Research Committee.[4] One purpose of this study was to determine the amount of aerial

[4] "Final Report on Extended Training Experiment," Headquarters, Victorville Army Air Field, Victorville, Calif., 12 p.

training required for students to achieve the limits of their proficiency in radar bombing. A group of 20 men, 10 bombardier students selected from Boca Raton and 10 navigator students selected from Victorville, were given approximately 150 hours of aerial practice beyond the 30 to 35 hours of training of the standard radar observer curriculum. The group was comprised of men of all levels of ability as determined by a printed proficiency test administered at the completion of the regular course. During the experiment, training was conducted in teams. Each radar observer student flew with the same bombardier for the duration of the experiment and, on a large proportion of his missions, was teamed with the same pilot and copilot. In planning the missions, provision was made for each student to analyze the photographs of his previous bomb runs. The purpose of this procedure was to acquaint each student thoroughly with the sources of his own errors so that he could correct them on later missions.

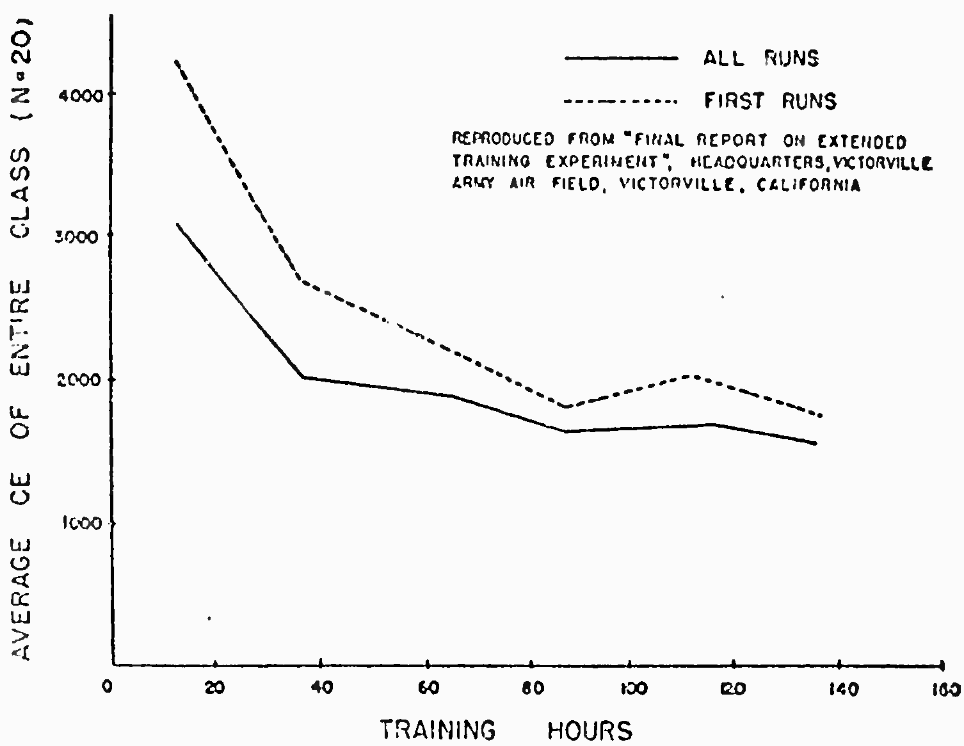


FIGURE 9.6.—NDRC extended training study learning curves.

The group results of this study are shown in figure 9.6. The bombing results of the group were averaged for successive 25-hour periods of training and plotted. The average CE's are plotted at the midpoints of each period. Two average CE's are shown, one for all runs and one for the first runs only. The average CE for all runs comprises every scoreable run in extended training made by the class and includes many repeated runs over the same target. In general, the more runs made on a single target in the same mission, the lower is the circular error for that mission. In order to eliminate the influence of this repetition of runs, the CE's of the first runs only of each mission were averaged. This average represents the student's ability to hit the target on his first attempt.[5]

Figure 9.6 reveals that approximately 85 hours beyond the usual training period were required for the entire class "to reach such a level of proficiency that the circular error on the first run of each mission was approximately equal to the average for the whole mission."[6] As noted earlier, we may assume that the students brought to the experiment approximately 30 to 35 hours of previous main scope training. If this is added to the extended training time, figure 9.6 may be translated to mean that the steepest part of the learning curves comes at about 50 hours of training (20 hours extended training), that learning begins to level off at about 65 hours of training (35 hours extended training), and does not reach its terminal plateau or peak until about 115 hours of training (85 hours of extended training). It must be noted again, however, that the extended training experiment differed from the basic training situation in at least two important respects. Improved opportunities were provided for analysis of one's own bombing errors, and bombing was carried out by relatively constant crew combinations which made it probable that teams rather than individuals were learning. It may be expected that a greater number of training hours would be required to reach the same degree of accuracy under customary training conditions. The conclusion assumes, of course, that other aspects of training are unchanged and that only the amount of aerial practice is varying. The exorbitant price in aerial training time poses the question of how other features of the course, e. g., ground school classes and trainers, may be improved in order to introduce economy in the number of hours required for a student to reach peak efficiency.

Figures 9.5 and 9.6 invite comparison because of the fact that the learning curves of figure 9.6 (extended training experiment) begin approximately at a point in training where the day-to-day curve of figure 9.5 ends, i. e., at about 30 hours of training. The sharp improvement at the early part of the curve for the extended training experiment is in striking contrast to the plateau seen in figure 9.5 for the preceding period of training. As a possible explanation for this difference, attention is again called to the features of the extended training experiment which favor improvement in CE, namely, team learning, and self-analysis of bombing errors. Had these conditions existed in regular training, it is possible that the two sets of curves might have been more continuous.

---

[5] A question arises in this connection regarding the effect of increasing familiarity with targets in the training area upon apparent improvement in bombing accuracy. For a more detailed description of this feature of the experiment, the reader is referred to "Final Report on Extended Training Experiment," *op. cit.*

[6] "Final Report on Extended Training Experiment," *op. cit.*, p. 5.

Navigator students were found in the extended training study to be consistently superior to bombardier students with respect both to learning progress and peak proficiency. The navigators reached near peak proficiency at only 35 hours of additional practice while bombardiers required approximately 85 hours of additional aerial time. Moreover, the difference between the two persisted even after plateaus had been reached. The average CE of navigators at peak performance was 1,585 feet; that of bombardiers was 1,678 feet. The average CE for the entire group was 1,627 feet. It should be mentioned that initial differences in favor of the navigator students might have resulted from the fact that all had been previously trained at Victorville where the study took place, whereas the bombardiers had all been trained at Boca Raton and were unfamiliar with the Victorville routes and targets.

RELATIONSHIP BETWEEN AVERAGE CE AND OTHER PROFICIENCY MEASURES

A matter of some interest from the point of view of the interrelationship of different proficiency measures is the degree to which CE is related to various test and course grades. All such relationships for which data were available were analyzed and are reported in this section. In the first and second studies, an analysis is made of the relationship between CE and course grade, with attention called to the effect of including extreme misses in CE and partialling out the effects of practice on CE. In the third and fourth studies, CE is correlated with proficiency measures presenting the radar observer's specific skills rather than his over-all proficiency. Throughout this section and in tables 9.7, 9.8, 9.9, and 9.10 the signs of correlation coefficients have been reversed so that positive correlations indicate the association of "good" performance in the two variables.

Relation Between Final Course Grade and Average CE, Computed With and Without Extreme Misses

The correlations computed in analyzing the relationship between final course grade and CE when the latter is computed with and without the extreme misses represented by TNV drops, are based upon data from Langley Field camera bombing records. A total of 179 students were available for Langley Field classes 45–30 through 45–36. Of these seven classes, four had entirely completed their bombing prior to VJ-day. For 174 of the 179 students, it was possible to obtain the final course grade assigned by the training station. A description of this course grade is given in chapter 11.

The Langley Field bombing records, partly as a result of the nonmosaic method of photo scoring used at that station, contained an unusual number of TNV drops. In order to examine the effect of these extreme bombing errors on the correlation between CE and other proficiency measures, two computations of average CE's were carried out. One of these was based solely on scored photo drops. The other included, in addition, the TNV drops each of which were assigned a circular error score of 8,000 feet. This figure was chosen as being somewhat higher than the highest scored circular error for this sample.

The effect of including the TNV drops upon the correlation with final course grade is shown in table 9.7 which gives the coefficients based upon both uncorrected and corrected average CE's. The inclusion of the TNV drops, as shown in the coefficient with the corrected average CE, has the effect of raising the correlation with course grade from 0.07 to 0.15, an increase of doubtful statistical significance.

TABLE 9.7.—*Relationship between average circular error and course grade for 174 students, Langley Field*

Variable	M	SD	r with course grade
Uncorrected average CE	2,993	1,092	0.07
Corrected average [1] CE	5,005	1,438	[2] .15
Course grade	40.7	29.3	....

[1] Each drop recorded as TNV (target not visible or not photographed) was given an arbitrary CE of 8,000 feet.
[2] Significant at the 5-percent level.

Relation Between Final Course Grade and CE, With and Without Bombing Practice Partialled Out

The analysis of the relationship between final course grade and CE, with and without bombing practice partialled out, were based on data obtained from three training stations, Boca Raton, Langley Field, and Victorville. The Langley Field data is described in the preceding section. The data from Boca Raton and Victorville are those analyzed earlier in the chapter in connection with the investigation of CE reliability. The correlations for all three schools are presented in table 9. 8. Since the difference among students with respect to bombing experience could conceivably affect both the average CE and final course grades, it appeared advisable to partial out this factor. For Boca Raton, the index of bombing practice was the total number of scored, photographed drops (insufficient information was available concerning unphotographed drops); for Victorville and Langley Field it was the total number of releases, including unphotographed as well as photographed drops.

It is seen from table 9.8 that practice effect has no influence on any of the relationships. Combining the camera bombing results for the three schools by Fisher's Z-method, the correlation between average CE and final course grade is 0.16 (N=670), which is significant at the 1-percent level.[7] The combined r of 0.16 is based on uncorrected

[7] The assumption is made in combining these results that the three samples are independent random samples from the same general population of radar observer students.

average CE's. It may be concluded that a low but stable relation exists between the final proficiency grades of each school and bombing skill as measured by average circular error. The higher a student's course grade, the lower the average CE he is likely to have.

TABLE 9.8.—*Correlations between average circular error and final course grade with and without bombing practice partialled out*

Training station	Type of bombing practice	N	r_{12} [1]	r_{13} [1]	r_{23} [1]	$r_{12.3}$ [1]
Boca Raton	Actual	141	0.02	0.17	0.08	0.03
Boca Raton	Camera	110	[2] .22	.15	.03	[2] .22
Langley Field	Camera	174	.07	.03	.03	.07
Victorville	Camera	395	[3] .18	.07	.03	[3] .18
All schools	Camera	679	[4] .16			

[1] 1 = average circular error.
2 = final course grade assigned by training station.
3 = number of bombing runs.
[2] Significant at the 5-percent level.
[3] Significant at the 1-percent level.
[4] Consists of the camera bombing coefficients for Boca Raton, Langley, and Victorville combined using Fisher's z-transformation method. This coefficient is significant at the 1-percent level.

It is interesting to note in table 9.8 that camera bombing appears to be more closely related to proficiency as measured by final course grade than is actual bombing. This difference is consistent with a point made earlier that actual bombing on a pin-point target which is easily identifiable involves fewer of the skills taught in the course than does camera bombing where industrial areas provide very complex targets.

Relation Between Phase Grades and Circular Error

In this analysis the average CE for Boca Raton and Langley Field students were correlated with phase grades computed by the Psychological Research Project (Radar) and based upon standardized proficiency measures. Three such grades were computed; a flight stanine based upon aerial performance check scores, a trainer stanine based upon bench set and supersonic performance check scores, and a classroom stanine based upon proficiency test scores. These tests and performance checks are described in chapters 5 and 6. The computation of the three phase grade stainines is described in chapter 11.

The Boca Raton students are drawn from the sample used in the analyses of CE reliability. The Langley Field students are those described in the two studies just represented. In computing CE for the Langley Field students, unscored extreme misses were included. Of the 179 Langley Field students for whom bombing records were available, phase grades could be obtained for only 97.

Table 9.9 shows the correlations obtained. The correlation of CE with course grades, based upon standardized proficiency measures, is included for comparative purposes. The coefficients in the table suggest again that actual bombing CE is less related to other proficiency measures than is camera bombing CE. For the course grade stanine, the flight stanine, and the trainer stanine the combined Boca Raton and Langley Field samples provide statistically significant correlations. The correlation with course grade stanine is significant at the 1-percent level, while those with flight and trainer stanines are significant at the 5 percent level. The correlation with the classroom stanine does not reach statistical significance.

TABLE 9.9.—*Correlations of average circular error with course grades and proficiency stanines computed by the radar project for Langley Field and Boca Raton students*

Training station	Type of bombing	N	Correlations			
			Course grade	Flight stanine	Trainer stanine	Classroom stanine
Boca Raton	Actual bombing	82	0.04	0.07	0.02	0.08
Boca Raton	Camera bombing	81	[1] .20	.14	.17	.06
Langley Field [1]	Camera bombing	97	[1,2] .19	.17	.12	.10
Combined Boca Raton and Langley Field.[3]	Camera bombing	178	[4] .20	[2] .15	[2] .14	.08

[1] Langley correlations based on corrected average CE (TNV's counted in average; each 8,000 feet).
[2] Significant at the 5-percent level.
[3] Combined, using Fisher's z-transformation method.
[4] Significant at the 1-percent level.

Relation Between Single Proficiency Measures and CE

The phase grade stanines described above were not available for Victorville students. However, scores for this group could be obtained on specific standardized proficiency tests and performance checks. Four such measures were correlated with average CE: they are the Final Aerial Performance Check, Radar Final Test I and II, Radar Navigation Intermediate Test, and Radar Bombing Intermediate Test. The number of students for whom scores could be obtained varied for the different measures.

The correlations obtained are shown in table 9.10. With the exception of Part II, Radar Final Proficiency Test, the correlations of all tests with CE attained statistical significance at or above the 5-percent level. The coefficients for the different tests are approximately equal in size. That obtained for the aerial performance check is also

TABLE 9.10.—*Correlations of average circular error from camera bombing with standardized proficiency measures for Victorville students*

Proficiency measures	N	r
Final Aerial Performance Check	99	0.18
Radar Final Proficiency Test:		
Part I	385	[1] .14
Part II	385	.06
Total score	385	[1] .15
Radar Navigation Intermediate Test	277	[1] .15
Radar Bombing Intermediate Test	373	[1] .14

[1] Significant at the 1-percent level.
[2] Significant at the 5-percent level.

of the same order of magnitude, but since it is based upon a smaller N, its stability is uncertain.

SYSTEMATIC ERROR IN RADAR BOMBING TRAINING: THE RANGE OVER ERROR

Ordinarily, one would expect that among experienced radar observers the directional errors in bombing would be unsystematic over a period of time, that errors would occur as frequently in one direction as in another. In basic radar observer training this was not the case. Table 9.11 gives the average amount and direction of the range error of camera bombing at each of the three schools. The direction of the error is consistently over or beyond the target. The data are taken from the population samples used in the other circular error studies reported in this chapter. Results comparable to those in table 9.11 were found for the first 25 hours of extended training in the NDRC extended training experiment. For 416 runs during this period an average range over error of 1,074 feet is reported. In addition to this data, 108 bombing runs scored at Langley Field by the radar method (SCR 584) yielded an over error of 1,420 feet. This finding minimizes the possibility that the directional error is a function of the photo-terrain method of scoring.

TABLE 9.11.—*Constant range error for camera bombing in radar observer training*

School	Number of scored runs	Range error
		Feet over
Langley Field	2,093	1,094
Boca Raton	758	1,107
Victorville	1,403	866

While the possible causes for this error are clearly not limited to misuse of his equipment by the radar observer, an exploratory analysis of these causes was undertaken. It was planned that this analysis be followed by experiments which might reveal the responsible factors. Although training was terminated before such experiments could be conducted, the hypotheses developed are of sufficient interest to warrant presentation. Discussion is divided into causes arising from equipment sources and those arising from psychological sources.

Equipment Sources

Trail error in photo-scoring of direct bombing runs.—Trail is the distance behind the aircraft at which a given bomb will strike. The bomb lags behind the aircraft because air resistance tends to overcome its forward speed, as shown in figure 9.7. In making an actual bomb release, compensation is made for this lag. In the calibration of the radar computer drum ground speed lines, which are used in direct bombing for determining the correct bomb release circle, this compensation is accomplished by inclusion of an average trail value.

In figure 9.7 it is seen that the line AB represents the axis of the vertical camera at the time of the impact picture; hence point B will be the center of the picture. The camera takes no account of trail since it photographs the impact point or the point directly under the aircraft. But since, as noted above, the radar computer is calibrated in such a way as to allow for trail, the aircraft will have traveled beyond the aiming point at the time of impact, and the photograph taken at this point will be in error in the over direction. It becomes apparent that the impact point of any simulated bomb release in direct bombing

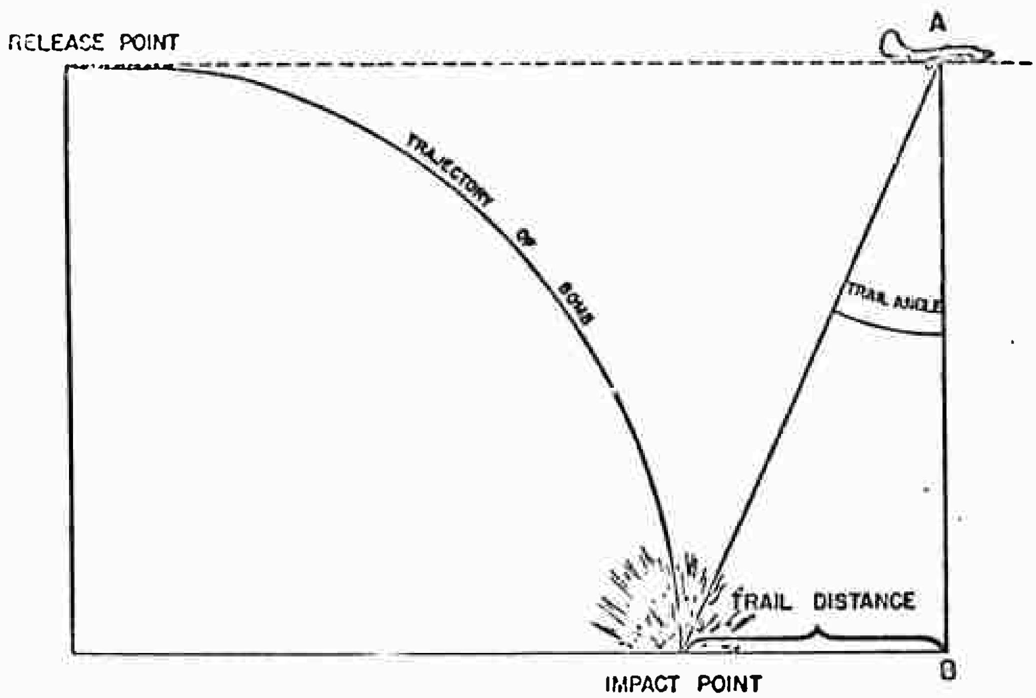


FIGURE 9.7.—Diagrammatic sketch of aircraft and bomb at instant of impact.

will be plotted trail distance too far over, unless this value can also be taken into account in scoring the impact photograph.

It has not been possible to determine the exact trail value included in the computer, but it is known to be approximately 60 mils at 12,400 feet, the average bombing altitude at Langley Field. Scored as zero trail, 60 mils of trail would amount to an over error of 744 feet. The smaller range error obtained at Victorville (see table 9.12 below) may possibly be accounted for by the fact that no direct bombing was done at that school.

Trail error in coordinated bombing.—In contrast to direct bombing, the bomb releases of coordinated bombing practice were not corrected for trail so as to make the impact point photograph correct. This required that the trail arm of the bombsight (index for setting proper trail value in mils) be set at zero. If a student neglected to do this and either set trail or left trail in the bombsight, an over error would result on the impact point photograph at an average rate of 12.4 feet per mil of trail. A similar error was that of preset trail. Preset trail is an internal misadjustment in the bombsight resulting in an actual trail value differing from the reading set by the trail arm. If preset trail is present, it is usually positive. It may be that students were not consistently briefed to preflight the bombsight for this malfunction.

Failure to use sector scan.—The time between renewals by the sweep of the target return on the scope may be a factor contributing to error. On normal rotation the aiming point is renewed every 2.5 seconds. Assuming a ground speed of 204 miles per hour, the aiming point would be approximately 750 feet nearer the center of the scope the next time the sweep passed. With the 30° sector scan the target return is renewed every 0.83 second with AN/APQ-13 and every 0.12 second with AN/APS-15 and AN/APS-15A. With 30° sector scan it would be only 126 or 249 feet nearer, depending upon the equipment used. Since most radar observer students would not release bombs before they saw the aiming point touch the bomb release circle, failure to use sector scan could result in over error. The movement of the target on a full revolution of a 360° scan is illustrated in figure 9.8.

Psychological Sources

Uncertainty delay.—When a student loses or is uncertain of his aiming point he may have to take too long for definite identification. At a ground speed of 204 miles per hour the effect of uncertainty delay upon direct bombing range error and displacement error in coordinated bombing is to create an over error at the rate of about 200 feet per second. In coordinated bombing, a delayed final sighting angle signal will set up a slow rate in the bombsight which also results in an over error.

Target difficulty.—Probably uncertainty delay increases as a function of target difficulty. In this case, the more difficult the target, the larger the over error. If it were possible to obtain a suitable criterion of target difficulty a test of this hypothesis could be made.

Delay in starting the camera intervalometer.—In either direct or coordinated bombing, delay in starting the camera intervalometer should account for very little error since only simple reaction time is involved. However, in many training aircraft, the intervalometer was out of the bombardier's reach and some delay in starting it resulted.

Failure to use leading edge of target and leading edge of bomb circle.—It was frequently reported at radar observer training stations that aerial instructors were advising procedures other than the use of the coincidence of the leading edges of bombing circle and aiming point for sighting bombs and calling sighting angles to the bombardier. Moreover, there was a tendency for bombardier students to alter the procedure by waiting until the bombing circle was over the center of the aiming point. Because of special characteristics of the radar equipment, maximum accuracy can be achieved only when the standard procedure is used. Using the middle of the bombing circle will

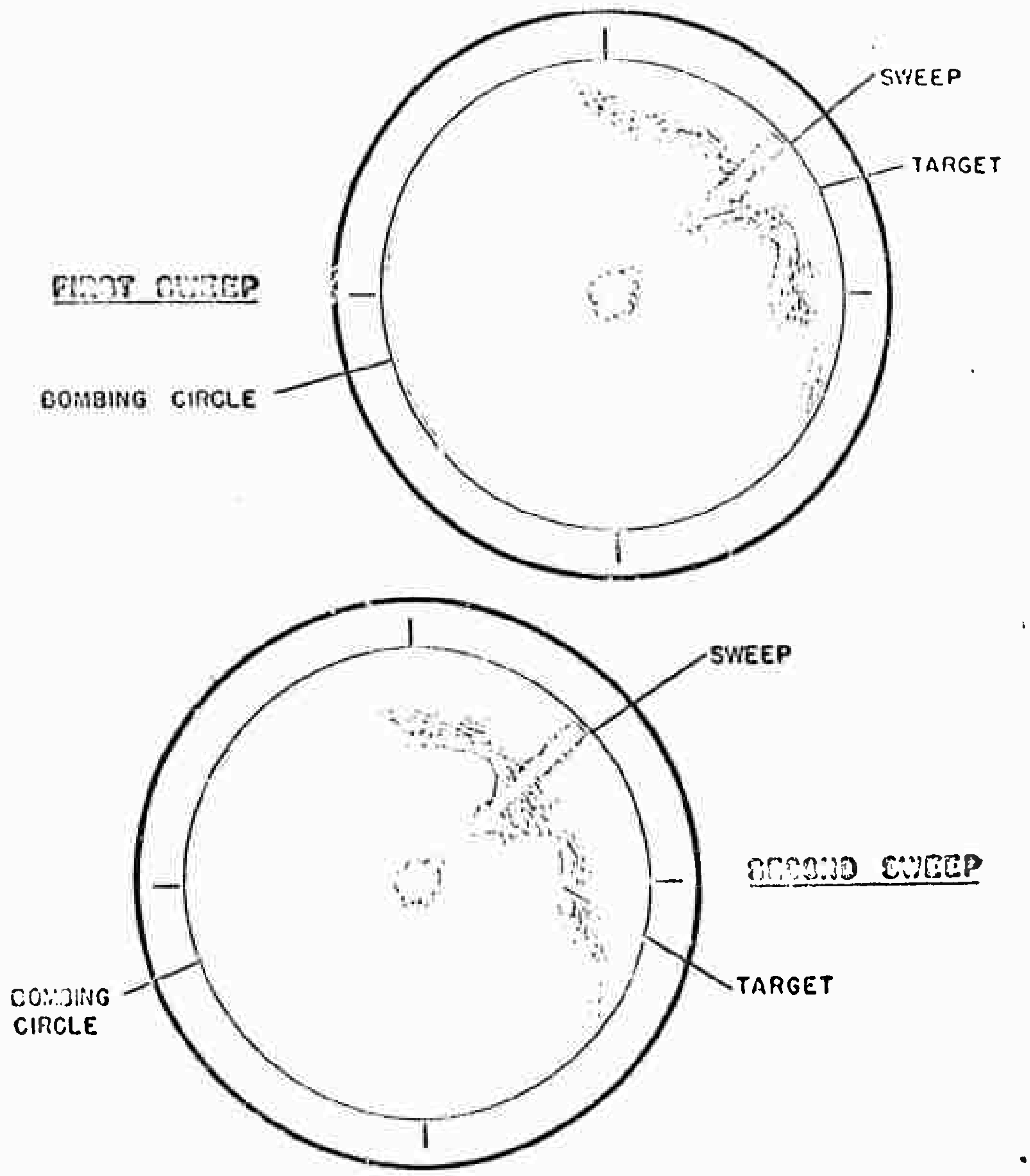


FIGURE 9.8.—Apparent movement of the target during one full revolution of the sweep

result in an over error of 250 feet (since the width of the circle on bombing range represents a ground distance of 500 feet).

Synchronization lag.—In a coordinated bomb run the bombardier is required to synchronize the rate mechanism of the bombsight (which in part controls the range of the bomb to be dropped) with the speed of approach to the target, as given by the radar operator who calls "mark" as the target reaches successive sighting angles. The bombardier does this by adjusting two knobs: when the radar operator calls a given angle, the bombardier moves the displacement knob so as to place the sighting angle index at that angle, and also the rate knob to correct the acceleration rate of the sighting angle index.

Because of the series of perceptual delays and the reaction time involved in the process of synchronization (particularly at the final sighting angle), it is apparent that, while in most cases an approximately correct rate may be established, the successive displacement settings will be behind the proper value. At an altitude of 12,400 feet, this results in an error of approximately 117 feet per degree of lag.

Controlling the Range Over Error

It is believed that the sources of error listed in the foregoing section, particularly those involving scope interpretation, are readily susceptible to elimination or reduction by additional instruction and practice. This conclusion is borne out in the results of the Extended Training Experiment conducted by NDRC which are presented in table 9.12.

Table 9.12 discloses that the range error decreases progressively up to 100 hours of training. Also, after 25 hours of extended training, the directional error is as likely to be short as it is over. Two factors may account for the decrease in over error after the first 25 hours of extended training: (*a*) The beginning student exhibits deficiencies in set operation, navigation and bombing procedure, or scope interpretation, which cause over error and are eliminated by further practice and instruction; (*b*) students were allowed to analyze their bombing errors over a considerable period of time, thus becoming increasingly aware of the systematic nature of their range error. This would aid them in developing compensating techniques which would tend to place the point of impact nearer the aiming point.

TABLE 9.12.—*Size and direction of range error for successive 25-hour periods of extended training*[1]

Period	Number of scored runs	Range error
0–25 hours	116	[2] 1,071
25–50 hours	454	[2] 131
50–75 hours	220	[3] 64
75–100 hours	278	[3] 77
100–125 hours	402	[2] 373
125–150 hours	229	[3] 138

[1] From Final Report on Extended Training Experiment, 11 July 1945, Headquarters, Victorville Army Air Field, Victorville, Calif., p. 6.
[2] Feet over.
[3] Feet short.

SUMMARY

In this chapter are discussed five topics related to bombing errors in radar observer training: methods of recording and scoring bombing circular error, the reliability of circular error, the learning curve for circular error, the relationship between circular error and training proficiency measures, and the systematic over error in range.

Recording and scoring procedures were adapted to score two types of bombing practice: simulated or camera bombing, the most frequent type, and actual bombing. The most commonly used recording and scoring system in the wartime training program was photo bomb scoring. In this technique, actual bomb drops are scored by taking a terrain photograph at the moment of bomb impact. To score simulated bomb drops, a series of vertical terrain photos are taken during the bomb run. From these photos the impact point of the simulated release is plotted and the amount of error scored. Scoring is facilitated by the use of a mosaic photograph which, since it encompasses a wider area than any single photograph, allows the measurement of larger errors.

The principal limitations of photo-terrain scoring are that weather severely restricts its use, that turbulence will frequently upset the vertical alignment of cameras which are not gyro-stabilized and that extreme bombing misses cannot be scored. These limitations result in an incomplete record of the student's bombing performance and contribute to unreliability of the circular error measure.

Simulated bomb runs may be recorded also by means of a gun-laying radar device. This method, which was developed near the end of wartime training, not only overcomes the weather handicap of terrain photography, but also makes it possible to score extreme misses. However, it presents difficulties in the amounts of equipment required, the radio liaison required between the aircraft and the ground radar installation, and the unwieldy length of the record for each bomb run.

Bomb releases are usually scored in terms of circular error, which is converted to a standard 12,000 feet altitude. However, additional components of the bombing run, such as heading, track, ground speed, and altitude, may be objectively scored. Measures such as these are as yet unexplored as indices of bombing performance.

An analysis was made of the reliability of circular error in order to evaluate its usefulness as a validation criterion. Circular error scores in training are affected by numerous influences unrelated to bombing ability. After considering the problem of the differing results yielded by several methods of computing reliability, it was pointed out that the most useful reliability coefficient was one which summarized the effect on circular error scores of all factors that cause a student's rank to vary from one day to another. Three methods of determining reliability were evaluated according to this standard: the correlation of odd and even drops over a single mission, the correlation of odd and even drops over several missions, and the correlation of drops from odd with those from even missions. It was concluded that the first two methods exaggerate the reliability of circular error because they fail in whole or in part to measure the influence of mission to mission variations. It was further concluded that the odd-even missions method provides the most accurate estimate of reliability since it takes into account all of the known mission-to-mission as well as drop-to-drop influences.

Application of the method of odd-even drops over several missions to available data for the Boca Raton school yielded coefficients of 0.46 for actual bombing (N=144) and 0.35 for camera bombing (N=110). When corrected for length by the Spearman-Brown prophecy formula, these coefficients rose to 0.63 and 0.52, respectively. Because practice (number of scored drops) was permitted to vary in the Boca Raton sample, its influence was partialled from the coefficients, and was found to be negligible in both actual and camera bombing. The higher odd-even drop reliability found for actual bombing was attributed to several causes, among which are its concentration on a few missions, the simple target used, and relative objectivity of scoring.

On data available for the Victorville training station, odd-even drop reliability was determined by two methods: (a) correlation of mean CE's of odd and even drops, and (b) correlation of median CE's of odd and even drops. The latter method supposes that large CE's reflect irrelevant factors in the testing situation and that these factors unduly affect mean CE. Consequently, median CE, which is less affected, may reflect to a greater extent the bombing ability of the student. Uncorrected r's obtained with the method employing means were 0.31 and 0.32 for two samples of 134 and 268 cases, respectively, as compared with 0.27 and 0.16 for the same samples with the odd-even median method. This result suggests that extreme scores generally are no less indicators of true bombing ability than are the lower circular error scores, and appears also to emphasize the importance of scoring accurately extreme bombing errors and including them in the average CE.

Extreme bombing misses, ordinarily not scoreable by the photo-terrain technique, were given an arbitrary score in one group of the Victorville sample. Inclusion of these drops had a favorable effect on odd-even mean drop reliability, the uncorrected coefficient for this sample being 0.39 (N=134) as compared to an r of 0.31 found when the extreme scores were not included.

Odd-even mission reliability was studied in two samples. For Boca Raton data, from camera bombing only, the obtained coefficient was 0.19 (N=122). This is increased to 0.31 when corrected by the Spearman-Brown formula. Odd-even mission reliability for the Victorville data was found to be 0.11 uncorrected (N=372) and 0.20 when corrected. From the odd-even mission coefficients obtained at the two stations, it may be concluded that the reliability of circular error in radar observer training is probably between 0.20 and 0.30. This reliability is thought to be sufficient to encourage further work to improve circular error as a criterion measure.

The odd-even mission coefficients obtained in this study appear to be somewhat higher than those found in comparable studies of visual bombing. A possible explanation of this difference is that there is less variability among students in visual bombing proficiency than in radar bombing. Consequently, the differences between students in visual bombing are more easily obscured by variations in factors unrelated to bombing ability.

An exploratory attempt was made in this chapter to describe certain sources of unreliability of circular error and to suggest means for reducing their influence. Among the variations in bombing conditions which probably affected scores were variations in (*a*) target difficulty and familiarity, (*b*) number of missions over which drops are spread, (*c*) bombing altitude, (*d*) weather conditions, (*e*) radar equipment, (*f*) crew efficiency, (*g*) working conditions and auxiliary instruments in aircraft, (*h*) briefing procedure, and (*i*) instructor. The effect of variations in practice and in learning rate was also discussed. Radar scope photography and recording by ground radar appear to be the two scoring developments which offer most promise of increasing reliability.

In a study of the relationship between number of bombing runs and average CE at three radar schools, correlations near zero were obtained. This suggestion that the amount of improvement in circular error with practice was negligible, was supported by an analysis of circular error learning curves. Group learning curves based on five classes at Boca Raton disclosed that improvement occurred only on the first two camera drops (decrease of approximately 1,000 feet in circular error) and on the first four actual bombing drops (decrease of approximately 700 feet in circular error). After these early drops, the evidence for future learning was ambiguous. On actual bombing missions there appeared to be some improvement from the first to the fourth mission, but camera bombing failed to show a similar trend. If it is assumed that the plotted data were representative of progress made during the training period, it may be concluded that 30 to 35 hours of main scope aerial training is insufficient to yield much improvement in bombing accuracy.

That 30 to 35 hours is inadequate to develop highly skilled radar observers is supported by the results of an extended training study of bombing progress conducted by NDRC. In this study the 20 participating students were provided an opportunity to analyze their own bombing errors and were teamed with other crew members for the major portion of the experiment. The study revealed that, under these conditions of training, the steepest part of the learning curve for the group came at about 50 hours of aerial training (20 hours of extended training), that learning began to level off at about 65 hours of practice (35 hours of extended training), and reached its terminal plateau or peak at about 115 hours of training (85 hours of extended training). It is likely that team training and self-analysis of bombing errors, facilitate learning. Since these were conditions special to the experiment, it is probable that the number of aerial hours required to reach equivalent proficiency would be greater under wartime training conditions.

Studies were made of the extent to which bombing performance may be predicted from classroom, trainer, and flight grades at three radar schools. A low but stable relation was found to exist between the final proficiency grades of each school and average CE for camera bombing. The combined r for three schools was 0.16, with a total N of 670. Actual bombing was found to be virtually unrelated to final course grade, a finding consistent with the fact that actual bombing involved fewer of the skills taught in the course than did camera bombing.

Average CE was also correlated with proficiency measures prepared by the Radar Project. The relationship between camera bombing CE and a final course stanine based on the entire battery of tests and checks was low but significant; the r for the combined Boca Raton and Langley Field samples was 0.20, N=178. For a stanine based upon aerial performance checks and one based upon trainer performance checks, the correlation with CE was somewhat higher than for a stanine based upon several printed proficiency tests. The r's between actual bombing CE and these measures are insignificant. In studies of the correlation between CE and individual tests, coefficients of the order of magnitude of 0.15 were obtained. A single performance check furnished a similar result.

In the earlier stages of radar observer training, the range error is systematic and in the over direction. Probably the major psychological factor contributing to this directional error is an uncertainty delay, which is the result of poor scope interpretation, target difficulty, and perhaps inefficient set operation. Secondary psychological factors apprear to be (a) failure to use the leading edges of both target and bombing circle, (b) synchronization lag, and (c) delay in starting the camera intervalometer. It is likely that these errors decrease with practice and instruction.

CHAPTER 10

History of Radar Observer Selection[1]

A history of the selection of radar observers involves four psychological organizations: Project SC–70, NS–146 of the National Defense Research Committee, which is referred to hereafter as the NDRC Project; the AAF Aircrew Evaluation and Research Detachment No. 1, referred to as AERD No. 1; the Psychological Research Project (Navigator), referred to as the Navigator Project; and Headquarters, AAF Training Command. This chapter reviews the contributions of each of these organizations, points out their interrelations and briefly summarizes the data presented. Some of the research described deals with operators of air-to-surface-vessel radar equipment, referred to as ASV. The work of ASV operators is roughly comparable to that of radar observers and has been included primarily because of its importance to later developments in radar-observer research.

SELECTION RESEARCH BY PROJECT SC–70, NS–146, NATIONAL DEFENSE RESEARCH COMMITTEE

The selection of radar observers originated with a research contract recommended to the National Defense Research Committee by the National Research Council. The NDRC project established for this work began its investigations at Camp Murphy, Fla. in February 1943. Later, it moved to Boca Raton Army Air Field, Fla. The facilities of the project were made available to both the Army and Navy, and emphasis in research was placed upon specific problems recommended to it by one or the other of these services.

One of the initial undertakings of the NDRC project was the construction of tests for selection of ASV radar operators. These operators used the SCR717A and SCR717B radar sets, which were similar to those later used by radar observers. A job analysis revealed a number of specific tasks involved in satisfactory ASV operation, such as the "ability to detect pips on an oscilloscope screen, to estimate relative pip heights, to set accurately a pip to a hairline or gate, to convert

[1] Written by S/Sgt. Harold F. Kunzman.

information from one form to another quickly, and to read and interpret accurately certain symbols and relationships involved on two-dimensional plots."[2] It was concluded that the measurement of these abilities would require tests which would determine the potential radar operator's speed and accuracy of perceptual discrimination, his alertness and persistence, and his capacity for making quick judgments.

Tests

On the basis of the results of this job analysis, along with related studies and observations of the radar operator and his equipment, the NDRC project prepared 13 printed tests and administered them experimentally at various operating stations under the AAF Tactical Center, Orlando, Fla., at Camp Murphy, Drew Field, and at the Naval Training School, Virginia Beach, Va. A brief description of each of the tests follows:[3]

Form Detection Test.—This test is designed to measure speed and accuracy in matching identical forms. It consists of 42 lines of 12 irregular forms, only 2 of which are the same size and shape. The task is to select from each line the two forms that are identical. Time limit, 8 minutes.

Form Conversion Speed Test I.—Twelve different irregular forms are presented, coded to represent a letter of the alphabet from A through L. The forms are grouped to constitute short words which must be decoded by the use of the figure-letter code key. Time limit, 3 minutes.

Form Conversion Speed Test II.—This is the same as Form Conversion Speed Test I, except that Arabic numerals from zero to nine are substituted for letters of the alphabet, and a different set of symbols are used. Time limit, 3 minutes.

Scale Reading Test.—In each of 48 items there is presented a meter with an indicator set at some point on the meter scale. The task is to determine the exact scale reading. A description of a machine-scored version of this test, Scale Reading, CPG37A, is included in appendix A. Time limit, 18 minutes.

Oscilloscope Reading Test.—This test contains 40 sets of 6 circles. In each circle there is a pattern of vertical lines of differing heights drawn on a horizontal base line. The task is to determine, in each set of six, which 2 patterns are exactly alike. Time limit, 10 minutes.

Spot Location Test.—This test consists of one large circle blocked off into lettered segments, and nine small circles each containing

[2] Selection and training of oscilloscope operators, Memorandum No. 3, Instruction Manual for Oscilloscope Operator Tests, Experimental Edition, NDRC Project SC-70, NS-146, Camp Murphy, Fla.

[3] For a complete description see: Selection and training of oscilloscope operators, Memoranda Nos. 2 and 3, Instruction Manual for Oscilloscope Operator Tests, Experimental Editions, NDRC Project SC-70, NS-146, Camp Murphy, Fla.

several dots. The task is to determine into which lettered segment of the large circle each of the dots in the small circle would fall if the latter circle were as large as the former. A description of a machine-scored version of this test, Spot Location Test, CP818A, is included in appendix A. Time limit, 4 minutes.

Course Location Test.—In this test there are 10 pairs of circles. One circle of each pair contains a pattern of connected lines which is to be duplicated in the other circle. The latter circle contains only dots upon some of which the duplicated pattern must fall. Time limit: 5 minutes.

Ratio Estimation Test.—Ninety-six pairs of two vertical lines are presented, each pair in a rectangular box. The task is to determine the ratio of the left hand line to the right hand line. The lines are spaced to simulate double pipping of some radar sets. A description of a machine-scored version of this test, Ratio Estimation Test, CP225A, is included in appendix A. Time limit: 5 minutes.

Target Course Analysis Test.—This test consists of five groups of small circles, each group arranged to form a rectangle. The circles at the top are numbered from 1 to 10. From each of the top circles a line originates, criss-crosses the rectangular field in random pattern, and ends at one of the blank circles along the bottom or sides of the rectangle. The task is to follow by eye the line from the numbered circle at the top to its end at one of the side or bottom circles, and then to place in the end circle the number of the circle from which the line originated. Time limit: 15 minutes.

Plot Reading Test.—A large square is blocked off into 25 equal parts by the use of perpendicular lines. Each side of the square is marked off equally into 50 numbered units. The task is to locate each of 50 points found within the square by giving the numbers for the pair of lines that describe its location. Time limit: 10 minutes.

Coordinate Plotting Test.—Around the outer edge of a large circle, degrees are marked off in intervals of five. From the center of the circle to the outer edge, nine concentric circles are drawn, with the distance between each circle representing 5 miles. Fifty pairs of numbers are presented, one number indicating degrees around the circle, the other, miles from the center of the circle. The task is to locate the point in the circle which is described by the pair of numbers, by placing a heavy dot at the correct spot. Time limit: 10 minutes.

Coordinate Reading Test.—This is the reverse of the Coordinate Plotting Test. Points are already marked inside the circle, and the task is to locate the point by giving a pair of numbers, one representing degrees around the circle, the other, miles from the center of the circle. A description of a machine-scored version of this test, Coordinate Reading Test, CP221A, B, is included in appendix A. Time limit: 15 minutes.

Polar Grid Coordinate Test.—The scaled circle described in the two preceding tests is presented with grid coordinate lines superimposed upon it. The test involves two tasks. The first is the same as the Coordinate Plotting Test: i. e., locating a point within a circle described by a pair of numbers. The second task is to determine the grid coordinates of this point. A description of a machine-scored version of this test, Polar Grid Coordinate Test, CP819B, in included in appendix A. Time limit: 18 minutes.

During the summer of 1943, three additional tests were constructed. These were: Air-borne Oscilloscope Reading Test, Oscilloscope Conversion Test, and Oscilloscope Interpretation Test. "The first two of the above tests are concerned with air-borne radar operator functions, especially the interpretation and transmission of symbolic directions from operator to pilot. The Oscilloscope Interpretation Test is a test of ability to detect different types of signals through varying degrees of background masking or jamming and is applicable to all types of radar presentations."[4] A description of a machine-scored version of this test, Oscilloscope Interpretation, CP817A, is included in appendix A.

Following experimental administration of the first 13 tests developed, the results were validated against officers' ratings of enlisted radar operators at 8 ground radar centers under the Tactical Center, Orlando, Fla. Since all studies of the validity of this group of tests were carried out on ground radar personnel, validity statistics are not included in this review.[5]

Following is a matrix of intercorrelations of 12 of the first 13 tests described above; the exception is the Course Location Test. The Army General Classification Test, referred to as the AGCT, is included in this matrix.

TABLE 10.1.—*Matrix of intercorrelations of selected radar operator aptitude tests*[1]

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Form Detection	.....	0.36	0.49	0.31	0.47	0.38	0.19	0.28	0.34	0.30	0.21	0.32	0.22
2. Form Conversion I	0.36	.....	.51	.45	.56	.41	.21	.35	.37	.47	.40	.55	.39
3. Form Conversion II	.49	.53	.....	.40	.41	.10	.22	.31	.34	.28	.32	.40	.45
4. Scale Reading	.31	.45	.40	.....	.41	.38	.29	.55	.41	.57	.58	.59	.63
5. Oscilloscope Reading	.47	.56	.41	.41	.....	.37	.23	.31	.35	.16	.30	.50	.30
6. Spot Location	.38	.41	.10	.38	.37	.....	.32	.38	.40	.48	.50	.52	.20
7. Target Course Analysis	.19	.21	.22	.29	.23	.32	.....	.21	.26	.30	.36	.35	.19
8. Plot Reading	.28	.35	.31	.55	.31	.18	.21	.....	.11	.50	.52	.59	.42
9. Ratio Estimation	.34	.37	.31	.41	.35	.40	.26	.11	.....	.50	.43	.47	.47
10. Coordinate Reading	.30	.47	.28	.57	.16	.18	.30	.50	.50	.....	.67	.59	.45
11. Coordinate Plotting	.21	.40	.32	.58	.30	.50	.36	.52	.43	.67	.....	.74	.43
12. Polar-Grid Coordinate	.32	.55	.40	.59	.50	.52	.35	.59	.47	.59	.74	.....	.47
13. A. G. C. T	.22	.39	.45	.63	.30	.20	.19	.42	.47	.45	.43	.47	.....

[1] From Memorandum No. 4, Preliminary Report of Results from Oscilloscope Operator Tests, Experimental Edition, NDRC Project, SC-70, NS-136, Southern Signal Corps School, Camp Murphy, Fla. N=100, selected randomly from 578 radar operators-in-training at Drew Field.

[4] Final Report in Summary of Work on the Selection and Training of Radar Operators, OSRD Report No. 5766, 24 Sept. 1945.

[5] A Validation Study of Oscilloscope Operator, OSRD Report No. 3712, NDRC Project SC-70, NS-146, Research Report No. 10. This report contains validation studies.

Table 10.2 presents the rotated factor loadings of these tests, based on the matrix of intercorrelations in table 10.1.

TABLE 10.2.—*Rotated factor loadings of selected radar operator aptitude tests* [1]

	I	II	III	IV	V	h^2
1. Form detection	65	15	01	11	18	49
2. Form conversion I	58	18	21	14	12	64
3. Form conversion II	50	19	28	42	−03	54
4. Scale reading	18	43	15	52	31	61
5. Oscilloscope reading	64	34	10	10	21	54
6. Spot location	38	02	32	21	49	53
7. Target course analysis	16	03	40	19	20	26
8. Plot reading	17	18	15	43	51	53
9. Ratio estimation	29	22	03	39	38	43
10. Coordinate reading	21	41	14	27	57	63
11. Coordinate plotting	01	48	48	29	50	79
12. Polar grid coordinate	24	46	30	28	47	72
13. A. G. C. T	11	46	03	72	03	74

[1] This analysis was carried out in the Psychological Research Project (Radar) by Capt. Lloyd G. Humphreys while on temporary duty assignment from PRU, SAACC, San Antonio, Tex.

NOTE.—I = perceptual speed; II = visual memory; III = spatial relations; IV = numerical facility; V = length estimation.

1. Factor I, best defined by Form Detection, Oscilloscope Reading, and Form Conversion I and II, appears to be the ability to note quickly and discriminate details in visual patterns. This is the factor usually called perceptual speed, found in all air-crew classification battery analyses.

2. Factor II seems to be the ability to recognize previously seen patterns. While this factor, frequently called visual memory, has never been definitely isolated and pure measures are not known, it is believed to be identical with Memory II identified in air-crew classification battery analyses. None of the present tests have loadings over 0.50 in this factor. Form Conversion I, Coordinate Reading, Polar Grid Coordinate, and the Army General Classification Test have the highest loadings.

3. Factor III seems to be the ability to move one's self mentally in space and predict the result of such movement in terms of position, view of terrain, etc. While this factor is here labelled spatial relations it is probably identified with Space I found in all air-crew classification battery analyses. All the present tests have relatively small loadings with Coordinate Plotting, Target Course Analysis, and Polar Grid Coordinate having the highest loadings.

4. Factor IV best represented by the Army General Classification Test and Scale Reading, seems to be the ability involved in carrying out simple arithmetic computations. This is the well-known numerical facility factor.

5. Factor V appears to be the ability to estimate lengths without the aid of measuring devices. The highest loadings belong to Coordinate Reading, Plot Reading, Coordinate Plotting, Spot Location, and Polar Grid Coordinate. This factor is probably identical with a

tentatively defined length estimation factor isolated in analyses of the air-crew classification batteries.

Later projects concerned with radar observer selection testing incorporated several of the tests developed by the NDRC project in their selection batteries. The Air-crew Evaluation and Research Detachment No. 1 included two such tests Oscilloscope Interpretation and Coordinate Reading, in the experimental battery they administered to Eighth Air Force radar observer students. Psychological Research Project (Navigator) formed a selection battery consisting of the Oscilloscope Interpretation, Scale Reading, and Polar Grid Coordinate Tests, which they administered at advanced navigator schools. The same two tests which the AERD No. 1 had validated in England were later utilized by the teams administering the radar observer selection battery to bombardiers and navigators in the Training Command. Psychological Research Project (Radar) included the Scale Reading, Polar Grid Coordinate, Ratio Estimation, and Spot Location Tests in a battery administered experimentally to radar observer students.

SELECTION RESEARCH BY THE AIRCREW EVALUATION AND RESEARCH DETACHMENT NO. 1[6]

The AERD No. 1, consisting of 6 officers and 15 enlisted men, spent 3 months on temporary duty with the Eighth Air Force in England during the summer of 1944. The purpose of this detachment was twofold: first, to conduct research activities in connection with the selection of lead personnel for very heavy bombardment aircraft, and second, to validate air-crew classification test data in a combat theater. The subsequent radar selection research which the AERD No. 1 conducted was in partial pursuance of this first objective.

At the time AERD No. 1 arrived in England, the demand for radar observers was steadily increasing. The number of radar observers required by the operational groups was of such magnitude that few students could be eliminated once they were entered in the Eighth Air Force radar observer school. Students enrolled in the school came from either the operational groups of the Eighth Air Force or the radar training schools at Langley Field and Boca Raton. The former group had little, if any, radar training before entering the school. Interviews made by AERD No. 1 revealed varying criteria for selecting such officers for radar observer training. Some, it was found, were sent to radar school because they were regarded as being generally competent. Others were sent because they were not members of a crew. Still others were selected because they were surplus.

[6] For a more complete report, see Tepley, W. M., *ed.*, *op. cit.*

The selection of men arriving from Langley Field or Boca Raton was equally subjective and varied.

In view of this situation, inquiry was made into the possibility of making preliminary recommendations on selection before a radar stanine became available. A fact often stressed by experienced radar observers was that the efficient use of radar techniques in combat operations depended upon skill in dead-reckoning navigation. They emphasized that the radar observer's place as a member of a lead crew warranted selection of men with superior aptitude for navigation as well as ability to operate radar equipment.

As a result, AERD No. 1 decided to investigate first the navigation aptitude of current radar observer students. Navigation stanines were obtained from cadet classification records for students in the radar observer class of June 1944. The average of these stanines was found to be 6.4. The average navigator stanine of Navigation school graduates in the AAF Training Command in the winter of 1942–43 was approximately 7.1 and during the winter of 1943–44, approximately 8.0. From these figures it was apparent that men selected for radar training were somewhat lower in aptitude for navigation than their contemporary graduates from navigation training. This finding was discussed with the training directors of the Eighth Air Force and the radar observer school. On the basis of these discussions two steps were taken to insure a better selection of potential radar observers in the period before a radar stanine could be developed. First, a TWX was coordinated with United States Strategic Air Forces and sent by Headquarters, Eighth Air Force, advising Headquarters, AAF, that to insure successful combat operation of radar equipment a minimum navigator stanine of eight was considered essential. Second, arrangements were made to have 44 navigators and bombardiers with navigator stanines of 8 and 9 sent to the radar observer school from one of the Eighth Air Force Replacement Centers. Eighteen officers from this source, nine navigators and nine bombardiers, arrived for instruction in the July class.

Following these preliminary recommendations, the AERD No. 1 turned to the preparation of an experimental battery of tests for the selection of radar observers. Personnel of the unit had already become familiar with the radar observer's job and equipment. By going on training missions, observing ground trainers which simulated airborne radar equipment, conferring with instructors, and studying manuals and other training literature, sufficient information was obtained to permit description of the work in general terms and to provide a basis for choosing tests for an experimental battery.

The job description suggested that the skills of the radar observer could be divided roughly into three categories: skill in operating and tuning, skill in interpreting the oscilloscope, and skill in dead reckoning navigation. These three categories of skills were further refined to indicate the more specific aptitudes and abilities needed for the job. Skill in operating and tuning the apparatus was believed dependent upon the operator's familiarity with mechanical devices, his speed of perception, and his motor coordination. Mechanical Principles, CI903B, was selected to test familiarity with mechanical devices; Spatial Orientation I, CP501B, was chosen as a test of speed of perception: classification test scores on six psychomotor tests were obtained for as many students as possible to give a measure of motor coordination. Skill in interpreting the oscilloscope screen, the second category, seemed to demand proficiency in matching elements from a map with those on a screen, in perceiving patterns, in determining distance and direction, and in interpreting the oscilloscope when jammed with interfering patterns. The operation of matching elements from a map with those on a screen suggested the use of Spatial Orientation II, CP503B. Three new tests developed by the AERD No. 1 were used to test speed of pattern perception: Pattern Identification (X-1, AERD), Pattern Orientation (X-1, ETO-AERD), and Orientation to Landmarks (ETO-AERD). Coordinate Reading (NDRC) and Dial and Table Reading, CP622A, were selected to measure ability at determining direction and distance. Interpreting the oscilloscope through interfering material suggested the use of Gottschaldt figures, AC 12 J, and Oscilloscope Interpretation (NDRC). Skill in dead reckoning navigation, the last of the three categories of skill, emphasized speed and accuracy of computation involving the E-6B computer and airplots. To measure the abilities known to be important for success in navigation training, the following tests were selected: Numerical Operations, CI702A, B; Mathematics B, CI206C; Reading Comprehension, CI614H; and Dial and Table Reading, CP622A. Also, an attempt was made to measure directly skill at dead reckoning by administering a test of navigation proficiency developed by AERD No. 1.

In the foregoing discussion, mention was made of four new tests developed by AERD No. 1: Pattern Identification, CP820A, Pattern Orientation, CP816A, Navigation Proficiency, and Orientation to Landmarks (X-1, ETO-AERD). A description of the first two of these tests may be found in appendix A of this report: the Navigation Proficiency Test is described in chapter 5; Orientation to Landmarks is described below.

The Orientation to Landmarks Test was designed to measure the ability to reorientate oneself quickly to a spatial pattern, seen first from one direction and then from another. Construction of the test was begun at Psychological Research Unit No. 3. It consists of pairs of aerial photographs, one referred to as the reconnaissance photograph and the other as the cockpit view. In the reconnaissance photograph the aerial view is always a vertical one. In the cockpit view the area is photographed from an oblique angle. Five landmarks are encircled and numbered in each reconnaissance photograph. In the cockpit view the same 5 points, along with 10 additional points, are labeled with the letters A through O. The answer for each item is the letter labeling the landmark which matches it. The test is divided into two separately timed parts, each containing 25 items.

Early work with the Orientation to Landmarks Test suggested that the initial orientation to the oblique photograph was accomplished by checking the relationship of outstanding landmarks to one another. It seemed probable that this was a type of pattern perception which was essentially similar to that involved in matching patterns from map to oscilloscope screen.

Following the selection of tests for the experimental battery, AERD No. 1 was faced with the problem of selecting a criterion for validation purposes. Several criteria were considered and rejected, some because complete records were not available and some for other reasons. For example, records of practice bombing missions and camera missions were incomplete because cloud formations frequently interfered with the taking of photographs. Results of combat missions were rejected for the same reason. An attempt was made to obtain ratings on radar observers in five groups of the Third Division of the Eighth Air Force. The effort was abandoned as impractical because only a few had been rated, most of the observers having been just recently assigned to the group. Pass-fail criterion in the radar observer school was discarded because, as mentioned before, the needs of the operational groups were such as to prevent any substantial number of students being failed. After eliminating these possibilities, AERD No. 1 decided to use as a criterion composite grades in radar observer school. The composite grade was a mean of grades obtained by assigning a weight of 2 to the mean flight grade and a weight of 1 to the mean ground trainer grade. Usually the mean flight grade consisted of 4 flight grades; the mean ground grade was composed of approximately 8 to 12 ground grades.

Table 10.3 presents correlations between the composite grade and the tests administered to classes 44–4 and 44–5. These correlations were based only upon students from Langley Field and Boca Raton because it was expected that subsequent classes at the school would be composed entirely of such students. In the statistical treatment of the data which follows, these students will be referred to as the Langley group. Table 10.3 indicates that the following seven tests had correlations of sufficient magnitude with the composite grade to be of predictive value: Oscilloscope Interpretation, Coordinate Reading I and II, Spatial Orientation I, Pattern Orientation, Two-Hand Coordination, and Complex Coordination. Since the intercorrelation of Parts I and II of Coordinate Reading was 0.74 they were combined in a single score.

Table 10.3—*Means, standard deviations, and correlations of tests with composite grades for Langley men only of classes 44-4 and 44-5*

Test	N	M	SD	r
Bombardier stanine	73	5.36	1.88	0.26
Navigator stanine	74	7.00	1.53	.03
Pilot stanine	71	5.27	1.80	.29
Oscilloscope Interpretation (NDRC)	121	36.44	8.70	.27
Coordinate Reading, part 1 (NDRC)	59	42.78	8.49	.38
Coordinate Reading, part 2 (NDRC)	59	36.76	8.46	.34
Coordinate Reading, total (NDRC)	59	79.55	15.20	.40
Numerical Operations, C1702A	123	24.72	6.60	.13
Numerical Operations, C1702B	123	21.60	7.20	.08
Mathematics B, C1266C	122	20.46	10.00	.08
Spatial Orientation I, CP501B	123	62.70	12.55	.33
Spatial Orientation II, CP503B	124	26.23	7.58	.13
Orientation to Landmarks (ETO-AERD)	119	22.38	8.32	.12
Pattern Orientation (X-1, ETO-AERD)	64	21.07	6.39	.37
Pattern Identification (X-1, ETO-AERD)	62	24.76	7.35	.02
Dial and Table Reading, CP622A, 621A	121	29.80	6.40	.08
Reading Comprehension, C1614H	121	33.78	13.00	.01
Mechanical Principles, C1903B	65	31.11	8.67	.10
Gottschaldt Figures, AC12	118	29.94	8.16	.14
Navigation Proficiency Test A (ETO-AERD), part 1	121	14.79	3.68	.05
Navigation Proficiency Test A (ETO-AERD), part 2	121	19.96	4.86	.07
S. A. M. Rotary Pursuit, CM803A	47	48.14	6.75	.02
S. A. M. Two-Hand Coordination, CM101A	52	43.39	10.08	.32
S. A. M. Discrimination Reaction Time, CP611D	69	51.52	8.61	.18
S. A. M. Steadiness, CE206B	67	49.66	12.06	.22
S. A. M. Finger Dexterity, CM116A	69	48.01	11.46	.19
S. A. M. Complex Coordination, CM701A	69	47.38	10.08	.29

It will be noted that the correlation of composite grade with navigator stanine is 0.03, with Navigation Proficiency Test, Part I (pilotage and instruments) 0.15, and with Navigation Proficiency Test, Part II (E-6B computer), 0.07. This seemed to provide a clear demonstration that composite grade for the type of training given at the radar observer school was a function of variables other than navigational aptitude and proficiency. Persons conducting the study emphasize that this should not be misinterpreted as evidence against the position stated earlier that navigators of high ability should be selected for radar observer training since the school criterion did not require the degree of navigational skill called for under operational conditions.

The 6 tests most highly correlated with composite grades were intercorrelated. The intercorrelations based on classes 44-4 and 44-5 are presented in table 10.4. The N varies from 85 to 170 because not all of the tests were administered to all classes.

Table 10.4

	1	2	3	4	5	6
1. Oscilloscope Interpretation (NDRC)	....	0.29	0.37	0.17	0.27	0.39
2. Coordinate Reading (NDRC)	0.29	....	.23	.34	−.08	.24
3. Spatial Orientation I, CP501B	.37	.23	....	.52	.13	.24
4. Pattern Orientation (X-1, ETO-AERD)	.47	.31	.52	....	.15	.26
5. S. A. M. Two-Hand Coordination, CM701A	.27	.08	.13	.15	....	.45
6. S. A. M. Complex Coordination, CM101A	.39	.24	.24	.26	.45	....

After taking account of magnitude of intercorrelations, correlations with the composite grade, and standard deviations, the tests were assigned tentative weights as follows:

Test:	Weight
Oscilloscope Interpretation	1
Coordinate Reading	1
Spatial Orientation I	1
Pattern Orientation	2

On the basis of this weighting, an aggregate score was obtained for each student in class 44–6, and a prediction made of composite grade in terms of above average, average, and below average. The two psychomotor tests were not included. Scores for them were available for only approximately 50 percent of class 44–6. The results were as follows:

> Above-average group: Of this group 74 percent were above the median composite grade, while 26 percent were below this grade. Two students were eliminated.
>
> Average group: Twenty-eight percent of this group were above the median composite grade, 72 percent were below, and one student was eliminated.
>
> Below-average group: Thirty percent of this group were above the median composite grade and 70 percent were below. Eight students were eliminated.

Thus, if the lower third on the tests had been excluded from training, only 3 instead of 11 students would have been eliminated.

Students from class 44–6 were added to the previous classes and validities on the six most promising tests calculated for the total numbers of available cases. The results are shown in table 10.5.

TABLE 10.5.—*Validity coefficients and N's for the six tests of highest validity in Air-Crew Evaluation Research Detachment No. 1 study*

Tests	r	N
Coordinate Reading (NDRC)	0.32	144
Oscilloscope Interpretation (NDRC)	.25	243
Pattern Orientation (X 1, ETO-AERD)	.26	149
Spatial Orientation I, CP501B	.21	207
S. A. M. Complex Coordination, CM701A	.26	150
S. A. M. Two-Hand Coordination, CM101A	.29	109

Test:	Weight (percent)
Coordinate Reading (NDRC)	36
Two-Hand Coordination, CM101A	29
Complex Coordination, CM701A	21
Spatial Orientation I, CP501B	14

Using these data, Psychological Research Project (Radar) later found that the correlation of the best possible weighted combination of these six tests with success in radar observer training would be

0.49. The best possible weighted combination of the following four of these tests would correlate 0.48 with the same criterion.

In conjunction with its use of the composite grade as a criterion, AERD No. 1 made several comparisons of the relative success of various groups of students within the school. As described earlier in this chapter, the radar students tested in the course of this study fell into two primary groups—those with previous radar training in the United States and those without such training. The latter, recruited from operational groups, will be referred to in discussing the statistical treatment of the data as the non-Langley sample. Students receiving radar observer training also differed with respect to their classification as either navigators or bombardiers. Data concerning such classification were available for classes 44–4, 44–5, and 44–6.

For the four classes shown in table 10.6 it will be seen that there are only small differences in composite grades between students with previous radar training at Langley and those without. The greatest difference is in class 44–4. Here the biserial correlation of composite grade with presence or absence of Langley training was 0.36. For class 44–5 the biserial correlation of the same variables was 0.06. Of interest in class 44–5 was the relatively favorable showing of the 18 non-Langley students. As pointed out earlier those students were selected to fill out class 44–5 on the assumption that only men having the highest aptitude for general navigation should receive training in the AN/APS–15 operation. Their mean composite grade was as high as that of students with previous radar experience or combat navigation.

TABLE 10.6.—*Means and standard deviations of composite grades, Langley and non-Langley students, for each class in Eighth Air Force Radar Observer School*

Class	Langley			Non-Langley		
	N	M	SD	N	M	SD
44-2	48	78.42	4.78	22	78.18	6.00
44-3	43	77.42	3.23	28	75.50	3.79
44-4	69	74.70	4.34	24	71.50	6.90
44-5	66	76.86	2.73	[1] 18	76.61	2.54

[1] Specially selected high stanine navigators and bombardiers.

Table 10.7 presents the mean composite grades and standard deviations for the navigators and bombardiers of the Langley group for two classes. Examination of this table indicates that the two categories of students did equally well. This comparison is partly vitiated by the fact that a sizable proportion of the bombardier students had received intensive training in dead reckoning navigation while serving as navigation instructors in bombardier schools.

TABLE 10.7.—*Means and standard deviations of composite grades for navigators and bombardiers (Langley students only) for classes 44-4 and 44-5 of Eighth Air Force Radar Observer School*

Class	Navigators			Bombardiers		
	N	M	SD	N	M	SD
44-4	31	74.50	4.36	28	74.43	4.28
44-5	40	76.35	2.99	26	77.77	2.08

Nevertheless, significant differences did exist between navigators and bombardiers with respect to navigation proficiency as measured by the Navigator Proficiency Test (ETO-AERD). Table 10.8 sum-

TABLE 10.8.—*Point biserial correlations for classes 44-5 and 44-6 between navigator-bombardier classification and various parts of the navigator proficiency test (ETO-AERD)*

Part of test	Class 44-5 point biserial r's N=66	Class 44-6 point biserial r's N=85
E-6B computer	0.40	[1] 0.93
Instruments	.78	....
Pilotage and pinpointing	.45	....
Airplot	....	.69

[1] The E-6B computer part of the test given to Class 44-6 differed from the same part given to class 44-5 in two respects: (1) The items were more difficult and (2) a greater variety of problems was included. The number of items, however, was the same for both classes.

marizes point biserial correlations between the navigator-bombardier classification and scores obtained on the various parts of this test.

STUDENT SELECTION BY PSYCHOLOGICAL RESEARCH PROJECT (NAVIGATOR)

In accordance with a directive from Headquarters, AAF, in July 1944, the Psychological Research Project (Navigator) adapted special screening tests developed by the NDRC project to aid in the selection of individuals for radar training in the AAF Training Command. At this time, students being trained were almost exclusively bombardiers, but upon the recommendation of Headquarters, Eighth Air Force that only men with navigator stanines of eight or more be given radar observer training, it was decided that these tests be administered in the advanced navigator schools.

The selection battery decided upon consisted of three tests developed by the NDRC project. They were Oscilloscope Interpretation, CP817A, Scale Reading, CP637A, and the Polar Grid Coordinate Test, CP819B. These tests were adapted to machine scoring which involved minor changes in a few of the items. Additional criteria for selection were a strong preference for this type of training and a navigator stanine of eight or nine. Preference for radar observer training was indicated by choosing one of four statements arranged in diminishing interest order from one to four (i. e., "1" represented highest degree of interest, "4" the lowest).

In August and September of 1944, representatives of the Navigator Project administered the test battery at San Marcos, Hondo, and Ellington Air Fields. After the initial testing at each of the advanced schools, suitable representatives of each school were designated to carry on the testing in order to allow the project to concentrate on its research activities. However, the Navigator Project still maintained a general, supervisory responsibility with respect to this work.

At each school a roster of students meeting the stanine and preference requirements was prepared listing them in order of their radar aptitude test scores. The radar aptitude score was a composite score arrived at by adding the raw scores on the Oscilloscope Interpretation (CP817A) and Scale Reading (CP637A) tests to one-half the raw score on the Polar Grid Coordinate Test (CP819B). Radar observer schools quotas were filled by selecting men from this roster in order of their radar aptitude score.

The Navigator Project terminated radar observer selection activities when Headquarters, AAF activated test teams for the specific purpose of radar selection testing.

STUDENT SELECTION BY HEADQUARTERS, AAF TRAINING COMMAND

A directive from Headquarters, AAF Training Command, Fort Worth, Tex., in September 1944, established six traveling testing teams, two in each Training Command, to carry out the selection testing begun by the Navigator Project. The number of teams was reduced to two in January 1945. Each team was composed of an orientation officer who was a combat returnee, an aviation psychologist, and two psychological assistants. When selection of potential radar observers was first undertaken by these teams, only rated pilots were tested. After 11 December 1944, pilots were replaced by rated bombardiers and navigators.

The test battery administered by the testing teams consisted of four of the same tests which AERD No. 1 had administered to students in the radar observer school of the Eighth Air Force. They were: Pattern Orientation, CP816A; Coordinate Reading, CP224A, B; Oscilloscope Interpretation, CP817A; and Spatial Orientation I, CP501B. The battery also included a nine point numerical interest scale ranging from "1" which represented little or no interest, to "9" which indicated exceptionally strong interest.

A composite radar score was computed by adding the raw scores on the four tests. Since the standard deviation of the Coordinate Reading Test was approximately twice the size of each of the other three tests, this test received a correspondingly high weight. For practical use the composite score was converted to a stanine.

Navigators and bombardiers assigned to radar training were selected in order of their radar observer aptitude stanines. To be eligible, a radar observer stanine of six or above was required. In addition, a navigator stanine of at least seven and an interest preference of three or above were required. Provisions were made, however, to progressively lower these standards if higher quotas necessitated the change. After 1 May 1945, most of the assignments of bombardiers to radar observer training were made from the upper half of the distribution on a standardized navigation proficiency test. This selection requirement operated in addition to those previously mentioned.

The table below summarizes the recommendations of bombardiers and navigators to radar observer school for the period 9 November 1944 through 30 June 1945. Because of an insufficient number of bombardiers and navigators available for radar observer training, selection testing was terminated 31 July 1945.

TABLE 10.9.—*Summary of radar observer screening and recommendations for the period 9 November 1944 through 30 June 1945*

Total number of bombardiers screened	6,957
Total number of navigators screened	6,859
Grand total screened	13,816
Total number of bombardiers recommended	2,051
Total number of navigators recommended	1,988
Grand total recommended	4,039
The bombardiers recommended included:	*Percent*
Combat returnee volunteers	4.7
Permanent party	9.1
Recent graduates	86.2
The navigators recommended included:	
Combat returnee volunteers	3.3
Permanent party	4.4
Recent gradutes	92.3

Preparations were made in May 1945, to extend the selection of radar-observer students to unclassified cadets. At the request of Headquarters, AAF, Psychological Research Project (Radar) formulated recommendations for tests and weights upon which a radar observer stanine was to be based in classification centers.

Previously, the Radar Project had found the best possible weighted combination of 4 of the 6 tests with appreciable validity in the AERD No. 1 study to be as follows:

Test:	Weight, percentage
Coordinate Reading, CP224A, B	36
Two-Hand Coordination, CM101A	29
Complex Coordination, CM701A	21
Spatial Orientation I, CP501B	14

While basing its recommendations primarily upon the validity statistics reported by AERD No. 1 the project took note of the fact that certain tests then in the classification battery had not been available to AERD No. 1 at the time of its validation studies. It was felt, moreover, that class grades used as a criterion in the AERD No. 1 study were determined primarily by the ability of the radar observer to use the radar set and did not reflect the navigation and bombing requirements which would be present in the Training Command criterion. A summary of the test recommended and the weights assigned to each is presented below:

Test:	Weight, percentage
Complex Coordination, CM701A	20
Two-Hand Coordination, CM101A	20
Instrument Comprehension II, CI616B	10
Spatial Orientation I, CP501B	10
Coordinate Reading, CP224A, B	20
Dial and Table Reading, CP622A and 621A	10
Arithmetic Reasoning, CI206C	10

In the original combination of four tests listed above, 50 percent of the weight had been assigned to factors measured by Two-Hand Coordination, CM101A and Complex Coordination, CM701A. Instrument Comprehension II, CI616B, however, was not in the classification battery at the time air-crew officers in this sample were classified. Factor analyses completed at the Psychological Research Unit indicated that Instrument Comprehension II, CI616B, was heavily loaded with a factor which was also an important component in Complex Coordination, CM701A, and Two-Hand Coordination, CM101A. This is the ability called spatial relations in chapter 4. Job analysis led to the conclusion that this was one of the more important abilities determining success in the radar observer's task. As a result, a weight of 10 percent was recommended for Instrument Comprehension II, and 20 percent each for Two-Hand Coordination and Complex Coordination.

It was recommended that Arithmetic Reasoning, CI206C, and Dial and Table Reading, CP622A and 621A, both tests of navigational aptitude, be included in the stanine because of the conviction that, against a Training Command criterion, abilities important to navigation would also determine success for radar observers. For each of these tests a weight of 10 percent was suggested. Some of the weight indicated by the AERD No. 1 results for Coordinate Reading, CP224A, B, was assigned to these tests partly because of the correlation of 0.70 found between Coordinate Reading and Dial and Table Reading in the AERD No. 1 study. The weight assigned to Spatial Orientation I, CP501B, was reduced from 14 percent to 10 percent because of the fact that two of the tests added, Instrument Comprehension II and Dial and Table Reading had substantial loadings in perceptual speed. Finally, it was recommended that Coordinate Reading, CP221A, B, be added to the classification battery with a suggested weight of 10 percent for the radar stanine.

After 1 June 1945, all students taking the air-crew classification tests were given a radar stanine computed in terms of these recommendations. However, because of the termination of hostilities, no students were assigned to training on this basis.

SUMMARY

The chief contributors to the development of a selection battery for radar observers were Project SC-70, NS-146, of the National Defense Research Committee and the Air-crew Evaluation and Research Detachment No. 1. The NDRC project conducted the initial research in the field, and prepared a number of tests for purposes of radar selection. Three of these tests, Oscilloscope Interpretation, CP817A, Scale Reading, CP637A, and the Polar Grid Coordinate Test, CP819B, comprised the radar observer selection test battery administered by Psychological Research Project (Navigator). Oscilloscope Interpretation, CP817A, and Coordinate Reading, CP224A, B, were used in the experimental battery administered by AERD No. 1 and also by traveling teams administering the radar selection battery at advanced navigator and bombardier schools. Coordinate Reading was later included in the Air-crew Classification Battery to aid in determining a radar stanine for students taking the classification tests.

As a part of its broad assignment to research on lead crew selection, AERD No. 1 validated an experimental battery of potential radar observer aptitude tests. Of the 24 tests which AERD No. 1 validated against composite grades in radar observer school, 6 had validity coefficients of sufficient magnitude to be of predictive value. They were: Oscilloscope Interpretation, CP817A, Coordinate Reading, CP224A, B, Spatial Orientation I, CP501B, Pattern Orientation, CP816A, Two-Hand Coordination, CP224A, B, and Complex Coordination, CM701A. The best possible weighted combination of these six tests had a correlation of 0.49 with success in radar observer training. Later, four of these tests, Oscilloscope Interpretation, Coordinate Reading, Spatial Orientation I, and Pattern Orientation formed the Radar Observer Selection Battery administered by traveling teams at advanced navigator and bombardier schools in the training command.

Psychological Research Project (Navigator) conducted the first selection testing for the purpose of assigning students to radar observer training. The selection battery used has been listed above. When Headquarters, AAF, expanded the selection program and activated traveling teams to test students at advanced navigator and bombardier schools, the Navigator Project terminated its radar observer selection activities.

The teams activated by Headquarters, AAF, in September 1944, administered a selection battery of four tests which AERD No. 1 had validated at the Eighth Air Force Radar Observer School in England. Requirements for assignment to radar school were a radar stanine of six or more and a navigator stanine of seven or above. The radar stanine was derived from the student's scores on the selection test battery and an interest indication of at least three on a nine-point scale. After 1 May 1945, a navigation proficiency test was included in the battery administered at advanced bombardier schools.

Psychological Research Project (Radar) in May 1945 made a number of recommendations relative to the tests and weights to be used in determination of a radar stanine from air-crew classification test scores. These recommendations were based partly on validity statistics reported by AERD No. 1 and partly on other considerations. The tests and their weights were as follows:

Test:	*Weights (percentage)*
Complex Coordination, CM701A	20
Two-Hand Coordination, CI101A	20
Instrument Comprehension II, CI616B	10
Spatial Orientation I, CP501B	10
Coordinate Reading, CP224A, B	20
Dial and Table Reading, CP622A	10
Arithmetic Reasoning, CI206C	10

After 1 June 1945, all students taking the air-crew classification tests were given a radar stanine computed in terms of these recommendations. However, because of the termination of hostilities, no students were assigned to training on this basis.

CHAPTER 11

Validation of Selection Tests for Radar Observer Training[1]

INTRODUCTION

From the time of its activation, the main goal of the Psychological Research Project (Radar) was the validation of tests for radar observer selection. Even though, as explained in chapter 3, initial emphasis was placed upon the development of proficiency measures, it was clear that the primary research role of these measures was to serve as criteria for validating selection tests. Similarly, it was for this reason primarily that the analysis of bombing errors described in chapter 9 was undertaken. This concern with the development of acceptable criteria seemed, in radar observer training, to be a necessary prerequisite to the potentially more significant research upon test validation.

In this chapter a report is made of the results of validating various selection tests against quantitative job criteria, both course grades and bombing scores. The methods and results of two extensive validation studies are described.[2] For convenience, the studies will be referred to as validation studies I and II. For each, descriptions are given of the variables validated, the criteria, and the samples as well as the resulting validity coefficients. In study II multiple regression statistics are presented for various combinations of the variables found to have appreciable validity. In each study two validation samples were used, one of bombardiers and one of navigators. The reliability of the course grades was determined for portions of the validation samples. The estimates of reliability are presented with considerable misgivings but are included because they are the best estimates available under the circumstances.

The results of the two studies are discussed jointly at the end of the chapter. An attempt is made to summarize the evidence regarding the importance of various abilities to success in radar-observer training.

[1] Written by Cpl. Harold H. Kelley.

[2] The following personnel were responsible for the planning of the validation studies: S/Sgt. Roland E. Johnston, Jr., Cpl. Kelley, Lt. Sol M. Roshal, and S/Sgt. Bernard C. Sullivan.

VALIDATION STUDY I

Validation study I consists of the validation of selection test scores which were available for radar-observer students before experimental selection testing was begun by the Radar Project. The criterion against which the scores were validated consisted of the radar-observer course grades assigned by the training stations. Two samples were used, a bombardier sample including mostly students from Boca Raton and Langley Field and a navigator sample including mostly students from Victorville.

Variables

This study validates the variables in two batteries of selection tests: the air-crew classification battery (November 1943) and the radar-observer selection battery. Appendix A describes each of the tests composing the two batteries.

The air-crew classification battery used in this study consists of 11 printed tests and 6 psychomotor tests. It was administered to aviation students before they entered preflight school. Test scores were differentially weighted and combined into three stanines or standard scores indicating aptitude for each of the three air-crew specialties: bombardier, navigator, and pilot. It was on the basis of stanines that aviation students were classified and assigned to one of the three air-crew specialties. Although the air-crew classification battery was changed at different times as new tests and validation information became available, variables for the present study are all taken from the November 1943 battery which was administered to aviation students from November 1943 to September 1944. Appendix A describes the composition of the three stanines computed from the battery in terms of how much each test contributes to each stanine.

The radar-observer selection battery consi s of four printed tests and an interest blank. It was administered to bombing students and navigation students just prior to their being commissioned and graduated from advanced training. Test scores were weighted and combined into one stanine for the air-crew specialty of radar observer (bombardment). The composition of this stanine is described in chapter 10. On page 224 is described the Interest Blank, Radar Preference I, which is an indication of the strength of preference for radar-observer training expressed at the time of taking the test battery.

In all, 27 selection variables are represented in validation study I. From the air-crew classification battery there are 21 variables, including the three stanines. From the radar-observer selection battery there are 6 variables including the one stanine and Radar Preference Rating I.

Validation Criteria

All selection variables were validated against the final course grades assigned students graduating from the radar-observer course. Table 11.1 presents the specific student classes from which the criterion grades in the present study were taken. Course grades at all three schools were determined as a composite of flight, trainer, and classroom grades. For all classes in the present study with the exception of six Victorville classes, the flight grades contributed 60 percent of the composite, and trainer and classroom grades each contributed 20 percent. For Victorville classes 45-10 through 45-15, flight grades made up 75 percent of the total course grades with the remaining 25 percent being half trainer and half classroom grades.

TABLE 11.1.—*Radar observer classes used from each school in Validation Study I*

School	Classes	Graduation dates
Langley	45-7 through 45-18	24 February 1945–5 May 1945.
Boca Raton	45-3 through 45-11	24 March 1945–19 May 1945.
Victorville	45-10 through 45-19	10 March 1945–12 May 1945.

The course grades for these classes represented ratings given by instructors, scores from informal classroom quizzes, and scores from standardized tests and performance checks developed by the Radar project. Only a rough estimate can be made of the extent to which the three schools based course grades upon instructor ratings and classroom quizzes as compared with the relatively more standardized tests and checks. At Boca Raton and Langley Field, about half the classes in the present study were given the complete battery of proficiency measures developed by the Radar Project. This battery is described in chapters 5 and 6. At Victorville, the last class was given most of the Radar Project tests and checks and the preceding three classes were given almost all of the tests. Of the remaining classes only two received one or more standardized tests.

The three schools made varying use of the scores from such standardized measures as were employed during this period. It is estimated that for the Boca Raton and Langley Field classes in study I, the course grades were based upon scores from standardized measures and informal grades in equal proportion. At Victorville, however, course grades were based almost completely upon instructors' daily grades and informal classroom tests.

For each training station, a distribution of course grades was made for all students in the classes studied.[3] Two distributions were necessary for Victorville due to a shift in the grading system from class 45-15 to class 45-16. To make the scores from the three schools com-

[3] This statistical work was carried out by Sgt. Gerald S. Blum and S/Sgt. Harold F. Kunsman.

parable, the scores from Langley Field and Victorville were converted to distributions corresponding to the Boca Raton distribution which has a mean of 84.08 and a standard deviation of 3.96. Distributions of the converted course grades were made separately for the bombardier sample and navigator sample finally used in obtaining the validity coefficients. Statistics from these distributions are presented in table 11.2. It may be noted, incidentally, that on the average navigators obtained higher course grades than bombardiers, the difference having a critical ratio of 4.33. This is significant at the 1 percent level.

TABLE 11.2.—*Statistics from distributions of converted course grades for bombardier and navigator samples, Validation Study I*

Sample	M	SD	N
Bombardiers	83.17	4.13	205
Navigators	84.65	3.55	381

Critical ratio of difference between means = 4.33.[1]

[1] Significant at the 1-percent level.

No precise measure could be made of the reliability of the course grades for the validation samples. However, a rough estimate was made, using the course grade and its components for Langley Field classes 45-11 through 45-15. Course grades for these classes were based partly upon standardized tests and performance checks, and partly upon daily grades and informal tests as were the course grades for other Langley Field and Boca Raton classes in study I. Whether or not the coefficient obtained is a fair estimate of the reliability of Boca Raton and Langley Field course grades used in the first validation study depends upon how representative these classes were of the larger group.

The reliability of classroom grades was estimated by correlating the average grades for odd weeks with the average grades for even weeks.[4] Similarly, for flight and trainer grades, average grades on odd missions were correlated with average grades on even missions. Odd and even course grades were computed by averaging the odd and even ground, trainer, and flight grades. Raw flight grades were weighted 60 percent while ground and trainer raw grades were each weighted 20 percent. The odd-even correlations and the correlations corrected for length are presented in table 11.3.

The auxiliary scope missions referred to in the table are flown at the auxiliary PPI scope in the nose of the aircraft. On such missions the student operated the bombsight, observed the auxiliary scope, did follow-the-pilot navigation, and maintained a navigation log. This initial aerial training, to which 6 to 10 flights were devoted,

[4] Sgt. John S. Harding planned the method used for estimating the reliability of the course grades. The statistical work was accomplished by Sgt. Samuel D. Morford.

served to familiarize the student with the bombsight, interpretation of returns on the PPI scope, and coordinated bombing technique. No reliability is reported for bench set trainer grades since only the final bench set check score was available for a large portion of the sample.

TABLE 11.3.—*Estimates of reliability of radar-observer-course grades and component grades for classes 45-11 through 45-15, Langley Field, Validation Study I*

Grade	N	r_{11}	r_{11}[1]
Classroom:			
Set operation	222	[2] 0.40	0.58
Navigation	227	[3] .18	.30
Bombing	225	−.07	−.16
Radar intelligence	226	.12	.21
Trainer:			
Supersonic	217	[2] .20	.33
Flight:			
Auxiliary scope missions	226	[2] .25	.40
Main scope missions	225	[2] .63	.77
Course grade	220	[2] .48	.65

[1] Corrected for length by the Spearman-Brown prophecy formula.
[2] Significant at the 1-percent level.
[3] Significant at the 5-percent level.

If this estimate is representative, the course grades for Boca Raton and Langley Field have a corrected reliability of 0.65. To the extent that the odd and even grades in any component of the aggregate were not independent, i. e., were systematically biased in the same direction, this figure is an overestimate. Because the grades assigned at Victorville are composed of subjective ratings to a greater extent, their reliability is thought to be somewhat lower.

Composition of Validation Samples

The cases used in computing validity coefficients were selected from the 1,339 students in classes shown in table 11.1. Only radar observer students who were rated bombardiers or navigators were used. A number of cases were eliminated because their specialty rating was not identified or because they were eliminated pilots. Other cases were eliminated because of incomplete scores for the November 1943 air-crew classification battery. The samples of navigators or bombardiers who had taken any other single battery were too small to warrant statistical analysis. Many other cases were lost because they could not be identified on the course grade rosters or located on the microfilm rosters of air-crew classification test scores.[4]

The validation samples consisted of 586 cases, 205 bombardiers and 381 navigators. The distribution of bombardiers and navigators by schools is presented in table 11.4. Comparing the cases from Boca Raton and Langley Field with the Victorville cases in the sample, it is to be noted that relatively more navigators are from Victorville. This

[4] Complete sets of air-crew classification test scores are recorded on microfilm rosters for all individuals who took any of the batteries. These rosters are prepared by Headquarters, AAF Training Command.

difference is significant at the 1 percent level. Since, as already pointed out, Victorville course grades are based upon standardized measures to a lesser extent than those from the other two schools, there is a systematic difference between bombardier and navigator course grades. Specifically, the grades for bombardiers are based to a greater extent upon Radar Project tests and performance checks than are the grades for navigators. This difference is actually greater than it appears because the navigators from Boca Raton and Langley Field were in the early classes in the sample. Because more tests and checks were put into use for later classes, navigators from these schools had fewer standardized measures contributing to their course grades than did bombardiers.

TABLE 11.4.—*Validation cases selected from 1,339 students in consecutive radar observer classes; Validation Study I*

School	All cases	Validation cases		
		Total sample	Bombardier sample	Navigator sample
Langley	490	159	52	107
Boca Raton	252	83	62	21
Victorville	597	314	91	253
Total	1,339	556	205	381

Validation Statistics

The validity coefficients for the variables in this study are presented in table 11.5.[6] They will be interpreted along with the results of validation study II in the section beginning on page 257 which presents the factors indicated to have appreciable validity. Examination of table 11.5 shows that two of the validity coefficients for the bombardier sample are significant at the 1-percent level while an additional five are significant at the 5-percent level. For the navigator sample, five coefficients are significant at the 1-percent level and an additional four are significant at the 5-percent level.

A multiple correlation was not computed for the valid variables in this study because the same tests are included along with a number of other tests in validation study II and are validated against what is thought to be a better criterion.

Correction for restriction of range.—The absolute sizes of the validity coefficients must be evaluated in the light of the various selection procedures to which the samples are known to have been subjected. Four major selection procedures which have operated to restrict the ranges of various abilities in the present samples are: the Aviation

[6] Processing of these validation data on International Business Machine punch cards was planned and carried out by Lt. Roshal with the assistance of S/Sgt. Johnston, Sgt. William J. Mangan, Sgt. Harold I. Raush, and S/Sgt. Sullivan. Sgt. Blum computed the correlation coefficients from these machine data.

TABLE 11.5.—*Product-moment correlations between the radar observer course grade and selection variables from Validation Study I*

Selection variable and code number	Bombardiers, N=205			Navigators, N=381			Unrestricted statistics [1]			
							MPEU		PRU	
	Validity	M	SD	Validity	M	SD	M	SD	M	SD
Stanines:										
Bombardier stanine	0.12	6.88	1.27	[2]0.15	7.21	1.45	5.13	2.12	5.66	2.09
Navigator stanine	[2].25	6.30	1.39	[2].15	8.22	1.10	4.50	1.81	5.38	1.86
Pilot stanine	[2].17	5.77	1.53	.07	5.38	1.75	4.65	1.79	5.48	1.82
Radar stanine	[4].08	5.20	1.62	......	.....	.....	[5]4.94	[5]1.76	[6]6.44	[6]1.77
Radar observer selection battery:										
Coordinate reading, CP224A or B	[4].05	101.06	18.63	......	.....	.....	.....	.....	.....	.....
Oscilloscope interpretation, CP817A	[4].08	22.63	6.22	......	.....	.....	.....	.....	.....	.....
Position orientation, CP816A	[4].05	27.45	7.36	......	.....	.....	.....	.....	.....	.....
Spatial orientation I, CP501B	[4].10	35.64	5.39	......	.....	.....	.....	.....	.....	.....
Radar preference, I	[4].00	7.85	2.01	......	.....	.....	.....	.....	.....	.....
Aircrew Classification Battery (printed tests):										
Biographical data blank (navigator score)	[4].16	22.64	2.90	−.01	23.77	2.86	21.79	3.03	22.30	3.01
Biographical data blank (pilot score), CE602D	.10	25.91	5.70	[2].12	27.33	7.22	26.76	6.47	27.22	6.79
Dial and table reading, CP622A, CP621A	[4].14	39.15	7.27	[2].20	44.42	7.26	32.36	9.85	34.49	8.89
General information, CE505E	[3].14	42.27	13.19	.04	45.96	15.28	32.33	13.50	45.64	11.95
Instrument comprehension I, CI615B	[3].14	9.66	2.78	−.03	8.14	2.75	8.30	3.38	10.62	3.28
Instrument comprehension II, CI616B	.11	31.73	10.86	.08	37.06	10.29	25.50	10.71	29.63	10.77
Mathematics A, CI702F	[2].21	11.28	7.86	.08	22.75	9.73	5.29	6.73	10.56	8.11
Mathematics B, CI206C	.10	16.50	9.98	[2].12	25.85	9.56	10.83	8.91	14.90	9.28
Mechanical principles, CP903B	.07	31.47	8.23	[2].18	36.86	9.81	30.39	8.86	32.34	9.33
Reading comprehension, CI614H	.12	22.41	13.32	[2].15	33.94	13.90	14.17	11.86	20.79	12.42
Spatial orientation I, CP501B	.08	30.12	5.18	.07	31.86	5.26	28.33	5.61	27.97	5.79
Spatial orientation II, CP503B	.09	21.14	6.51	[2].10	25.15	6.35	19.07	6.49	20.45	6.57
Air-crew classification battery (psychomotor tests):										
Complex coordination CM701A	.04	54.73	9.08	−.01	55.27	10.03	49.55	9.66	51.83	10.67
Discrimination reaction time, CP611D	−.06	56.29	8.96	.07	56.53	6.26	49.82	10.08	50.96	9.83
Finger dexterity, CM116A	.05	54.22	9.97	[2].13	52.62	10.59	51.25	10.44	51.34	10.07
Rotary pursuit, CP410B	.03	52.50	8.84	.03	51.06	10.18	49.19	9.97	49.81	10.07
Rudder control, CM120A	.05	50.75	9.56	−.01	49.79	10.94	43.64	10.33	50.22	10.07
Two-hand coordination, CM101A	.12	51.39	10.22	.08	53.45	9.66	50.53	10.14	51.07	10.13

[1] In these columns are presented the M's and SD's obtained from administrations to unclassified aviation students. Results of testing at Psychological Research Unit and Medical and Psychological Examining Units are presented separately because of significant differences in the obtained statistics. It was not easily possible to determine for validation samples where the individuals were given their classification tests. It is thought that the majority of them were tested at Psychological Research Units. The Medical and Psychological Examining Unit data are based upon 1,520 cases tested at Units 4 to 10, reported in Research Note 44-1, Psychological Section, Office of the Surgeon, Hq. AAFTC, Fort Worth, Tex., 10 Jan. 1944. The Psychological Research Unit data are based upon 1,500 cases tested at Units 1 to 3, reported in Research Note 44-3, Psychological Section, Office of the Surgeon, Hq. AAFTC, Fort Worth, Tex., 1 Feb. 1944.

[2] Significant at the 1-percent level.

[3] Significant at the 5-percent level.

[4] These statistics are based on 137 bombardiers who had taken the Radar Screening Battery.

[5] These statistics are based upon all bombardiers (N=6,813) who took the Radar Observer Selection Battery prior to 1 Apr. 1945.

[6] These statistics are based upon all navigators (N=3,703) who took the Radar Observer Selection Battery prior to 1 Apr. 1945.

Cadet Qualifying Examination administered prior to aviation student status, the air-crew classification battery, proficiency grades in bombardier or navigator training, and the radar stanine and other variables described in chapter 10. It would be highly desirable to be able to compare the validities obtained in this study with those obtained in other air-crew specialties for unclassified aviation students. This would constitute comparison with samples that were not restricted on the bases just described. However, this comparison could not be made because the data required for satisfactorily correcting validities for such a complex set of restricting variables were not available. To enable the reader to make his own estimates of the amount of restriction that occurred in the various test scores due to the restricting influences, means and standard deviations obtained from administrations of the selection variables to samples of unclassified cadets are presented in table 11.5 parallel to the same statistics for the present validation samples. The reader will note that while most of the standard deviations for the validation samples are somewhat smaller than the unrestricted standard deviations, several tests, notably Reading Comprehension, Mathematics A, and Mathematics B, have larger standard deviations. These increases constitute an artifact which arises because the tests in question are too difficult for the general aviation student group. Scores for such a group pile up at the low end of the frequency distribution. After selection occurs and this pile-up of low scores is eliminated, the standard deviation increases. It will also be noted that the unrestricted means for tests which were weighted in computing the bombardier or navigator stanine (see appendix A) are lower than the means in the corresponding bombardier or navigator validation sample.

Formulas are available for correcting correlations for restriction of range based upon a single variable.[7] Using these, the validity coefficients obtained for 137 bombardiers who had taken the radar observer

[7] Two formulas were used in the present study. Both were derived by Karl Pearson (Mathematical Contributions to the Theory of Evolution—XI. On the Influence of Natural Selection on the Variability and Correlation of Organs. *Philosophical Transactions of the Royal Society of London, Series A*, vol. 200 (March 1903), pp. 1–66).

Notation:

S_1, S_2, S_3 = standard deviations in unrestricted distributions of variables 1, 2, and 3.

s_1, s_2, s_3 = standard deviations in restricted distributions of variables 1, 2, and 3.

R_{12} = correlations between variables 1 and 2, neither of them restricted.

r_{12}, r_{13}, r_{23} = correlations between variables 1 and 2, 1 and 3, 2 and 3, respectively, either or both variables restricted directly or indirectly.

Formula 1: Used to correct the correlation between variables 1 and 2 when the restriction is based on variable 1 and the ratio of the unrestricted to the restricted standard deviations of variable 1 is known. This is used in the present study to correct stanine validities since the restriction is based on the stanine and the ratio of the unrestricted and restricted stanine standard deviations can be determined.

$$R_{12}=\frac{r_{12}\frac{S_1}{s_1}}{\sqrt{1-r_{12}^2+r_{12}^2\frac{S_1^2}{s_1^2}}}$$

Formula 2: Used to correct the correlation between variables 1 and 2 when the restriction is based upon variable 3 and the ratio of the unrestricted to the restricted standard deviations of variable 3 is known. This is used in the present study to correct the correlation between a test and criterion, r_{12}, knowing the correlation between test and stanine, r_{13}, the correlation between criterion and stanine, r_{23}, and the ratio of the restricted and unrestricted stanine standard deviations.

$$R_{12}=\frac{r_{12}+r_{13}r_{23}\left(\frac{S_3^2}{s_3^2}-1\right)}{\sqrt{\left[1+r_{13}^2\left(\frac{S_3^2}{s_3^2}-1\right)\right]\left[1+r_{23}^2\left(\frac{S_3^2}{s_3^2}-1\right)\right]}}$$

selection battery were corrected for restriction of range based upon radar stanine. These corrected coefficients might be useful in providing better estimates of the validity of tests for selecting radar observers from among graduating bombardiers and navigators. Correction for range restriction might, for example, change the relative order of the tests in terms of validity. The corrected coefficients are not presented, however, because they differ less than 0.01 from the uncorrected coefficients. Even though a radar stanine of 5 or higher was required for entrance to later radar observer classes, there was little restriction of radar stanine variability in this sample. The standard deviation of radar stanines for the 137 bombardiers was 1.62, only slightly smaller than the unrestricted standard deviation of radar stanines (1.76) for all bombardiers who took the radar observer selection battery up to 1 April 1945 (N=6,813).[8] Furthermore, the correlations of radar stanine with the criterion (0.08) and with the selection tests other than these included in the radar observer selection battery (median $r=0.09$, range $=-0.11$ to 0.27) were too low to make the test validity corrections worth while. No correction for restriction due to selection in radar stanine was possible for navigators since no navigator in this study had been given the radar observer selection battery.

VALIDATION STUDY II

Validation Study II consists of the validation of experimental selection tests as well as tests from the air-crew classification and radar observer selection batteries described in Validation Study I. The criteria consist of radar observer course grades for classes graduating prior to or during the week of the cessation of hostilities on 14 August 1945. The course grade criterion for classes from Victorville is similar to that used in Validation Study I. Course grades for Boca Raton and Langley Field were computed by Psychological Research Project (Radar) wholly on the basis of scores from standardized tests and performance checks. An additional study used bombing accuracy as a criterion for the Victorville sample.

Variables

This study validates the variables in three batteries of selection tests: the air-crew classification battery (November 1943), the radar observer selection battery, and an experimental battery. Appendix A describes each of the tests of which the three batteries are composed. The report of Validation Study I describes the content of the first two batteries and their use in classification and assignment.

The experimental battery consists of 19 printed tests, 6 psychomotor tests, and a second indication of preference for radar observer

[8] Sgt. Albert Pepitone computed the unrestricted means and standard deviations of the radar stanine for bombardiers and navigators. These are presented in table 11.5; see footnotes (5) and (6).

training, Radar Preference II. The battery was administered by the Radar Project to bombardiers and navigators in the week prior to their entrance into radar-observer training. The experimental battery was not used in classification or assignment.

In all, 72 selection variables are represented in Validation Study II. From the air-crew classification battery and the radar observer selection battery come the same 27 variables validated in Study I. From the experimental battery there are 45 variables, since for the 19 printed tests, both right and wrongs were validated.

The experimental battery.—It will be remembered from chapter 3 that the first experimental selection battery was assembled entirely from tests already available. Plans at that time called for the validation of additional batteries which, is was expected, would include newly developed tests. In assembling the first battery, the decision was made to limit it primarily to tests of intellectual, perceptual, and motor abilities, reserving the major part of the projected second battery for measures of interest, personality, background, etc.

The choice of printed tests for the first experimental battery was made in terms of two contrasting approaches to selection test research. Those approaches are described in chapter 4 as the factor test approach and the job analogy test approach. It will be remembered that factor tests are those tests which are relatively independent of other measures. Individually they tend to have low validities, but because of their low intercorrelations they may in combination produce high validity. It is believed by some psychometricians that eventually a limited number of factor tests will be developed with which it will be possible to predict success on any job merely by differentially weighting the various scores.

Job analogy tests, on the other hand, individually tend to have higher validities than do factor tests. They also usually have relatively high intercorrelations since they test complex functions. Consequently, a combination of job analogy tests often yields little higher validity than the most valid single test.

It was decided to choose printed tests for the experimental battery on the basis of both the factor and job analogy view.[9] The decision to work from both points of view was based upon two considerations. First, since in the present stage of selection-test research neither approach is self-sufficient, the use of either alone was not justified. Second, considerable interest was felt in a comparison of the relative efficiency of the two approaches at their current level of development.

Psychomotor tests for the experimental battery were not selected within the above framework. Their choice was based upon an inde-

[9] Capt. Lloyd G. Humphreys and Sgt. Hyman Heller selected these tests.

pendent job description of the radar observer's task by representatives from the School of Aviation Medicine.[10]

Also in the battery was Radar Preference II, an indication of interest in radar observer training made on the Radar Student Information Blank, described in appendix A.

The printed and psychomotor tests constituting the experimental battery are listed below.

Tests chosen as factor tests.—The following tests were included in the experimental battery because of evidence that they are relatively pure measures of factors that have been definitely or tentatively found in analyses of test batteries in the Aviation Psychology Program.[11] Each test is presented with the factor it measures and the loading it has in that factor.[12]

Aerial Orientation, CP520A, is thought to be a relatively pure measure of the space I or spatial relations factor. Loading is unknown.

Area Visualization, CP815A, is thought to measure the visualization factor with a loading of 0.50.

Air Corps Vocabulary (1942) has a loading of approximately 0.70 in the verbal comprehension factor.

Compass Orientation, CI660A, is thought to represent a hypothetical factor by the same name. Loading is unknown.

Estimation of Length, CP631A, represents a postulated length estimation factor. Loading is unknown.

Flight Orientation, CP528A, is thought to be a relatively pure measure of the space I or spatial relations factor, but its loading is unknown.

Mechanical Information, CI905B, measures the mechanical experience factor, with a loading of 0.75.

Memory for Landmarks, CI510AX2, measures the memory I or rote memory factor. An almost identical form, CI510AX1, has a loading of 0.60 in this factor.

Numerical Operations, CI702BX1, represents the numerical facility factor. An older form of this test, CI702A and B, has a loading of 0.80.

Pattern Comprehension, CP803A, has a loading of approximately 0.50 on the visualization factor.

---

[10] This job description was carried out by Capt. Judson Brown and Capt. Glenn Finch.

[11] Of the tests in the November 1943 Air-Crew Classification Battery, Instrument Comprehension II, CI616B, was considered to be a factor test. It is probably the best single measure of space I with a loading of approximately 0.50. Several factors known to be measured more or less adequately by tests in the classification battery were not measured by tests included in the experimental battery. These include pilot interest, measured by General Information, CE505E; psychomotor coordination, measured by Complex Coordination, CM701A, and Rotary Pursuit, CP410B; and psychomotor precision, measured by Finger Dexterity, CM116A, and Discrimination Reaction Time, CP611D.

[12] The factor loadings presented in this section are taken primarily from analyses of the July 1943 and November 1943 air-crew classification batteries reported by Psychological Research Unit No. 3.

Position Orientation, CP526A, represents a postulated space II or rotational space factor. An earlier version, Hands, CP512A, has a loading of 0.45 on this factor.

Spatial Reasoning, CI211BX2, measures the reasoning I or general reasoning factor with a loading of 0.55.

Speed of Identification, CP610C, represents the perceptual speed factor. An older form of this test, CP610B, has a loading of approximately 0.65 on this factor.

Visual Memory, CI511A, is thought to measure a postulated factor called memory II or visual memory. A roughly similar test, Map Memory, CI505AX2, which contains line drawings instead of photographs, has a loading of 0.60 in memory II.

Tests chosen as job analogy tests.—The following tests were included in the experimental battery on the basis of similarity to a task carried out by the radar observer.[13]

Pattern Identification, CP820A, developed by the Air-Crew Evaluation and Research Detachment No. 1, resembles the observer's task of identifying patterns of returns on the PPI scope.

Polar Grid Coordinate, CP819B, presents the subject with tasks similar to those involved in reading fix data from the PPI scope and plotting it on a map.

Scale Reading, CP637A, requires reading scales that are very similar to the numerous scales read by the radar observer in carrying out his navigation and bombing tasks.

Spot Location, CP818A, presents a task which is analogous to the radar observer's task of quickly finding points in a pattern on the map which he sees in a similar pattern on the PPI scope.

Numerical Operations, CI702BX1, included in the experimental battery on the basis of relatively pure factor content, was also considered to be a job-analogy test. It requires simple arithmetic calculations similar to those carried out by the radar observer in navigation and bombing.

Ratio Estimation, CP225A, was included in the experimental battery even though it is not clearly analogous to a task of the radar

[13] Of the tests in the radar-observer selection battery, three were considered to be job-analogy tests.

Coordinate Reading, CP224A or B, requires operations identical with reading the range and bearing of a target on the PPI scope.

Oscilloscope Interpretation, CP817A, involves perceptual tasks similar to detecting returns on the PPI and A scopes.

Pattern Orientation, CP816A, requires the identification of rotated patterns which is analogous to recognizing patterns of returns on the PPI scope when azimuth stabilization is not functioning on the radar set

Of the tests in the air-crew-classification battery, two were considered to be job-analogy tests. Dial and Table Reading, CP622A and CP627A, was thought to be analogous to the radar observer's dial, scale, and table reading tasks. Spatial Orientation I, CP501B, seemed to present a problem similar to that of identifying patterns of scope returns on a target photo and vice versa.

observer. It is not included in the comparison of the factor and job-analogy approaches to be presented below.

Psychomotor Tests.—The following psychomotor tests were included in the experimental battery on the basis of the recommendation of the School of Aviation Medicine:

Check List Dial Setting, Model A (no code number).
Complex Coordination, CM701E.
Rate Control Test, CM825A.
Self-Pacing Discrimination Reaction Time, CP611E modified.
Thurstone Two-Hand Pursuit, CM810A.
Visual Coincidence, CP613B3 modified.

This battery was administered only to radar-observer students at Langley Field.

Plans for further experimental batteries.—As pointed out above, no measures of personality, attitudes, or interests were included in the experimental battery. It was planned to validate several successive batteries with increasing emphasis upon such measures. The specific tests which were planned for inclusion in a second experimental validation battery [14] are as follows:

Indices of Self-Confidence, CE427E, is a self-rating scale of psychomotor test performance.

Technical Information, CE509A, is a collection of five sports and hobbies tests including items on hunting, firearms, radio, photography, and electricity. This test was to be included on the hypothesis that amount of information of the type asked for is indicative of strength of interest.

Satisfactions Test, CE409C, is concerned with the subject's likes and dislikes.

Biographical Data Blank, CE607B, consists of personal data items taken from the CE602E and CE602F forms of the Biographical Data Blank.

The Humm-Wadsworth test measures personality traits by means of subjective questions. It was planned to validate the epileptoid, hysteroid, and autistic scales.

Directional Orientation, CP515D and E, measures ability to maintain orientation to compass directions.

Camouflaged Figures, CP801A, requires the subject to distinguish patterns from confused backgrounds.

Object Completion, CP811A, is constructed to measure ability to perceive the form of objects when only portions of their elements are seen.

---

[14] Sgt. Harding and Sgt. Heller were primarily responsible for planning the second experimental validation battery.

Penetration of Camouflage, CP817A, very similar to Camouflaged Figures, requires the detection of patterns concealed in confusing backgrounds.

Validation Criteria

Bombardier sample: Course grades.—The bombardier sample consists of students from Boca Raton and Langley Field. The radar observer course grades forming the criterion for this sample were computed by the Radar Project, entirely from scores on standardized tests and performance checks.[16] For the Langley Field students, scores from a complete battery of proficiency measures constructed by the project were available. A complete set of tests and checks was also available for the Boca Raton bombardier students, but all checks containing navigation items had been modified by personnel at the training station. This modification relieved the radar observer of many of his navigational tasks, limiting his work primarily to taking radar fixes.

Before the various test and performance check scores were weighted and combined, their raw scores were converted into stanine scores by the following steps. For each form of every test and check, a distribution of raw scores was made separately for each school. These raw scores were then converted to single-digit scores on a one-to-nine scale. The conversion was made on a percentage basis assigning a score of 9 to the highest 4 percent, 8 to the next highest 7 percent, 7 to the next 12 percent, 6 to the next 17 percent, 5 to the middle 20 percent, 4 to the next lower 17 percent, 3 to the next 12 percent, 2 to the next 7 percent and 1 to the lowest 4 percent. For distributions where it was possible closely to approximate those percentages, the conversion yielded a distribution with a mean of 5.00 and standard deviation of 2.00.

The 10 converted test and performance check scores for each student were combined into three part-course grades which, in turn, were combined into an over-all course grade. A flight grade was computed by determining the average of the final aerial check score given a weight of 2 and the intermediate aerial check score given a weight of 1. A trainer grade was computed by finding the average of the final supersonic check score, given a weight of 3, and the radar navigation intermediate supersonic, radar bombing intermediate supersonic, and final bench set check scores, each given a weight of 1. A ground grade consisted of the average of the final test score (total of Final Test I and II), given a weight of 3, and the radar navigation, radar bombing, and set operation intermediate test scores, each given a weight of 1.

---

[16] Sgt. Heller, with the assistance of S/Sgt. William J. Woywod, supervised the computation of course grades from the Langley Field data. Cpl. Robert J. Patterson, with the assistance of Cpl. Irving Fudeman and Cpl. James C. Holt, did the same work on the Boca Raton data.

The part-course scores were combined into an over-all course grade by computing the average of the flight grades, given a weight of 6, the trainer grade, given a weight of 2, and the ground grade, given a weight of 2. In computing the validity statistics course grades for Langley Field and Boca Raton students were grouped into a single bombardier sample of 629 cases. The distribution of course grades for this sample had a mean of 4.94 and a standard deviation of 1.17.

Bombardier Sample: Reliability of Course Grades.—Because no method was available for accurately determining the reliability of the course grades computed for the students at Langley Field and Boca Raton, an attempt was made to estimate the reliability. It will be noted, from the weights given the various tests, checks, and part-course grades in computing the final course grade, that the final aerial check contributes 40 percent of the course grade, the intermediate aerial check contributes 20 percent, and final supersonic check and final test each contribute 10 percent, and the remaining tests and checks each contribute 3.33 percent. To estimate the reliability of the course grade, two scores were obtained for each student. Each part score was based upon test and check scores which together made up 50 percent of the course grade. Also, the total test contribution and total check contribution was divided equally between the two part scores. On this basis, score "A" for each individual consisted of the final aerial check score weighted by 40 and the three intermediate test scores each weighted by 3.33. Score "B" consisted of the intermediate aerial check score weighted by 20, the final supersonic check score weighted by 10, the final test score weighted by 10, and the two intermediate supersonic check scores and single bench set check score, each weighted by 3.33.[16]

Only those students were included in the sample who had all test and check scores. For the Langley Field group, the bench set check scores for two radar sets, the AN/APS-15 and AN/APS-15A, were averaged and treated as a single score. The correlations between "A" and "B" scores were computed for the Langley Field and Boca Raton groups separately.

The correlations between the "A" and "B" scores for 278 students from Langley Field was 0.27. When corrected for double length, the coefficient become 0.43 which is probably a better estimate of course grade reliability than the uncorrected part-grade intercorrelation. Corresponding coefficients obtained from 117 Boca Raton students were 0.23, uncorrected, and 0.38, corrected,

Navigator Sample: Course Grades.—The navigator sample was composed of radar observer students from Victorville. The course grades used in this validation study were those assigned by the train-

[16] This work was done by Cpl. Fudeman and Cpl. Owen R. Munger.

ing station since too few test and performance check scores were available to form by themselves an adequate basis for computing course grades. Practically all the classes used in this study had taken the final and three intermediate tests. However, the only performance check scores available were for the bench set and final aerial checks, these being given to half the classes used. Consequently, the course grades used for the navigator sample are based primarily upon instructor ratings and informal written tests. Aerial, ground, and trainer grades were weighted by the training station in the ratio of 6 : 2 : 2 in computing the over-all course grade.

The aggregate scores obtained from the training station were converted to standard scores with a mean of 50 and standard deviation of 10.[17] Two such conversions were necessary, one for classes 45-27 through 45-29 and another for classes 45-30 through 45-33. This was necessitated by a shift in the grading system from class 45-29 to class 45-30. No data are available for computing the reliability of the navigator sample course grades. The distribution of 226 converted course grades for the navigator sample had a mean of 50.58 and a standard deviation of 9.94.

Navigator Sample: Circular Error.—Records of bombing accuracy were available for the navigator sample and were used as a validation criterian. The records were based upon camera bombing and summarize all drops that were photographed and scoreable. The data are analyzed from the point of view of reliability as developed in chapter 9. Various information important to an understanding of the characteristics of the circular error data used in this study and the conditions under which they were obtained is presented in table 11.6. It will be noted that there is great variability in the number of scored drops from which the average circular error is computed. The range of from 1 to 19 drops suggests the possibility that differences among students in average circular error may be due to their having had different amounts of practice. If this were true it would tend to

TABLE 11.6.—*Distribution statistics for circular error data and data descriptive of the conditions under which circular error data were obtained; navigator sample, Victorville, Camera Bombing, Validation Study II*

Data	Mean	Median	SD	Range	Number of students
Average circular error in feet	4,350	4,000	1,700	1,400 to 14,300	226
Number of drops scored per student	9.2	9.1	3.6	1 to 19	245
Number of missions per student from which scored drops were obtained	4.1	4.1	1.4	1 to 9	242
Highest altitude from which each student made a scored drop	15,669	15,808	1,076	11,000 to 19,800	242
Lowest altitude from which each student made a scored drop	11,116	10,972	1,579	7,250 to 15,000	242
Number of different targets upon which each student made scored drops	5.9	5.8	1.9	1 to 11	245

[17] S/Sgt. Richard T. Mitchell with the assistance of Cpl. James C. McClure, Jr., carried out this statistical work.

mask any relationship otherwise existing between circular error and selection test scores. To determine whether practice leading to improvement in circular error was indicated by the number of scored drops, the correlation was computed between average circular error and total number of bombing runs, both scored and unscored. The coefficient obtained, based on 240 cases, was found to be only 0.01.

The evidence presented in chapter 9 indicates the reliability of average circular error to be approximately 0.20. Such an estimate is not inconsistent with a correlation of 0.20 found between circular error and course grades. This coefficient is based upon the 226 students in the navigator sample and is significant at the 1 percent level.

Composition of Validation Samples

Only those radar observer students were included in the validation samples who were either rated navigators or bombardiers and who had taken all of their proficiency tests before 14 August 1945. Only those classes were used that had been given the experimental selection battery on entering radar observer training. The radar observer classes used from each training station are listed together with their graduation dates in table 11.7. Only bombardier students were used from Langley Field and Boca Raton and only navigators from Victorville. These restrictions left approximately 1,100 cases. This number was further reduced by 20 percent, half being eliminated because they had taken air-crew classification batteries other than the November 1943 battery and the remainder because they did not have both criterion data and radar stanines. This final restriction was made necessary by the time limitations under which the validation statistics were computed. Since each of the 72 selection variables was correlated with the criterion data and radar stanine, a great saving of time was accomplished in the International Business Machine sorting and tabulating work by using a single sample for the course grade and radar stanine correlations in the case of the bombardiers and a single sample for the course grade, circular error, and radar stanine correlations in the case of the navigators. The final

TABLE 11.7.—*Radar observer classes used from each training station in Validation Study II*

Training station	Classes	Graduation dates
Langley	45-19 through 45-33	12 May 1945-18 August 1945.
Boca Raton	335 through 625	23 May 1945-18 August 1945.
Victorville	45-27 through 45-33	7 July 1945-18 August 1945.

samples consisted of 629 bombardiers and 226 navigators. It is important to note that paralleling the difference between bombardiers and navigators is a difference in kind of criterion used in the validation. The bombardiers' course grades were computed by the Radar Project wholly on the basis of test and check scores. The navigators were assigned grades by the Victorville training authorities using the few available test and check scores to an unknown extent.

Validation Statistics

The results of the validation against course grades for the bombardier sample are presented in table 11.8.[18]

TABLE 11.8.—*Product-moment correlations between the radar observer course grade and selection variables for the bombardier sample in Validation Study II*

Variable	r_{1t} [1]	r_{2t} [2]	ρ_{1t} [3]	*N*	Mean [3]	*SD* [4]
Stanines:						
Bombardier stanine	[4] 0.13	0.18	0.14	628	7.14	1.22
Navigator stanine	[4] .14	.28	.16	629	6.43	1.36
Pilot stanine	[4] .08	.17	.09	629	5.87	1.60
Radar observer selection battery:						
Coordinate reading, CP221A or B	[4] .08	.77	.11	629	117.09	15.63
Oscilloscope interpretation, CP817A	.01	.39	.01	629	27.08	5.88
Pattern orientation, CP816A	.04	.52	.07	629	30.79	7.55
Spatial orientation I, CP501B	.07	.45	.09	629	38.35	4.56
Radar preference I	.00	−.04	.00	629	8.27	1.20
Radar stanine	.07	....	.10	629	6.62	1.19
Air-crew classification battery (printed tests):						
Biographical data blank, navigator score, CE602D	.03	.01	.04	629	22.34	3.[illegible]
Biographical data blank, pilot score, CE602D	−.05	.06	−.01	628	26.59	6.[illegible]
Dial and table reading, CP622A and CP621A	[4] .10	.26	.12	629	40.18	6.97
General Information, CE505E	[4] .10	.07	.10	629	39.14	13.72
Instrument comprehension I, CI615B	−.02	−.19	−.03	628	9.39	2.89
Instrument comprehension II, CI616B	.02	.20	.04	629	34.04	10.34
Mathematics A, CI702F	[4] .10	.12	.10	628	10.16	8.23
Mathematics B, CI206C	[4] .14	.12	.14	629	16.18	9.18
Mechanical principles, CI903B	.01	.06	.01	629	32.33	9.00
Reading comprehension, CI614H	[4] .10	.02	.10	629	21.83	12.94
Spatial orientation I, CP501B	.06	.29	.08	629	31.04	5.21
Spatial orientation II, CP503B	.03	.25	.05	629	22.81	6.24
Air crew Classification Battery (Psychomotor tests)						
Complex Coordination, CM701A	.07	.11	.08	629	54.59	9.39
Discrimination Reaction Time, CP611D	[4] .11	.08	.11	629	56.60	5.54
Finger Dexterity, CM116A	−.06	.01	−.06	629	56.22	8.96
Rotary Pursuit, CP410B	−.06	.09	−.05	629	53.40	10.03
Rudder Control, CM120B	[4] .08	.02	.08	629	50.23	9.50
Two-Hand Coordination, CM101A	.00	.02	.00	629	52.52	9.76
Experimental Printed Tests:						
Aerial Orientation, CP520:						
Rights	[4] .12	.13	.13	456	22.70	4.38
Wrongs	−.09	−.04	−.09	456	7.97	3.50
Area Visualization, CP815A:						
Rights	.02	.16	.03	410	34.15	6.97
Wrongs	.03	.05	.03	410	12.87	5.23
Air Corps Vocabulary (1942):						
Rights	.04	.07	.05	456	57.96	20.41
Wrongs	.04	.12	.05	456	26.77	11.06
Compass Orientation, CI660A:						
Rights	[4] .10	.10	.10	456	86.85	31.85
Wrongs	−.02	−.17	−.04	456	4.95	10.19
Estimation of Length, CP631A:						
Rights	.08	.04	.09	359	29.82	8.00
Wrongs	.09	.02	.09	359	23.14	9.06
Flight Orientation, CP528A:						
Rights	.07	.11	.08	457	38.00	8.58
Wrongs	−.08	−.02	−.08	457	8.72	6.23

See footnotes at end of table.

[18] Processing of these validation data on International Business Machine punch cards was planned and carried out by Lt. Roshal with the assistance of S/Sgt. Johnston, Sgt. Mangan, Sgt. Sheldon H. Nerby, and Cpl. Wilbert H. Schwotzer. The correlation coefficients were computed by Cpl. Arlene E. Babcock, Sgt. Blum, S/Sgt. Johnston, Sgt. Nerby, Sgt. Raush, and Cpl. Schwotzer.

TABLE 11.8.—*Product-moment correlations between the radar observer course grade and selection variables for the bombardier sample in Validation Study II*—Continued

Variable	r_{12} [1]	r_{13} [1]	r_{23} [1]	N	Mean [3]	SD [3]
Experimental Printed Tests—Continued						
Mechanical Information, CI905B:						
Rights	0.04	0.05	0.04	444	15.62	6.54
Wrongs	−.01	−.05	−.02	444	12.14	6.10
Memory for Landmarks, CI510AX2:						
Rights	.10	.17	.11	361	22.70	7.57
Wrongs	−.06	−.03	−.06	361	8.32	6.52
Numerical Operations, CI702BX1:						
Rights	[4] .16	.28	.18	411	35.75	9.01
Wrongs	−.03	−.02	−.04	412	4.88	3.12
Pattern Comprehension, CP803A:						
Rights	.02	.31	.04	407	18.20	5.88
Wrongs	−.01	−.05	−.01	407	9.13	5.67
Pattern Identification, CP820A:						
Rights	.07	.23	.09	407	28.65	11.29
Wrongs	[5] −.11	−.10	−.12	407	10.70	7.33
Polar Grid Coordinate, CP819B:						
Rights	[5] .10	.42	.12	432	52.35	12.40
Wrongs	.02	.19	.03	432	27.52	8.93
Position Orientation, CP526A:						
Rights	.08	.32	.10	455	211.29	47.27
Wrongs	−.02	.03	−.02	456	12.22	10.79
Ratio Estimation, CP225A:						
Rights	[4] .16	.19	.17	407	37.41	12.14
Wrongs	.02	.05	.02	407	26.55	9.80
Scale Reading, CP637A:						
Rights	[4] .15	.36	.17	431	44.84	8.53
Wrongs	−.05	−.05	−.05	431	8.65	4.94
Spatial Reasoning, CI211BX2:						
Rights	.08	.19	.10	390	30.84	14.68
Wrongs	−.07	.08	−.06	390	14.41	11.20
Speed of Identification, CP610C:						
Rights	[5] .12	.25	.13	446	73.34	11.29
Wrongs	.03	.04	.04	446	3.47	4.63
Spot Location, CP818A:						
Rights	.08	.33	.10	408	53.63	11.97
Wrongs	.08	.01	.08	408	9.97	6.05
Visual Memory, CI514A:						
Rights	[4] .13	.07	.14	416	70.57	12.54
Wrongs	[5] −.10	−.10	−.11	445	33.62	8.60
Radar Preference II	.00	.11	.01	428	1.78	0.62
Experimental Psychomotor Tests:						
Check List Dial Setting, Model A (No Code No.)	.06	.06	.07	359	77.60	9.13
Complex Coordination, CM701E	.07	.12	.08	358	55.16	7.17
Rate Control, CM825A	.04	.07	.05	358	402.40	62.01
Self-Pacing Discrimination Reaction Time, CP611E Modified	.07	.06	.07	319	413.68	44.25
Thurstone Two-Hand Pursuit, CM810A	.04	.01	.04	361	627.57	70.32
Visual Coincidence CP613B3 Modified	−.03	.07	−.03	359	110.73	12.12

[1] 1=selection variable, 2=course grade, and 3=radar stanine.

[2] Corrected for range restriction based on radar stanine. See footnote 7 on page 236 for formulas used. The unrestricted standard deviation of radar stanine was 1.76 based on 6,513 bombardiers who took the radar observer selection battery prior to 1 April 1945.

[3] The means and standard deviations of the stanines and air-crew classification battery tests obtained from samples of unclassified aviation students are presented in table 11.5. These may be compared with the means and standard deviations of the present samples to indicate the amount of selection that operated on the students prior to entering radar observer training. The majority of the students in the present samples were tested at the Medical and Psychological Examining Units so the statistics should be compared with those obtained from Medical and Psychological Unit administrations of the battery.

[4] Significant at the 1-percent level.

[5] Significant at the 5-percent level.

[6] The significance of the validity coefficients was tested before they were rounded to two figures. This coefficient and later ones appear to be significant after rounding but were not significant when tested at three decimal places.

Similar data for the navigator sample are presented in table 11.9 and the statistics from the validation against average circular error are given in table 11.10. Signs of all correlations involving circular error have been changed so that a positive coefficient indicates association of goodness of circular error with goodness of test score. The significance of these data, as well as of the multiple regression data to follow, will be discussed in the last section of this chapter. There,

TABLE 11.9.—*Product moment correlations between the radar observer course grade and selection variables for the navigator sample in Validation Study II*

Variable	r_{t1}[1]	r_{t1}[1]	r_{t1}[1]	*N*	Mean[1]	*SD*[1]
Stanines:						
Bombardier stanine	[4] 0.15	0.14	0.18	226	7.36	1.45
Navigator stanine	.05	.18	.10	226	7.85	1.65
Pilot stanine	.05	.12	.08	226	6.35	1.72
Radar Observer Selection Battery:						
Coordinate Reading, CP224A or B	.08	.78	.19	226	133.44	16.27
Oscilloscope Interpretation, CP817A	.08	.32	.17	226	31.85	6.02
Pattern Orientation, CP816A	[4] .18	.39	.25	226	35.46	6.92
Spatial Orientation I, CP501B	.10	.43	.18	226	40.82	4.20
Radar Preference I	−.01	−.24	−.09	226	8.88	.44
Radar stanine	.13	....	.22	226	8.06	1.00
Aircrew Classification Battery (printed tests):						
Biographical Data Blank, Navigator Score, CE602D	−.12	.05	−.10	226	23.76	3.93
Biographical Data Blank, Pilot Score, CE602	.01	.04	.02	226	27.67	6.64
Dial and Table Reading, CP622A and CP621A	[4] .16	.30	.22	226	44.27	7.45
General Information, CE505E	.03	−.09	.01	226	41.69	15.12
Instrument Comprehension I, CI615B	.00	−.15	−.04	226	8.08	2.95
Instrument Comprehension II, CI616B	.04	.14	.08	226	36.97	11.30
Mathematics A, CI702F	−.01	−.02	−.02	226	18.10	9.04
Mathematics B, CI206C	[4] .16	.04	.17	226	23.15	10.51
Mechanical Principles, CP903B	−.02	.05	−.01	226	35.45	9.51
Reading Comprehension, CI614H	.08	.07	.10	226	29.70	14.65
Spatial Orientation I, CP501B	.06	.32	.13	226	33.88	4.72
Spatial Orientation II, CP603B	−.08	.26	.00	226	26.13	6.00
Aircrew Classification Battery (psychomotor tests):						
Complex Coordination, CM701A	.11	.13	.14	226	55.18	10.51
Discrimination Reaction Time, CP611D	−.04	.10	−.01	226	55.87	5.99
Finger Dexterity, CM116A	.12	.02	.13	226	51.72	10.00
Rotary Pursuit, CP410B	−.01	.02	.00	226	51.33	9.59
Rudder Control, CM120B	.09	.09	−.01	226	50.60	9.78
Two-Hand Coordination, CM101A	.06	.03	.07	226	53.42	9.83
Experimental Printed Tests:						
Aerial Orientation, CP520A:						
Rights	.00	.17	.05	179	23.12	3.59
Wrongs	−.04	−.04	−.05	179	8.28	3.49
Area Visualization, CP815A:						
Rights	....	....	....	34	....	....
Wrongs	....	....	....	34	....	....
Air Corps Vocabulary (1942):						
Rights	....	....	....	0	....	....
Wrongs	....	....	....	0	....	....
Compass Orientation, CI660A:						
Rights	....	....	....	0	....	....
Wrongs	....	....	....	0	....	....
Estimation of Length, CP631A:						
Rights	−.07	.36	.03	179	33.50	8.71
Wrongs	.10	.10	.12	179	27.93	9.57
Flight Orientation, CP523A:						
Rights	−.11	−.16	−.15	35	40.74	7.87
Wrongs	.19	.09	.21	35	6.60	4.13
Mechanical Information, CI905B:						
Rights	.12	.04	.13	226	16.95	6.67
Wrongs	[4] −.14	−.01	−.14	226	11.08	5.94
Memory for Landmarks, CI510AX2:						
Rights	....	....	....	0	....	....
Wrongs	....	....	....	0	....	....
Numerical Operations, CI702BX1:						
Rights	.15	.16	.19	146	44.55	8.76
Wrongs	−.09	.13	−.06	146	5.34	2.43
Pattern Comprehension, CP803A:						
Rights	.06	.16	.10	178	21.44	5.66
Wrongs	−.03	.02	−.03	178	8.20	5.65
Pattern Identification, CP820A:						
Rights	.06	.20	.11	44	40.41	7.73
Wrongs	.10	−.21	−.05	44	7.18	6.14
Polar Grid Coordinate, CP819B:						
Rights	.04	[4] .44	.14	115	62.36	12.10
Wrongs	.09	.05	.10	115	32.26	7.61
Position Orientation, CP626A:						
Rights	.02	.25	.08	179	219.16	39.42
Wrongs	−.01	.00	−.01	178	11.65	13.25
Ratio Estimation, CP223A:						
Rights	[4] .48	.31	.24	150	43.95	11.86
Wrongs	−.09	.04	−.08	150	26.74	8.73
Scale Reading, CP637A:						
Rights	.09	.45	.17	146	51.79	9.02
Wrongs	−.13	.01	−.12	146	8.04	4.70
Spatial Reasoning, CI211BX2:						
Rights	[4] .26	.11	.28	75	37.03	15.04
Wrongs	−.21	−.06	−.22	76	17.03	10.79

See footnotes at end of table.

TABLE 11.9.—*Product-moment correlations between the radar observer course grade and selection variables for the navigator sample in Validation Study II*—Continued

Variable	r_{12} [1]	r_{13} [1]	r_{23} [1]	N	Mean [3]	SD [3]
Experimental Printed Tests—Continued						
Speed of Identification, CP810C:						
Rights	−0.10	0.26	−0.03	146	75.99	8.13
Wrongs	−.01	.07	−.02	146	1.57	1.10
Spot Location, CP818A:						
Rights	.01	.44	.11	146	60.25	9.57
Wrongs	.02	.03	.03	146	10.41	7.30
Visual Memory, CI514A:						
Rights	....	....	....	40	....	....
Wrongs	....	....	....	40	....	....

[1] 1=selection variable, 2=course grade, and 3=radar stanine.
[2] Corrected for range restriction based on radar stanine. See footnote 7 on page 234 for formulas used. Unrestricted standard deviation of radar stanine was 1.77 based on 3,703 navigators who took the radar observer selection battery prior to 1 April 1945.
[3] See footnote 3 to table 11.8.
[4] Significant at the 5-percent level.
[5] Significant at the 1-percent level.

TABLE 11.10.—*Product-moment correlations between average circular error and selection variables for the navigator sample in Validation Study II*

Variable	r_{12} [1]	r_{13} [1]	r_{23} [1]	N	Mean [3]	SD [3]
Stanines:						
Bombardier stanine	[4]0.16	0.14	0.10	226	7.30	1.45
Navigator stanine	[5].25	.18	.24	226	7.85	1.08
Pilot stanine	.10	.12	.11	226	6.35	1.72
Radar Observer Selection Battery:						
Coordinate Reading, CP221A or B	.00	.78	.02	226	133.14	16.27
Oscilloscope Interpretation, CP817A	.09	.32	.09	226	31.85	6.02
Pattern Orientation, CP816A	.03	.39	.04	226	38.46	6.02
Spatial Orientation I, CP501B	−.04	.43	−.02	226	40.82	4.20
Radar Preference I	.11	−.24	.10	226	8.88	.44
Radar Stanine	.02	....	.03	226	8.00	1.00
Aircrew Classification Battery (printed tests):						
Biographical Data Blank, Navigator Score, CE602D	.10	.05	.10	226	23.76	3.93
Biographical Data Blank, Pilot Score, CE602D	.09	.04	.10	226	27.07	6.64
Dial and Table Reading, CP622A and CP621A	[5].18	.30	.17	226	41.27	7.45
General Information, CE505E	.10	−.09	.10	226	41.69	15.12
Instrument Comprehension I, CI615B	.00	−.15	.00	226	8.08	2.95
Instrument Comprehension II, CI616B	.12	.14	.12	226	36.97	11.20
Mathematics A, CI702F	[4].16	−.02	.16	226	18.10	9.04
Mathematics B, CI206C	[4].14	.04	.14	226	24.15	10.51
Mechanical Principles, CP903B	.11	.05	.11	226	35.45	9.51
Reading Comprehension, CI614H	.11	.07	.11	226	29.70	14.65
Spatial Orientation I, CP501B	.09	.32	.09	226	33.88	4.72
Spatial Orientation II, CP503B	−.01	.26	.00	226	26.13	6.00
Aircrew Classification Battery (psychomotor tests):						
Complex Coordination, CM701A	.03	.13	.03	226	28.18	10.81
Discrimination Reaction Time, CP611D	.01	.10	.02	226	55.87	5.99
Finger Dexterity, CM116A	.03	.02	.01	226	54.72	10.00
Rotary Pursuit, CP410B	.05	.02	.05	226	51.23	9.89
Rudder Control, CM120B	.01	.00	.01	226	50.60	9.78
Two-Hand Coordination, CM101A	.11	.03	.11	226	53.42	9.83
Experimental Printed Tests:						
Aerial Orientation, CP520A:						
Rights	.12	.17	.13	179	23.12	3.59
Wrongs	−.04	−.04	−.04	179	8.25	3.49
Area Visualization, CP815A:						
Rights	....	....	....	34	....	....
Wrongs	....	....	....	34	....	....
Air Corps Vocabulary (1942):						
Rights	....	....	....	0	....	....
Wrongs	....	....	....	0	....	....
Compass Orientation, CI670A:						
Rights	....	....	....	0	....	....
Wrongs	....	....	....	0	....	....
Estimation of Length, CP631A:						
Rights	−.04	.35	−.03	179	33.70	8.71
Wrongs	.12	.10	.12	179	27.93	9.37
Flight Orientation, CP528A:						
Rights	.02	−.16	.01	35	40.74	7.37
Wrongs	.25	.09	.26	35	6.60	4.13

See footnotes at end of table.

TABLE 11.10. —*Product-moment correlations between average circular error and selection variables for the navigator sample in Validation Study II*—Con.

Variable	r_{12}[1]	r_{13}[1]	$_cr_{12}$[2]	*N*	Mean[3]	*SD*[3]
Experimental Printed Tests—Continued						
Mechanical Information, CI905B:						
Rights	0.10	0.04	0.11	226	16.95	6.67
Wrongs	-.10	-.01	-.10	226	11.98	5.94
Memory for Landmarks, CI510AX2:						
Rights	....	....	....	0	....	....
Wrongs	....	....	....	0	....	....
Numerical Operations, CI702BX1:						
Rights	-.04	.16	-.04	146	41.55	8.76
Wrongs	.06	.13	.06	116	5.34	2.43
Pattern Comprehension, CP803A:						
Rights	-.09	.16	-.08	178	21.44	5.66
Wrongs	.03	.02	.03	178	8.20	5.65
Pattern Identification, CP820A:						
Rights	.04	.20	.04	44	40.41	7.73
Wrongs	-.02	-.21	-.02	44	7.18	6.14
Polar Grid Coordinate, CP819B:						
Rights	.09	.44	.09	115	62.36	12.10
Wrongs	-.11	.05	-.11	115	32.26	7.61
…n Orientation CP520A:						
…ghts	.05	.25	.05	179	219.16	39.42
Wrongs	-.01	.00	-.01	178	11.65	13.25
Ratio Estimation, CP225A:						
Rights	.13	.31	.13	150	43.95	11.66
Wrongs	.02	.04	.02	150	26.74	8.73
Scale Reading, CP037A:						
Rights	.05	.45	.06	146	51.79	9.02
Wrongs	.05	.04	.05	146	8.04	4.70
Spatial Reasoning, CI211BX2:						
Rights	.06	.11	.06	76	37.03	15.06
Wrongs	-.09	-.06	-.09	76	17.03	10.79
Speed of Identification, CP610C:						
Rights	-.16	.23	-.14	146	75.09	8.13
Wrongs	.02	.07	.02	146	1.67	1.10
Spot Location, CP818A:						
Rights	.02	.44	.03	146	60.25	9.57
Wrongs	-.08	.03	-.08	146	10.41	7.30
Visual Memory, CI514A:						
Rights	....	....	....	40	....	....
Wrongs	....	....	....	40	....	....
Radar Preference II	....	....	....	....	....	....

[1] 1=selection variables, 2=average circular error, and 3=radar stanine.
[2] Corrected for range restriction based on radar stanine. See footnote 7 on page 236 for formulas used. Uncorrected st[illegible]rd deviations of radar stanine was 1.77 based on 3,701 navigators who took the radar observer selection battery prior to 1 April 1945.
[3] See footnote 3 to Table 11.8.
[4] Significant at the 5-percent level
[5] Significant at the 1-percent level.

evidence will be summarized in terms of which abilities are indicated to be important to success in radar observer training.

A comparison of tables 11.8 and 11.9 shows that more statistically significant correlations were found for the bombardier sample than for the navigators. Nine coefficients are significant at the 1-percent level for bombardiers and an additional 11 are significant at the 5-percent level. Only one of the navigator coefficients reached the 1-percent level and an additional six reached the 5-percent level. In interpreting these tables, several differences between the two samples must be remembered. The bombardier sample is relatively large (N=629) while the navigator sample is much smaller (N=226). Also, the criterion for the bombardiers is probably more reliable than that for the navigators since the former is based to a greater extent upon scores from standardized tests and performance checks. The less reliable criterion would cause an attenuation of the validity coefficients for the navigator sample. It is further to be noted that the navigators are probably more highly selected than bombardiers on the basis of psychological functions that are important to success in radar observer training. The restriction of range on valid functions would attenuate validity coefficients for tests measuring those functions. There is some evidence for this in the job analysis of the radar observer which indicates that the functions involved are more closely related to navigation than to bombing. Further evidence comes from the statistically significant difference in average course grade reported in Validation Study I, which is in favor of the navigators.

Multiple Regression Statistics

Selection of variables.—Two multiple regression equations were computed, one for the bombardiers and another for the navigators, both with the course grade criterion.[19] Variables were selected for inclusion in these regression equations primarily on the basis of having statistically significant correlation with the criterion. The bombardier and navigator stanines were included for both samples although the navigator stanine did not have significant validity for the navigator sample against the course grade criterion in the second study. The navigator stanine did correlate significantly with average circular error and, in the first validation study, with course grade.

Tests were included in the equations if their uncorrected correlation with the course grade criterion was significant at the 5-percent or 1-percent level and was based upon more than 145 cases. Experimental tests, for which both total right and total wrong scores were validated, were included if either the rights or wrongs correlated significantly with the criterion. The scores obtained from the administration of Spatial Orientation I, CP501B, in the radar observer selection battery were included in both equations because they almost reached the 5-percent level of significance, in both validation studies and for all samples. Compass Orientation, CI660A, and Pattern Identification, CP820A, were not included in the computations for the bombardier sample although they had significant validities. Their inclusion would have greatly increased the number of International Business Machine tabulator runs. They were eliminated as the tests nearest the 5 percent level of significance. Three tests without significant validity, Coordinate Reading, CP224A or B, Numerical Operations, CI702BX1, and Scale Reading, CP637A, were included for the navigator sample partly because they were the only tests whose validities became relatively large when corrected for range restriction based on radar stanine and partly because it was impossible to include them with little increase in calculating time.

---

[19] These statistics were computed by S/Sgt. Johnston, Cpl. Kelley, and Capt. William F. Long.

Treatment of rights and wrongs scores.—It is well known that a test score composed of weighted rights and wrongs scores often has higher validity than the rights score alone. To investigate this possibility, correlations between rights and wrongs scores were obtained for all experimental tests used in the bombardier regression equation. Zero-order multiple correlations were computed for each test yielding the validity that would be obtained by giving rights and wrongs scores optimal weights. The data used in these computations and the resulting multiple correlations are presented in table 11.11. It will be noted that, with the exception of Visual Memory, the validity coefficients for the best combinations of rights and wrongs scores are equal to or only very little higher than the validities of the rights scores alone. In view of these results, it was decided to use simply the number of right responses as the score on each experimental test.

TABLE 11.11.—*Multiple correlations between radar observer course grade and the best weighted combinations of rights and wrongs scores from selected tests used for bombardier sample in Validation Study II*

Variable	Uncorrected coefficients				Corrected coefficients [1]			
	r_{23}[2]	r_{13}[2]	r_{12}	$R_{1.23}$	r_{23}	r_{13}	r_{12}	$R_{1.23}$
Aerial Orientation, CP520A	−0.89	−0.09	0.12	0.13	−0.88	−0.09	0.13	0.14
Numerical Operations, CI702HX1	.04	−.03	.16	.17	.03	−.04	.18	.18
Polar Grid Coordinate, CP819B	.03	.02	.10	.10	.11	.03	.12	.12
Ratio Estimation, CP225A	−.08	.02	.16	.16	−.07	.02	.17	.17
Scale Reading, CP637A	−.13	−.05	.15	.15	−.14	−.05	.17	.17
Speed of Identification, CP610C	.03	.03	.12	.12	.04	.04	.13	.14
Visual Memory, CI514A	.03	−.10	.13	.17	.02	−.11	.14	.18

[1] Corrected for range restriction based on radar stanine. See footnote (2) to table 11.8.
[2] 1=course grade, 2=rights score, and 3=wrongs score.

Multiple regression statistics for bombardier sample.—The uncorrected intercorrelations and validities for all variables used in the bombardier multiple regression equations are presented in table 11.12. The corresponding coefficients, corrected for range restriction based on radar stanine, are presented in table 11.13. The beta weights for three regression equations were computed for the bombardier sample, one including the bombardier and navigator stanines, a second including 15 tests, the selection of which is described above, and a third including both tests and stanines. The beta weights and multiple correlation coefficients resulting from each of these combinations are presented in table 11.14. The equations were computed both from uncorrected coefficients and from coefficients corrected for range restriction based on radar stanine. As pointed out in a footnote to this table and to tables 11.17 and 11.19 to follow, the multiple coefficients have been corrected for expected shrinkage. However, the correction made is an underestimate, since the available formula does not cover the case in which the multiple is based upon a smaller number of more valid variables selected from a larger total group. The corrections are presented in the absence of a more accurate estimate.

TABLE 11.12.—*Uncorrected intercorrelations and validities of variables used in computing multiple regression statistics for bombardier sample, Validation Study II*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Course grade
1. Bombardier stanine	.....	0.45	0.11	0.01	0.32	0.43	0.15	0.18	0.25	0.07	0.05	−0.07	0.25	0.10	−0.12	−0.05	−0.02	0.13
2. Navigator stanine	0.45	.....	.19	.15	.50	.14	.25	.66	.52	.28	.16	.11	.49	.28	.12	.07	.08	[illegible]
3. Aerial Orientation, CP501A	.11	.19	.....	−.16	.00	−.08	.13	.14	.11	−.05	.08	.06	.17	.10	−.02	−.37	.06	[illegible]
4. [illegible], CP[illegible]A or B	.01	.15	−.16	.....	.16	−.31	−.05	.11	.14	.19	.29	.06	.00	.14	−.03	−.04	−.10	[illegible]
5. Dial and Table Reading, CP622A, CP621A	.32	.50	.00	.16	.....	−.07	−.05	.20	.19	.32	.18	.14	.10	.28	.05	.19	.03	.10
6. Discrimination Reaction Time, CP611D	.43	.14	−.08	−.31	−.07	.....	−.07	.13	.00	−.11	−.14	−.17	.02	−.05	−.29	−.24	−.12	.11
7. General Information, CE505E	.15	.25	.13	−.05	−.05	−.07	.....	.12	.14	−.14	.02	−.02	.32	−.08	−.07	−.22	−.04	.10
8. Mathematics A, CI502F	.18	.66	.14	.11	.20	.13	.12	.....	.46	.21	.04	.08	.47	.21	.02	.03	.04	.10
9. Mathematics B, CI206C	.25	.52	.11	.14	.19	.00	.14	.46	.....	.31	.02	.11	.48	.25	−.11	−.16	.00	.14
10. Numerical Operations	.07	.28	−.05	.19	.32	−.11	−.14	.21	.31	.....	.24	.18	.12	.42	.01	.25	.17	.16
11. Polar Grid Coordinate, CP810B	.05	.16	.08	.29	.18	−.14	.02	.04	.02	.24	.....	.17	.18	.34	.02	.14	.15	.10
12. Rate Estimation, C[illegible]22A	−.07	.11	.06	.06	.14	−.17	−.02	.08	.11	.18	.17	.....	.11	.22	−.01	.22	.09	.16
13. Reading Comprehension, CI614H	.25	.49	.17	.00	.10	.02	.32	.47	.48	.12	.18	.11	.....	.18	−.07	−.04	.06	.10
14. Scale Reading, C[illegible]A	.10	.28	.10	.14	.28	−.05	−.08	.21	.25	.42	.34	.22	.18	.....	.02	.23	.17	[illegible]
15. Spatial Orientation I[1] CP501B	−.12	.12	−.02	−.03	.05	−.29	−.07	.02	−.11	.01	.02	−.01	−.07	.02	.....	.08	−.01	.07
16. Speed of Identification, CP610C	−.05	.07	−.37	−.04	.19	−.24	−.22	.03	−.16	.25	.14	.22	−.04	.23	.08	.....	.08	.12
17. Visual Memory, CP234A	−.02	.08	.06	−.10	.03	−.12	−.08	.04	.00	.17	.15	.09	.06	.17	−.01	.08	.....	.13

[1] Based on scores from administration of Spatial Orientation I in the Radar Observer Selection Battery.

TABLE 11.13.—*Corrected intercorrelations and validities of variables used in computing multiple regression statistics for bombardier sample, Validation Study II*[1]

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Course grade
1. Bombardier stanine	.....	0.50	0.13	0.14	0.36	0.44	0.16	0.30	0.27	0.13	0.13	−0.03	0.25	0.16	−0.02	.10	.09	0.14
2. Navigator stanine	0.50	.....	.22	.30	.54	.16	.26	.67	.53	.35	.26	.17	.47	.26	.24	.14	.10	.16
3. Aerial Orientation, CP501A	.13	.22	.....	−.03	.04	−.06	.14	.15	.13	−.01	.12	.08	.17	.14	.04	.10	.07	.13
4. [illegible], CP[illegible] or B	.14	.30	−.03	.....	.29	−.18	.01	.17	.19	.33	.47	.18	.02	.34	.26	.14	−.03	.11
5. Dial and Table Reading, CP622A, CP621A	.36	.54	.01	.29	.....	−.04	−.03	.23	.22	.37	.27	.19	.10	.35	.16	.24	.05	.12
6. Discrimination Reaction Time, CP611D	.44	.16	−.06	−.18	−.04	.....	−.06	.14	.01	−.08	−.09	−.15	.02	−.02	−.22	−.21	−.11	.11
7. General Information, CE505E	.16	.26	.14	.01	−.03	−.06	.....	.12	.15	−.11	.05	−.01	.32	−.04	−.03	−.19	−.04	.10
8. Mathematics A, CI[illegible]	.30	.67	.15	.17	.23	.14	.12	.....	.47	.21	.09	.11	.47	.24	.08	.07	.05	.10
9. Mathematics B, CI[illegible]	.27	.53	.13	.19	.22	.01	.15	.47	.....	.33	.08	.13	.48	.27	−.04	−.12	.06	.14
10. Numerical Operations, CI[illegible]	.13	.35	−.01	.23	.37	−.08	−.11	.24	.33	.....	.33	.23	.12	.48	.14	.31	.18	.15
11. Polar Grid Coordinate, CP810B	.13	.26	.12	.47	.27	−.09	.05	.09	.08	.33	.....	.25	.06	.43	.30	.23	.17	.12
12. Rate Estimation, C[illegible]A	−.03	.17	.08	.15	.19	−.15	−.01	.11	.13	.23	.23	.....	.11	.27	.08	.26	.10	.17
13. Reading Comprehension, CI614H	.25	.47	.17	.02	.10	.02	.32	.47	.48	.12	.06	.11	.....	.17	−.06	−.03	.06	.10
14. Scale Reading, C[illegible]A	.16	.26	.14	.23	.35	−.02	−.04	.24	.27	.48	.43	.27	.17	.....	.18	.20	.15	.17
15. Spatial Orientation I[2] CP501B	−.02	.23	.04	.26	.16	−.22	−.03	.08	−.04	.14	.30	.08	−.06	.18	.....	.14	[illegible]	.09
16. Speed of Identification, CP610C	.10	.14	.10	.14	.24	−.21	−.19	.07	−.12	.31	.23	.26	−.03	.20	.14	.....	.09	.13
17. Visual Memory, CP234A	.09	.10	.07	−.03	.05	−.11	−.04	.05	.06	.18	.17	.10	.06	.15	.03	.09	.....	.14

[1] Coefficients are corrected for range restriction based on radar stanine. See footnote (2) to table 11.8.
[2] Based on scores from administration of Spatial Orientation I in the Radar Observer Selection Battery.

TABLE 11.14.—*Multiple regression statistics for bombardiers, Validation Study II*

Variable	Stanines beta weights		Tests beta weights		All variables beta weights	
	Uncorrected data [1]	Corrected data [1]	Uncorrected data [1]	Corrected data [2]	Uncorrected data [1]	Corrected data [1]
Bombardier stanine	0.08	0.08	......	......	−0.05	−0.01
Navigator stanine	10	.11	......	......	−.45	−.28
Aerial Orientation, CP530A	......	......	0.17	0.11	.24	.14
Coordinate Reading, CP221A or B	......	......	.26	.09	.36	.11
Dial and Table Reading, CP622A, CP621A	......	......	.01	.00	.18	.10
Discrimination Reaction Time, CP611D	......	......	.40	.24	.50	.31
General Information, CE505E	......	......	.20	.14	.32	.20
Mathematics A, CI702F	......	......	−.14	−.08	−.01	.01
Mathematics B, CI206C	......	......	.15	.12	.28	.19
Numerical Operations, CI702BX1	......	......	.07	.07	.07	.07
Polar Grid Coordinate, CP819B	......	......	−.04	−.04	−.06	−.04
Ratio Estimation, CP225A	......	......	.13	.11	.12	.10
Reading Comprehension, CI614B	......	......	−.01	−.01	.04	.02
Scale Reading, CP637A	......	......	−.01	−.01	−.02	−.02
Spatial Orientation I, CP501B [3]	......	......	.22	.10	.34	.15
Speed of Identification, CP610C	......	......	.25	.16	.33	.19
Visual Memory, CP514A	......	......	.18	.14	.23	[illegible]
Multiple R	.16	.17	.45	.38	.51	.41
Shrunken multiple R [4]	.15	.16	.41	.33	.47	.35

[1] Computed from the coefficients given in table 11.12.

[2] Computed from the coefficients given in table 11.13.

[3] Based on scores from administration of Spatial Orientation I in the Radar Observer Selection Battery.

[4] Corrected for shrinkage by the formula $R'_{1\,n\ldots m} = \frac{R^2_{1\,n\ldots m}(N-1)-(m-1)}{N-m}$ given in Croxton, F. E. & Cowden, D. J., *Applied General Statistics*, Prentice-Hall, Inc., N. Y., 1939, p. 775. This correction gives the best estimate of the correlation in the population from which the samples was drawn. It corrects for the tendency of correlations obtained from samples to be larger than the correlation existing in the population.

Multiple regression statistics for navigator sample.—The uncorrected intercorrelations and validities for all variables used in the navigator regression equations are presented in table 11.15. The corresponding coefficients corrected for range restriction, based on radar stanine, are presented in table 11.16. The beta weights for three regression equations were computed for the navigator sample, one including the bombardier and navigator stanines, a second including nine tests, and a third including both the tests and stanines.

TABLE 11.15.—*Uncorrected intercorrelations and validities of variables used in computing multiple regression statistics for navigator sample, Validation Study II*

Variable	1	2	3	4	5	6	7	8	9	10	11	Course grade
1. Bombardier stanine	....	0.51	0.13	0.42	0.35	0.21	0.31	0.19	0.20	0.22	0.07	0.15
2. Navigator stanine	0.51	....	.02	.53	.49	−.01	.35	.24	.18	.25	.07	.05
3. Coordinate reading, CP221A or B	.13	.02	....	.39	.05	.02	.48	.15	.23	.64	.22	.08
4. Dial and table reading, CP622A, CP621A	.42	.53	.39	....	.16	−.06	.32	.22	.10	.40	.06	.16
5. Mathematics B, CI206C	.35	.49	.05	.16	....	.17	.32	.18	.16	.17	−.01	.16
6. Mechanical information, CI905B	.21	−.01	.02	−.06	.17	....	−.07	.23	.15	.06	.04	.12
7. Numerical operations, CI702BX1	.31	.35	.48	.32	.32	−.07	....	.17	.08	.40	.03	.15
8. Pattern orientation, CP816A	.19	.24	.15	.22	.18	.23	.17	....	.40	.34	.20	.15
9. Ratio estimation, CP225A	.20	.18	.23	.10	.16	.15	.08	.40	....	.31	.08	.15
10. Scale reading, CP637A	.22	.25	.64	.40	.17	.06	.40	.34	.31	....	.22	.02
11. Spatial Orientation I,[1] CP501B	.07	.07	.22	.03	−.01	.01	.03	.20	.08	.22	....	.10

[1] Based on scores from administration of Spatial Orientation I in the Radar Observer Selection Battery.

TABLE 11.16.—*Corrected intercorrelations and validities of variables used in computing multiple regression statistics for navigator sample, Validation Study II* [1]

Variable	1	2	3	4	5	6	7	8	9	10	11	Course grade
1. Bombardier stanine	...	0.56	0.24	0.46	0.36	0.25	0.31	0.27	0.26	0.30	0.16	0.13
2. Navigator stanine	0.56	...	.20	.57	.49	−.02	.39	.33	.26	.34	.19	.10
3. Coordinate reading, CP224A or B	.24	.20	...	.54	.07	.06	.48	.46	.45	.77	.52	.19
4. Dial and table reading, CP622A, CP621A	.46	.57	.54	...	.17	−.02	.38	.37	.25	.53	.26	.22
5. Mathematics B, CP206C	.36	.49	.07	.17	...	.17	.32	.19	.17	.17	.02	.17
6. Mechanical information CP905B	.25	−.02	.06	−.02	.17	...	−.05	.21	.17	.09	.07	.13
7. Numerical operations, CI702BX1	.31	.39	.48	.38	.32	−.05	...	.26	.17	.45	.14	[illegible]
8. Pattern orientation, CP816A	.27	.33	.46	.37	.19	.21	.26	...	.62	.52	.41	[illegible]
9. Ratio estimation, CP225A	.26	.26	.45	.25	.17	.17	.17	.52	...	.46	.28	.21
10. Scale reading, CP637A	.30	.34	.77	.53	.17	.09	.45	.52	.46	...	.45	.17
11. Spatial orientation I,[2] CP501B	.16	.19	.52	.26	.02	.07	.14	.41	.28	.45	...	.18

[1] Coefficients are corrected for range restriction based on radar stanine. See footnote 4 to table 11.8.

[2] Based on scores from administrations of Spatial Orientation I in the Radar Observer Selection Battery.

These weights and the multiple correlation coefficients resulting from each combination are listed in table 11.17. These data and the beta weights computed for the bombardier sample will be discussed in the last section of this chapter in a consideration of the abilities necessary to successful completion of the radar observer's course.

TABLE 11.17.—*Multiple regression statistics for navigators, Validation Study II*

	Stanines beta weights		Tests beta weights		All variables beta weights	
	Uncorrected data [1]	Corrected data [2]	Uncorrected data [1]	Corrected data [2]	Uncorrected data [1]	Corrected data [2]
Bombardier stanine	0.18	0.19	...	...	0.06	[illegible]
Navigator stanine	−.04	−.01	...	...	−.23	[illegible]
Coordinate reading, CP224A or B	...	...	−0.03	−0.02	−.15	[illegible]
Dial and table reading, CP622A, CP621A	...	...	.13	.14	.25	[illegible]
Mathematics B, CI206C	...	...	.07	.07	.15	.13
Mechanical information, CI905B	...	...	.10	.09	.06	[illegible]
Numerical operations, CI702BX1	...	...	.12	.13	.18	[illegible]
Pattern orientation, CP816A	...	...	.06	.08	.07	[illegible]
Ratio estimation, CP225A	...	...	.14	.15	.19	[illegible]
Scale reading, CP637A	...	...	−.10	−.11	−.05	[illegible]
Spatial orientation I,[3] CP501B	...	...	.09	.11	.11	[illegible]
Multiple R	.16	.18	.31	.34	.36	[illegible]
Shrunken multiple R [4]	.13	.16	.19	.20	.21	[illegible]

[1] Computed from the coefficients given in table 11.15.

[2] Computed from the coefficients given in table 11.16.

[3] Based on scores from administration of Spatial Orientation I in the Radar Observer Selection Battery.

[4] See footnote 4, Table 11.14.

Evaluation of job analogy and factor approaches to selection.—While data from a single investigation can provide no definite comparison of the job analogy and factor approaches to selection test research, it appeared of interest to test the relative efficiency of the two methods in this study of radar observer selection. Two additional sets of multiple regression statistics were computed. One of these included the valid tests among those considered to be factor tests, listed on pages 206 to 207 and in footnote 11 on page 206. The other set included the valid tests among those considered to be job analogy tests, listed on pages 207 to 208 and in footnote 13 on page 207. The comparison was made only for the bombardier sample since too few cases were available for seven of the experimental factor tests to provide for their validation in the navigation sample. Omission of that number of tests would have put the factor approach at a decided disadvantage.

The intercorrelations and validities of the two groups of tests are given in table 11.18. The derived multiple regression statistics are given in table 11.19.

TABLE 11.18.—*Intercorrelations and validities of job analogy and factor test batteries, bombardier sample, Validation Study II*

Variable	Uncorrected coefficients							Corrected coefficients [1]						
	1	2	3	4	5	6	Course grade	1	2	3	4	5	6	Course grade
1. Coordinate Reading, CP224A or B	.....	0.16	0.19	0.29	0.14	−0.03	0.08	....	0.29	0.33	0.47	0.33	0.26	0.11
2. Dial and Table Reading, CP621A	0.16	....	.32	.18	.28	.05	.10	0.29	....	.37	.27	.35	.16	.12
3. Numerical Operations, CI702BX1	.19	.32	....	.21	.42	.01	.16	.33	.37	....	.33	.48	.14	.18
4. Polar Grid Coordinate, CP819B	.29	.18	.24	....	.34	.02	.10	.47	.27	.33	....	.44	.20	.12
5. Scale Reading, CI637A	.14	.28	.42	.34	....	.02	.15	.33	.35	.48	.44	....	.18	.17
6. Spatial Orientation, I ([2]) CP501B	−.03	.05	.01	.02	.02	.....	.07	.26	.16	.14	.20	.18	....	.09

	1	2	3	4	Course grade	1	2	3	4	Course grade
1. Aerial Orientation, CP520A	......	−0.05	−0.04	0.06	0.12	......	−0.01	0.00	0.07	0.13
2. Numerical Operations, CI702BX1	−0.05	......	.25	.17	.16	−0.01	......	.31	.18	.18
3. Speed of Identification, CP610C	−.04	.25	......	.08	.12	.00	.31	......	.09	.13
4. Visual Memory CI514A	.00	.17	.08	......	.13	.07	.18	.09	......	.14

[1] Corrected for range restriction based on Radar Observer Selection Battery. See footnote (2) to table 11.8.
[2] Based on scores from administration of Spatial Orientation I in the Radar Observer Selection Battery.

TABLE 11.19.—*Multiple regression statistics for job analogy and factor test batteries, bombardier sample, Validation Study II*

Variable	Job analogy tests, beta weights		Factor tests, beta weights	
	Uncorrected data [1]	Corrected data [2]	Uncorrected data [1]	Corrected data [2]
Coordinate Reading, CP224A or B	0.01	0.02	....	....
Dial and Table Reading CP622A, CP621A	.03	.03	....	....
Numerical Operations, CI702BX1	.10	.10	....	....
Polar Grid Coordinate CP819B	.03	.02	....	....
Scale Reading CP637A	.08	.08	....	....
Spatial Orientation I, CP501B [3]	.06	.05	....	....
Aerial Orientation, CP520A	....	....	0.13	0.12
Numerical Operations, CI702BX1	....	....	.13	.13
Speed of Identification, CP610C	....	....	.08	.08
Visual Memory CI514A	....	....	.10	.10
Multiple R	.21	.21	.24	.25
Shrunken Multiple R [4]	.16	.17	.22	.23

[1] Computed from the uncorrected coefficients given in table 11.18.
[2] Computed from the corrected coefficients given in table 11.18.
[3] Based on scores from administration of Spatial Orientation I in the Radar Observer Selection Battery.
[4] See footnote 4 Table 11.14.

It will be noted that the four factor tests yielded a slightly higher multiple correlation than did the six job analogy tests. More important, however, is the fact that neither multiple correlation approaches the value produced by all 15 tests used for the bombardiers, shown in table 11.14. Also, the beta weights in table 11.14 of two of the job analogy tests, Coordinate Reading, CP224A or B, and Spatial Orientation I, CP501B, are large even though the valid factor tests are included in the same regression equation. This latter fact suggests that the validated factor tests do not adequately sample the factor content of the criterion.

The evidence from this comparison of the factor and job analogy approaches to selection test research suggests that neither should be depended upon exclusively. The conclusion indicated is that, for the present, joint use of both approaches should be made. Until additional factors have been isolated and a greater number of factor measures developed, it will be wise to supplement the latter with job analogy tests.

INDICATED VALIDITY OF FACTORS

The validation statistics and multiple regression data presented in the foregoing sections provide a basis for some tentative conclusions regarding the factors involved in the task of the radar observer. Test validity is a result of factorial content common both to the test and the criterion task. Given the test validities and factor loadings, it is possible to make preliminary estimates as to the relative importance for success in radar observer training of the factors considered in this chapter and chapter 4. Such conclusions will suggest which selection tests measure important or valid factors and therefore merit further research.

This section summarizes the data pertaining to validity for each factor. It should be remembered that the conclusions reached are limited by the tests validated and that, consequently, the list of factors is necessarily incomplete. Other qualifying limitations of the data should also be kept in mind. The most important of these is the undetermined reliability of the criteria used in the validation studies. Since the degree to which the validity coefficients are attenuated by the unreliability of the criteria is unknown, the conclusions to be drawn from them are necessarily doubtful. A second limitation is due to the fact that the validity data are obtained from a sample of students who have already undergone selection resulting in what is, to a large extent, an unknown amount of restriction of range in ability. As will be pointed out, it is probable that the apparent lack of validity of some factors may be a reflection of this restriction. A third limitation is the factorial complexity of certain of the valid tests; for such tests the factor content contributing to the test validity is difficult to infer. A fourth limitation lies in the fact that the factor content of some of the validated tests is neither accurately nor completely described. It is impossible to conclude definitely from the validity or nonvalidity of such tests whether or not the factors they are thought to measure are valid. Finally, the validation sample is small and sampling errors will necessarily influence some of the conclusions reached. Nevertheless, the analysis has been made in this form because of its interest as a general approach to the task of summarizing the implications of a group of validity coefficients.

The factor summaries are presented in alphabetical order. Definitions of the factors discussed will be found in chapter 4. In each summary, a table is presented which lists significant factor content and validity coefficients for all validated tests which have loadings in the factor larger than the arbitrarily selected lower limit of 0.30. The factor loadings given in the tables are, for the most part, approximations based upon factor analyses of the July 1943 and November 1943 air-crew classification batteries. Pertinent information from the multiple regression equations will be presented in the text. The beta weights referred to will be found in table 11.14 and 11.17. For reasons already given, validity coefficients against course grade in Validation Study II will receive somewhat greater emphasis in the analysis than those in Validation Study I.

Evidence for Factor Validity

Length estimation.—Estimation of Length, CP613A, is thought to measure the length estimation factor. Its loadings are unknown because it has never been included in a factor analysis. However, since it showed no validity in the samples to which it was administered, it is likely that the length estimation factor has no validity for radar observer success.

Mechanical experience.—Table 11.20 presents the factor loadings and validities of tests measuring the mechanical experience factor. While the evidence is to some extent contradictory, it, in general, favors the conclusion that this factor has some validity for radar observer training. Mechanical Information, CI905B, when included in the navigator regression equation, makes a unique contribution as shown by its sizable beta weight (table 11.17). This can be explained only by its mechanical experience factor content. This same test, on the other hand, shows no validity in the bombardier sample of Validation Study II. The validity of General Information, as shown under the discussion of pilot interest below, is probably best explained by its mechanical experience content. The validity of Mechanical Principles and Biographical Data Blank, pilot score, in only one of the 5 samples in which they were validated makes doubtful any conclusions based upon them.

TABLE 11.20.—*Summary of evidence pertaining to validity of the Mechanical Experience factor for radar observer training*

Variable	Mechanical experience loading	Other major factor loadings	Significant uncorrected validities			
			Bombardiers		Navigators	
			Study I	Study II	Study I	Study II (course grade criterion)
Mechanical Information, CI905B.	0.75	None	(1)	....	(1)	[2] [3] 0.14
Mechanical Principles, CI903B.	0.60	Visualization 0.45; Reasoning I 0.25; Space I, 0.25.	....	....	[4] 0.18	....
General Information, CE505E.	0.40 to 0.50	Pilot Interest, 0.40; Space I, 0.30; Verbal Comprehension, 0.30; Perceptual Speed, 0.25.	[2] 0.14	[2] 0.10	....	....
Biographical Data Blank, Pilot Score, CE602D.	0.40 to 0.50	Mathematics Background 0.30 to 0.40; Numerical Facility, −0.20.	....	....	[2] .12	....

[1] Not validated.
[2] Significant at the 5-percent level.
[3] Wrong sign.
[4] Significant at the 1-percent level.

Memory I.—Table 11.21 presents the factor loadings and validities of the test thought to measure the memory I or rote memory factor. No significant validity was obtained for the test, Memory for Landmarks, CI510AX2. Since, in Validation Study I, none of the navigators had scores, the test was validated only for the bombardiers. This meager evidence indicates a lack of validity of memory I for radar observer training.

TABLE 11.21.—*Summary of evidence pertaining to validity of the Memory I factor for radar observer training*

Variable	Memory I loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG [1]	CE [2]
Memory for Landmarks, CI510AX2.	[3] 0.60	[3] None	Not validated	0.10	Not validated	No cases	No cases

[1] Course grade criterion.
[2] Circular error criterion.
[3] Estimated from the loading of an earlier form, CI510AX1.

Memory II.—Table 11.22 presents the factor loadings and validities for the test thought to measure memory II or visual memory. This test, Visual Memory, CI514A, was validated in Validation Study II for the bombardier sample only, because too few navigators had scores to warrant computing a validity coefficient. For the bombar-

TABLE 11.22.—*Summary of evidence pertaining to validity of the Memory II factor for radar observer training*

Variable	Memory II loadings	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG [1]	CE [2]
Visual Memory, CI514A...	[3] 0.60	[3] None	Not validated.	[4] 0.13	Not validated.	Too few cases.	Too few cases.

[1] Course-grade criterion.
[2] Circular-error criterion.
[3] Estimated from a similar test, Map Memory, CI505AX2.
[4] Significant at the 1-percent level.

dier sample, the rights score of Visual Memory had a validity coefficient of 0.13, significant at the 1 percent level, while the wrongs score had a coefficient of −0.10, significant at the 5 percent level. This evidence, although based only on one sample, suggests strongly that memory II has slight but stable validity for radar observer training.

Numerical Facility.—Table 11.23 presents the factor loadings and validities of the tests measuring the numerical-facility factor. The validity of Numerical Operations, CI702BX1, a relatively pure measure, is good evidence for the validity of the numerical-facility factor. That the test did not have significant validity for the navigator sample, Validition Study II, is probably explained by the great selection in numerical ability known to have operated on the navigators. In all regression equations in which it is included (tables

TABLE 11.23.—*Summary of evidence pertaining to validity of the Numerical Facility factor for radar observer training*

Variable	Numerical facility loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG [1]	CE [2]
Numerical Operations, CI702BX1.	[3] 0.80	None........................	Not validated.	[4] 0.10	Not validated.	0.15	−0.04
Mathematics A, CI702F.	.60	Verbal Comprehension 0.40; Mathematics Background, 0.25 to 0.35.	[4] 0.21	[5].10	0.09	−.01	[4].16
Mathematics B, CI206C.	.50	Reasoning I, 0.50; Verbal Comprehension, 0.30.	.10	[4].14	[4].12	[4].16	[4].14
Dial and Table Reading, CP622A and CP621A.	.50	Space I, 0.40; Perceptual Speed, 0.20 to 0.30.	[4].14	[5].10	[4].20	[4].16	[4].18

[1] Course grade criterion.
[2] Circular error criterion.
[3] Estimated from the loading of an earlier form, CI702A and B.
[4] Significant at the 1-percent level.
[5] Significant at the 5-percent level.

11.14 and 11.17), Numerical Operations maintains a sizable beta weight which can only be explained by the validity of numerical facility.

It is not clear what is indicated by the validity of the factorially complex tests which have loadings in numerical facility. Mathematics A, CI702F, besides its numerical facility loading, has loadings in verbal comprehension and in a factor tentatively named "mathematics background." The validity of verbal comprehension is doubtful, as will be shown below. This, plus the fact that no measure of "mathematics background" is in the bombardier regression equation (table 11.14), makes it appear that the negative beta weight of Mathematics A in that equation is best explained in numerical facility content held in common with Numerical Operations. This is based on the assumption that a valid test will not contribute to a multiple correlation if purer measures of its valid factors are also in the regression equation. The validity of Mathematics B, CI206C, on the other hand, seems to come, in part, from the Reasoning I content since it has a sizable beta weight (tables 11.14 and 11.17) even when included with Numerical Operations. Other evidence for the validity of Reasoning I is given on page 264. The valid factor content of Dial and Table Reading, CP622A and CP621A, may be inferred from its beta weights in the bombardier and navigator regression equations. In the former, it has a zero beta weight when included with relatively pure measures of numerical facility (Numerical Operations), Space I (Aerial Orientation), and perceptual speed (Speed of Identification). In the navigator equations, however, perceptual speed is represented by Spatial Orientation I and numerical facility is represented by Numerical Operations, but no measure of Space I is present. In this case, Dial and Table Reading has a sizable beta weight which seems to indicate that its Space I content yields the bulk of its validity. The data seem to warrant the conclusion that the numerical facility factor has validity for radar observer training.

Perceptual speed.—Table 11.24 presents the factor loadings and validities of tests measuring the perceptual speed factor. The validity of Speed of Identification, CP610C, for the bombardier sample, Validation Study II, supports the validity of the perceptual speed factor since this factor accounts for a substantial portion of the test's variance. In the bombardier regression equation (table 11.14), Speed of Identification has a large beta weight. On the other hand, no suitable explanation can be given for its lack of validity in the navigator sample of Validation Study II. Spatial Orientation I (scores from administration in the radar observer selection battery) was included in both regression equations even though it did not have uncorrected validity coefficients significant at the 5-percent level. In both cases, it obtained sizable beta weights. This was somewhat surprising in

TABLE 11.24.—*Summary of evidence pertaining to validity of the Perceptual Speed factor for radar observer training*

Variable	Perceptual speed loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG [1]	CE [2]
Speed of Identification, CP610C.	[3] 0.65	None	([5])	[4] 0.12	([5])	−0.10	[4] −0.16
Spatial Orientation I, CP501B.	.60	None	0.08	.06	0.07	.06	.09
Spatial Orientation II, CP503B.	.50	Reasoning I, 0.20	.09	.03	[4] .10	−.08	−.01

[1] Course grade criterion.
[2] Circular error criterion.
[3] Estimated from an earlier form, CP610B.
[4] Significant at the 5-percent level.
[5] Not validated.

the case of the bombardier equation, since significant factor content other than perceptual speed is not known for either Speed of Identification or Spatial Orientation I. While the available evidence is not uncontradictory, it seems in general to favor the tentative conclusion that the perceptual speed factor is valid.

Pilot Interest.—Table 11.25 presents the factor loadings and validities of the single test measuring the pilot-interest factor. The factorial complexity of this test, General Information, CE505E, precludes drawing any clear-cut conclusion as to its factor validity. It yielded a relatively large beta weight in the bombardier regression equations (table 11.14) even though better measures of space I and perceptual-speed each had sizeable weights in the same equation. This may be taken to mean that the validity of General Information is due to factor content other than space I or perceptual speed. If the validity of General Information were completely due to its space I or

TABLE 11.25.—*Summary of evidence pertaining to validity of the Pilot Interest factor for radar observer training*

Variable	Pilot interest loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG [1]	CE [2]
General Information, CE505E.	0.40	Mechanical Experience, 0.40 to 0.50; Space I, 0.30; Verbal Comprehension, 0.30; Perceptual Speed, 0.25.	[3] 0.14	[3] 0.10	0.04	0.03	0.10

[1] Course grade criterion.
[2] Circular error criterion.
[3] Significant at the 5-percent level.

perceptual-speed content, it could not be expected to add to the predictive value of an equation which includes better measures of these factors. Its total valid variance in this case would be explained by content held in common with the other measures and it would add nothing to their combined predictive power. This view, plus the lack of supporting evidence for the validity of verbal comprehension content, leads to the conclusion that the validity of General Information is due either to its mechanical experience or its pilot-interest loadings. On the basis of job-analysis information, mechanical experience seems to be a more likely explanation for validity than does pilot interest. However, until these deductions are supported by additional empirical evidence the validity of the pilot-interest factor must remain in question.

Psychomotor Coordination.—Table 11.26 presents the factor loadings and validities of tests measuring the psychomotor coordination factor. Neither Rotary Pursuit, CP410B, nor Complex Coordination, CM701A, produced a significant validity coefficient. This apparent lack of validity indicates that individual differences in psychomotor coordination are not important to the radar observer's task in training.

TABLE 11.26.—*Summary of evidence pertaining to validity of the Psychomotor Coordination factor for radar observer training*

Variable	Psychomotor coordination loading	Other major factor loadings	Significant uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CO [1]	CE [2]
Rotary Pursuit, CP410B	[3] 0.55	None	0.03	−0.06	0.03	−0.01	0.05
Complex Coordination, CM701A.	.40	Space I, 0.45	.04	.07	−.01	.11	.03

[1] Course grade criterion.
[2] Circular error criterion.
[3] Estimated from an earlier form, CM803A, which did not require divided attention.

Psychomotor Precision.—Table 11.27 presents the factor loadings and validities of tests measuring the psychomotor precision factor. Finger Dexterity, CM116A, yielded validity significant at the 5 percent level, for the navigator sample, validation study I, but for none of the remaining four samples in which it was validated. Discrimination Reaction Time, CP611C, is also valid in only one of the samples. Its validity in this instance is difficult to explain because of its factorial complexity. In the bombardier regression equation (table 11.14) the test has a sizable beta weight even though measures of space I (Aerial Orientation), numerical facility (Numerical Operations), and perceptual speed (Speed of Identification) are also in the equation. This seems to indicate the validity of its psychomotor precision loading.

TABLE 11.27.—*Summary of evidence pertaining to validity of the Psychomotor Precision factor for radar observer training*

Variable	Psychomotor precision loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CO [1]	CE [2]
Finger Dexterity, CM116A.	Unknown.....	Psychomotor Coordination, 0.50.	0.05	−0.06	[3] 0.13	0.12	0.03
Discrimination Reaction Time, CP611D.	Unknown.....	Space I, 0.45; Numerical Facility, 0.25; Perceptual Speed, 0.25.	−.06	[3] .11	.07	−.04	.01

[1] Course grade criterion.
[2] Circular error criterion.
[3] Significant at the 5-percent level.

However, in view of the small proportion of the possible samples in which either of the tests considered was valid, no definite conclusion about the validity of this factor could be reached.

Reasoning I.—Table 11.28 presents the factor loadings and validities of tests measuring the reasoning I factor. The validity of Spatial Reasoning, CI211BX2, in the navigator sample, validation study II, indicates some validity for reasoning I. This test, on the other hand, showed no validity in the bombardier sample of validation study II. When Mathematics B, the other test with reasoning I loading is included in regression equations with Numerical Operations, both receive sizable beta weights (tables 11.14 and 11.17). This seems to indicate that the validity of Mathematics B is accounted for either by its reasoning I or its verbal comprehension loadings in addition to its numerical facility loading. As pointed out later, there is almost no evidence for the validity of the verbal comprehension factor; hence the validity of Mathematics B, over and above that due to

TABLE 11.28.—*Summary of evidence pertaining to validity of the Reasoning I factor for radar observer training*

Variable	Reasoning I loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CO [1]	CE [2]
Spatial Reasoning, CI211BX2.	0.55	Unknown..................	([3])	0.08	([3])	[5] 0.26	0.06
Mathematics B, CI206C.	.50	Numerical, 0.50; Verbal Comprehension, 0.30.	0.10	[4] .14	[4] .12	[4] .16	[4] .14

[1] Course grade criterion.
[2] Circular error criterion.
[3] Not validated.
[4] Significant at the 5-percent level.
[5] Significant at the 1-percent level.

numerical facility seems most likely to be due to its reasoning I loading. It is possible that the validity of another test, Reading Comprehension, CI614H, is also due primarily to its reasoning I loading. This test, which is omitted from table 11.28 because its reasoning I loading is only 0.25, has a verbal comprehension loading of about 0.65 and a mechanical experience loading of 0.25. When included in a regression equation that also includes Mathematics B (table 11.14), it has a beta weight of zero. Since these two tests have in common only the verbal comprehension and reasoning I factors and verbal comprehension is assumed to be invalid, the conclusion indicated is that the validity of Reading Comprehension is due to reasoning I. In summary, the preponderance of evidence indicates that reasoning I is valid for radar-observer training.

Space I.—Table 11.29 presents the factor loadings and validities of tests measuring the space I or spatial relations factor. The clearest evidence for the validity of the space I factor is the validity of Aerial Orientation, CP520A, which is thought to have a major loading only in space I. Supporting evidence is yielded by the validity of Dial and Table Reading, CP622A and CP621A, which has been explained on page 220 primarily by its space I loading. Greatest doubt comes from the lack of validity of Instrument Comprehension II, CI616B, Complex Coordination, CP701A, and Flight Orientation, CP528A. The lack of validity for Instrument Comprehension II is particularly

TABLE 11.29.—*Summary of evidence pertaining to validity of the Space I (spatial relations) factor for radar observer training*

Variable	Space I loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG [1]	CE [2]
Instrument Comprehension II, CI616B.	0.50	Perceptual Speed, 0.30; Reasoning I, 0.30.	0.11	0.02	0.08	0.04	0.12
Complex Coordination, CM701A.	.50	Psychomotor Coordination, 0.40.	.04	.07	−.01	.11	.03
Discrimination Reaction Time, CP611D.	.40	Numerical Facility, 0.25; Perceptual Speed, 0.20; Psychomotor Precision (unknown).	−.06	[3].11	.07	−.04	.01
Aerial Orientation, CP520A.	([4])	Unknown	([5])	[3].12	([5])	.09	.12
Flight Orientation, CP528A.	([4])	Unknown	([5])	.07	([5])	.11	.02
Two-Hand Coordination, CM101A.	.40	Mechanical Experience, 0.25 to 0.40; Psychomotor Coordination, 0.30.	.12	.00	.08	.06	.11
Dial and Table Reading, CP622A and CP621A.	.40	Numerical Facility, 0.40; Perceptual Speed, 0.20 to 0.30.	[6].14	[6].10	[3].20	[6].16	[3].13

[1] Course-grade criterion.
[2] Circular-error criterion.
[3] Significant at the 1-percent level.
[4] Unknown.
[5] Not validated.
[6] Significant at the 5-percent level.

troublesome since it has major loadings in two other factors for both of which there is good evidence of validity. The data from Discrimination Reaction Time, CP611D, and Two-Hand Coordination, CM101A, are ambiguous because these tests are complex factorially. The validity of Discrimination Reaction Time has already been interpreted as probably due primarily to psychomotor precision. The ambiguity of this evidence prohibits any conclusions at this time regarding the validity of space I. It is concluded that space I may have low, positive validity for predicting success in radar observer training.

Space II.—Table 11.30 presents the factor loadings and validities of the test thought to measure the Space II or rotational space factor. This test, Position Orientation, CP526A, yielded no significant validity for either sample of Validation Study II. A lack of validity of Space II for radar observer training is indicated.

TABLE 11.30.—*Summary of evidence pertaining to validity of the Space II factor for radar observer training*

Variable	Space II loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG[1]	CE[2]
Position Orientation, CP526A....	[3] 0.45	None..............	[4]	0.08	[4]	0.02	0.05

[1] Course grade criterion.
[2] Circular error criterion.
[3] Estimated from an earlier form, the Hands test, CP512A.
[4] Not validated.

Verbal comprehension.—Table 11.31 presents the factor loadings and validities of tests measuring the verbal comprehension factor.

TABLE 11.31.—*Summary of evidence pertaining to validity of the Verbal Comprehension factor for radar observer training*

Variable	Verbal comprehension loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG[1]	CE[2]
Air Corps Vocabulary (1942).	0.75	None......................	[3]	0.04	[3]	[4]	[4]
Reading Comprehension, CI614H.	.60	Mechanical Experience, 0.25 to 0.30; Reasoning, 0.25.	0.12	[5].10	[6]0.15	0.08	0.11

[1] Course grade criterion.
[2] Circular error criterion.
[3] Not validated.
[4] No cases.
[5] Significant at the 5-percent level.
[6] Significant at the 1-percent level.

The fairly pure measure, Air Corps Vocabulary (1942), failed to produce significant validity. However, this test was only validated for the bombardier sample, Validation Study II, since none of the navigators used in the study had Air Corps Vocabulary scores. The validity of Reading Comprehension, the only other adequate measure of verbal comprehension, is probably explained by factor content other than verbal comprehension, specifically, by Reasoning I. This is deduced from the fact that its beta weight is near zero in a bombardier regression equation (table 11.14) in which it is the only test with a significant verbal comprehension loading. All evidence taken together seems to indicate that verbal comprehension has no validity for radar observer training.

Visualization.—Table 11.32 presents the factor loadings and validities of the tests measuring the visualization factor. The validity of 0.18, significant at the 1-percent level, produced by Mechanical Principles, CI903B, in the navigator sample, Validation Study I, is the only validity yielded by a test having a major loading in the visualization factor. Since this validity may be explained by the high mechanical experience loading of Mechanical Principles, it appears likely that visualization has no validity for radar observer training.

TABLE 11.32.—*Summary of evidence pertaining to validity of the Visualization factor for radar observer training*

Variable	Visualization loading	Other major factor loadings	Uncorrected validities				
			Bombardiers		Navigators		
			Study I	Study II	Study I	Study II	
						CG [1]	CE [2]
Pattern Comprehension, CP803A.	0.50	Reasoning I, 0.25	[3]	0.02	[3]	0.08	−0.09
Area Visualization, CP815A	.50	None	[3]	.02	[3]	[4]	[4]
Mechanical Principles, CI903B.	.50	Mechanical Experience, 0.60.	0.07	.01	[5] 0.18	−.02	.11

[1] Course grade criterion.
[2] Circular error criterion.
[3] Not validated.
[4] Too few cases.
[5] Significant at the 1-percent level.

Summary of Validity of Factors.—On the basis of the preceding summaries of the data for each factor, the factors may be grouped under several headings on the basis of estimated validity for radar observer training. Five of the factors considered have been judged to be valid; these are mechanical experience, memory II (visual memory), numerical facility, perceptual speed, and reasoning I (general reasoning). The evidence in the case of three of these factors, memory II, numerical facility, and reasoning I, is relatively clear-cut but varies in amount. The evidence for the validity of the remaining two factors, mechanical experience and perceptual speed, while not altogether uncontradictory, is still preponderently favorable. The weight of available evidence points to a lack of validity for the following six factors: Length estimation, memory (rote memory), psychomotor coordination, space II (rotational space), verbal comprehension, and visualization. For three factors, pilot interest, psychomotor precision, and space I (spatial relations), the evidence from different sources is too ambiguous to make any conclusion possible.

SUMMARY

The results of two validation studies are presented in this chapter. In the first study, the air-crew specialty stanine scores, the tests in the air-crew classification battery, and the tests in the radar observer selection battery were validated against course grades assigned by the training authorities at Boca Raton, Langley Field, and Victorville. The course grades were based partly upon informal tests and instructor ratings and partly upon standardized tests and performance checks. The validation samples consist of 205 bombardiers and 381 navigators.

For the bombardier sample in the first study, six variables had positive validity coefficients, significant at the 5-percent level or 1-percent level; the navigator and pilot stanines; navigator score of Biographical Data Blank, CE602D; Dial and Table Reading, CP622A and CP621A; General Information, CE505E; and Mathematics A, CI702F. For the navigator sample, nine variables had positive validity coefficients, significant at the 5-percent or 1-percent level; the bombardier and navigator stanines; pilot score of Biographical Data Blank, CE602D; Dial and Table Reading, CP622A and CP621A; Mathematics B, CI206C; Mechanical Principles, CI903B; Reading Comprehension, CI604H; Spatial Orientation II, CP503B; and Finger Dexterity, CM116A.

The second study included the variables in the two batteries used in Validation Study I and also an experimental selection battery. The primary criteria consisted of radar observer course grades obtained for classes later than those used in Validation Study I, but before VJ-day. Course grades for the 629 bombardier students from Boca Raton and Langley Field were computed by the Radar Project wholly from standardized test and performance check scores. The course grades for the 226 Victorville students, all of whom were navigators, were similar to those used in the first study, being based primarily upon subjective ratings and informal quizzes. The selection variables were also validated against average circular error scores for the Victorville sample.

For the bombardier sample, 16 variables had positive correlations with course grades, significant at the 5-percent or 1-percent level: the bombardier and navigator stanines, Aerial Orientation, CP520A; Coordinate Reading, CP224A or B; Dial and Table Reading, CP622A and CP621A; Discrimination Reaction Time, CP611D; General Information, CE505E; Mathematics A, CI702F; Mathematics B, CI206C; Numerical Operations, CI702BX1; Polar Grid Coordinate, CP819B; Ratio Estimation, CP225A; Reading Comprehension, CI614H; Scale Reading, CP637A; Speed of Identification, CP610C; and Visual Memory, CP514A. These variables along with Spatial Orientation I, CP501B, produced an uncorrected multiple correlation of 0.51 with the course grade criterion.

For the navigator sample, six variables had positive correlations with course grades, significant at the 5-percent or 1-percent level; the bombardier stanine; Dial and Table Reading, CP622A and CP621A; Mathematics B, CI206C; Mechanical Information, CI905B; Pattern Orientation, CP816A; and Ratio Estimation, CP225A. These variables together with the navigator stanine, Coordinate Reading, CP224A or B, Numerical Operations, CI702BX1, Scale Reading, CP637A, and Spatial Orientation I, CP501B, yielded an uncorrected multiple correlation of 0.36.

For the navigator sample, five variables had negative correlations with average circular error scores, significant at the 5-percent or 1-percent level: the bombardier and navigator stanines; Dial and Table Reading, CP622A and CP621A; Mathematics A, CI702F; and Mathematics B, CI206C.

The samples of bombardiers and navigators for which the validity coefficients were computed had been subjected to selection at four points in their training: before attaining aviation student status, at classification center, at bombing or navigation school, and prior to entering radar observer training. Although it would have been desirable to correct the coefficients for more of the restrictions, it was possible to correct them only for the last which was based on the radar stanine. The corrected coefficients are estimates of the validities the tests would have for the population of graduating bombardiers and navigators. In most cases the corrections made little difference. Both corrected and uncorrected coefficients were used to compute the multiple regression statistics. The corrected coefficients yielded slightly higher multiple correlations in most cases.

Validity and multiple regression data were summarized in terms of the indicated importance of different factors for radar observer training. Five factors were judged to be valid for training success: mechanical experience, memory II (visual memory), numerical facility, perceptual speed, and reasoning I (general reasoning).

CHAPTER TWELVE

An Evaluation, With Suggestions for Future Research[1]

In this chapter, psychological research carried out in radar-observer training will be reviewed both from the point of view of its interest to psychologists and its potential value to the radar-observer program. Before undertaking such a review, it will be helpful to recall that the orientation of the Radar Project, like that of similar projects, was primarily to "war psychology" and that its first concern, consequently, was to produce for wartime use. Within this limitation, however, the greatest possible effort was made to insure the collection of data which would make possible evaluation of the reliability and validity of its instruments.

The presentation within the chapter is organized in the following terms: (1) The development and validation of selection tests, (2) the development of proficiency tests and performance checks, (3) the development of methods for instructor evaluation and selection, and (4) research on trainers and training methods.

JOB ANALYSIS AND SELECTION-TEST RESEARCH

Chapter 4 of the report presents a discussion of a job description and job analysis of the task of the radar observer in training. In chapter 11 the results of the validation of selection tests are analyzed; in part, this analysis is summarized in terms of the predicted validities of psychological abilities resulting from the job analysis. To personnel psychologists concerned with the process of studying a job from the point of view of anticipating which tests will predict success, the approach in chapter 4 will be of considerable interest. In describing the radar observer's task, a distinction is made between "job description," in which the description is made in terms specific to a single job, and "job analysis," where description is made in terms of general abilities. It is believed that a reading of chapter 4 will present convincing evidence of the value of distinguishing between the two. As is pointed

[1] Written by Capt. Stuart W. Cook.

out there, a job description is more useful in the preparation of proficiency measures and in planning training research and research into techniques and equipment. Both the job description and job analysis are useful as a basis for selection-test research, although it is predicted that such research will more and more come to be based upon the latter. Opinion on this point will be a function of one's judgment as to the eventual value of factor theory for prediction of job success. Factor theorists hold that performance on most of the tasks in contemporary technology can be explained by a limited number of independent functions or factors taken in various combinations and amounts. If this proves to be true, occupational analysis and selection-test research in the future should consist of description of jobs in terms of statistically isolated abilities, followed by validation of tests known to measure these abilities.

The difficulty commonly experienced by factor analysts in preparing an adequate verbal description of a newly isolated factor gives rise to a parallel difficulty for the job analyst. Verbalizing one's introspections following experience with the job or observation of persons undertaking the job, in itself presents difficulties, but these are greatly increased by the problem of converting these introspections into vaguely described psychological functions.

In chapter 11 the results of two validation studies are presented. In the first study, the air-crew specialty stanine scores, the tests in the air-crew classification battery, and the tests in a selection battery used to screen radar observer students, were validated against course grades assigned by the training authorities at three radar observer training stations. These grades were based partly upon informal tests and instructor ratings and partly upon standardized tests and performance checks. The validation samples consisted of 205 bombardiers and 381 navigators. The second study included the variables in the two test batteries used in Validation Study I, plus tests from an experimental selection battery. The validation criterion in this study consisted of radar observer course grades for classes susbsequent to those used in Validation Study I, plus tests from an experimental selection battery. The validation criterion in this study consisted of radar observer course grades for classes subsequent to those used in Validation Study I, but before VJ-day. Course grades for one part of the sample, 629 bombardier students, were computed by the Radar Project wholly from standardized tests and performance check scores. Course grades for the navigator part of the sample, which consisted of 226 students, were similar to those used in the first study, being based primarily upon subjective ratings and informal quizzes. For the navigator students, the selection variables were also validated against average circular error scores.

The most interesting multiple correlation coefficients are believed to be those resulting from the second study, since, in this study, a large number of experimental selection tests were validated against a proficiency criterion composed primarily of standardized proficiency tests and performance checks. For the bombardier sample in this second study, an uncorrected multiple validity coefficient of 0.51 was obtained. The parallel figure for the navigator sample was 0.36. These figures must be qualified in several ways. In the first place, the samples of students for which the validation coefficients were computed had been subjected to selection at four points in their training prior to administration of the selection tests in this study. They had been tested as a prerequisite to being granted aviation student status; they were screened further at air-crew classification centers; additional persons were eliminated through training failure at bombing or navigation schools; and still others were eliminated by tests prior to their entrance to radar observer training. Selection of this sort undoubtedly gave rise to curtailment of abilities, for which it was not possible to correct. Such corrected coefficients as are presented consist only of estimates of the validities the tests would have had for the total population of graduating bombardiers and navigators. It is not possible, consequently, to compare test validities presented in this report with those obtained in air-crew classification centers. For the same reason it would be possible from these results to make only approximate estimates of the validity of tests for selecting radar observers from the population of students admitted to aviation student status.

The second qualification has to do with the shrinkage to be expected in the multiple correlations reported. While an attempt has been made to correct for this shrinkage, it is known that the correction affords an underestimate of the amount of shrinkage which will occur. With this qualification, the shrunken coefficients which correspond to the figures of 0.51 and 0.36 given above are 0.47 and 0.24 respectively.

A third qualification has to do with the reliability of the validation criterion. According to the best evidence available, this reliability is probably not higher than 0.40 or 0.50.

Plans had been made for further experimental validation of selection tests. However, these were interrupted by the cessation of training which accompanied the end of the war. It had been decided that the most fruitful areas to test beyond those already sampled would be those of interests, attitudes, and emotions. Biographical data tests which had proven to be valuable in other air-crew specialties were to have been tried out. Future selection test research with radar observers should probably emphasize tests in these fields.

As indicated above, the validation statistics were analyzed, in part, from the point of view of hypotheses advanced in connection with the job analysis. The tests validated were divided into three groups: one, composed of tests measuring known factors; a second, of tests especially constructed to simulate a part of the radar observer's job; and a third, of tests which could be clearly fitted into neither of these categories. Of the group of factor tests, four had validity coefficients significantly different from zero. For the second group of tests, called job analogy tests, there were six significant coefficients. The four factor tests yielded a multiple correlation of 0.21, while the six job analogy tests gave a multiple of 0.21. Neither of these values approach the multiple of 0.45 obtained with the total group of tests. Thus, while the approach through available factor tests was equally successful with that of construction of new job analogy tests, it is clearly advisable to use both in combination. Two inferences may be drawn. One is that an inadequate coverage of factor tests was available; the second is that complex selection tests of the job analogy type measure abilities not covered by available factor tests. In the latter case, a factor study of the new job analogy tests should be illuminating.

A note of interest in connection with the use of factor tests in selection test research has to do with the customary attitude toward the expected size of individual test validities. Factor tests consistently yielded low validities, often from 0.10 to 0.30. These values alone would be discouraging, were it not for the fact that batteries of relatively uncorrelated tests with such validities yield useful multiple correlations. For example, for the bombardier sample in the second validation study, reported in chapter 11, 15 tests with validities ranging from 0.07 to 0.16 yielded a multiple correlation of 0.45. Similarly, for the navigator sample, 9 tests with validities ranging from 0.08 to 0.18 produced a multiple of 0.31.

Also of interest to occupational analysts is the degree of success which accompanied attempts to predict the validity of various psychological abilities on the basis of the job analysis. In chapter 11, all possible evidence was accumulated which related to the validity of each factor. This evidence unfortunately is very slight in many cases and conflicts in others. In addition, it is based upon rather small samples. With these qualifications, a rough summary of success in predicting factor validity may be made as follows: Of the five factors for which there seemed the most positive evidence of appreciable validity, one had been predicted to be among the most valid, three had been predicted to be among those with relatively high validity, one had been predicted to be among those with relatively low validity, none had been predicted to be among those with lowest validity, and none had been predicted to have no validity. Of the six factors for which there seemed to be most definite evidence for lack of validity, three had been predicted to have no validity, one had been predicted to be among those with lowest validity, one had been predicted to be among those with relatively low validity, one had been predicted to be among those with relatively high validity and none had been predicted to be among the most valid.

PROFICIENCY MEASUREMENT

The Radar Project was established shortly after the Army Air Forces embarked upon a greatly expanded program for the training of radar observers. This expansion occurred in the fall of 1944 and followed a year of increased success in strategic bombing with radar in the European Theater. Supervisory training personnel under the leadership of Col. William M. Garland were dissatisfied with the methods of proficiency measurement that had been hurriedly improvised and were desirous that new types of proficiency measures be devised. This situation was of interest to the Radar Project not only because of the expressed need for adequate methods of evaluating student proficiency but also because it appeared there was available no adequate criterion against which to validate selection tests. Of the possible criteria, one, the pass-fail criterion, was eliminated because the demand by the operational air forces for radar observers was so great as to prohibit the failure of all save the most inferior students. Another, instructor grades, appeared likely to be of doubtful value because rapid expansion of training necessitated the use of many instructors with no previous teaching experience and others with little or no motivation to teach. A third, bombing accuracy, was made impractical by the lack of sufficient photographic equipment at the radar training stations.

As a result, a comprehensive battery of standardized proficiency measures was constructed and put into wide use. This battery, which is described in chapters 5 and 6, 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 a fifth served as a final comprehensive examination. Four of the performance checks measured proficiency on ground trainers, while two measured aerial performance. Of the latter, one was administered midway through aerial training, while the other served as a final check on aerial performance.

Practical circumstances were such that it was possible to compute the reliability only of the unspeeded sections of printed proficiency tests. The reliability of the speeded tests and of the performance checks is not known. Estimates of the reliability of the course grade based upon the battery of standardized measures indicate that it lies between 0.40 and 0.50. On the basis of other evidence, which also must be qualified considerably, it seems doubtful that this value is higher than the reliability of the grades assigned before the standardized battery was put into use. While recognizing the reservations with which this comparison must be made, it is, on the other hand, clear that the reliability indicated for the standardized measures leaves much to be desired.

Analyses of the interrelations between various proficiency measures are suggestive of probable differences in reliability between the printed tests and the performance checks. It was found in these studies, reported in chapter 8, that tests covering dissimilar material from different phases of the course correlated more highly with each other than did tests and performance checks covering similar subject material. While more than one inference is possible from this finding, it is at least consistent with the hypothesis that printed tests are more reliable than are performance checks.

The experience gained by the Radar Project in attempting to solve difficulties of standardized administration in a large-scale achievement-testing program may be of value in other similar situations. Difficulties fell primarily into two categories. One of these followed from the necessity to depend for test and check administration upon operating personnel with no formal measurement training. The number and extent of the difficulties which can hinge partly upon the motivation and partly upon the comprehension of measurement problems by such personnel were underestimated. After many of these difficulties became apparent, a system was established whereby selected administrators were organized at each training station into specialized examiner boards. This was done, partly, in order to limit the number of personnel that would have to be trained in standardized administration and, partly, to provide for more systematic supervision of examiner personnel. Observation of the operation of this plan suggested that it would produce adequate administration, providing that examiners were selected carefully and indoctrinated thoroughly as to the importance of their task.

A second category of difficulties had to do with the standardization of testing conditions, particularly in connection with the administration of individual-performance checks on ground trainers and in the air. Initial contacts with the situation and with training personnel led the project to believe that the more important variables could be controlled. Further experience showed, however, that control of variables such as weather, route, crew, radar equipment, aircraft, etc., was administratively difficult and, under wartime-training conditions, probably impossible. Adjustment to this realization took the form of "standardized" treatment of instances in which departure from standard conditions occurred. A procedure was developed whereby an examiner was required not only to check each item as the performance check progressed but also to keep a log of the conditions under which he found it impossible to evaluate an item according to the standardized specifications. This log served as a basis for a joint decision on the questionable item by the examiner and the head of the examiner board. The latter individual developed a set of rules governing typical exceptions to routine administration, in terms of which he could make what appeared to be consistent and reliable decisions. Unfortunately, it was never possible to evaluate systematically the efficacy of this procedure from the point of view of improvement in performance-check reliability.

While some hope is held for improvement of reliability through adjustments such as those described above, it now seems clear that another safeguard should also be attempted. It is well known that the reliability of a composite of separate correlated measures, each with low reliability, can be increased by increasing the number of measures in the composite. It is probable that this could have been accomplished in the radar observer training program by adding the battery of standardized proficiency measures to existing grades, rather than substituting it for these grades.

An unexpected and beneficial outcome of the effort to standardize proficiency measurement was improved standardization of the curriculum and of instruction. Supervisory personnel faced with the task of providing for standard testing conditions found themselves applying pressure for increased attention to a thorough and systematic training routine. A standardized measurement program cannot function in a training situation where the instructional material varies irregularly from time to time and from school to school.

Several measuring instruments were developed which will be of special interest to test technicians. One section of the final comprehensive examination was built around the concept of a simulated radar observer bombing mission. This section consisted of interrelated items which required the 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 student's computations. This section of the test was prepared in response to a strong conviction of training authorities that continuous navigation on a mission requires something more than can be measured by tests of the separate skills involved. They believe that organizational abilities and possibly emotional qualities are required for the integrative aspects of conducting the mission.

In addition to this section of the comprehensive examination, there were three other sections, each measuring separate navigational skills such as use of computers, use of charts, and interpretations of scope photographs. These three subtests, when combined, correlated from 0.30 to 0.60 with the subtest containing the simulated mission. However, it was not possible to determine whether they accounted for all of its nonerror variance, since the reliability of the various subtests was not known. It would have been of practical interest to be able to answer this question, since the simulated mission was very time consuming per item in comparison with the simpler subtests. Considerable time could have been saved if the simulated mission were to have been eliminated with assurance that the shorter sections adequately measured the total task. Since all of the subtests were speeded, it would have been necessary to have prepared equivalent forms in order to obtain a satisfactory reliability estimate.

A second point of interest to test technicians is the type of performance check developed for use on the supersonic ground trainer and in the air. Practical circumstances in these two instances forced the measurement of proficiency while the student carried out a typical performance of the total job for which he had been trained. Physical and temporal limitations of the training situation made it impossible to break up this total task into a number of independent items. In addition, the project was faced with a firm conviction on the part of training personnel that any such division of the total task into independent items would remove from it many of the important skills for which measurement was desired. Consequently, during a performance check given on supersonic trainer or in the air, the radar-observer student simply flew a typical mission. Standardization was attempted by having the mission flown over one of two matched standard routes on the assumption that this would present each student with an equivalent set of navigational and bombing problems. Standardized directions to the student were provided. The examiner was furnished standardized instructions for evaluating performance in terms of specific items, each of which was accompanied by defined standards and tolerances. The items in such checks consist of evaluations of performance at convenient and crucial points in the course of a mission. Adjustment to this type of measurement situation led to the development of several techniques which are described in chapter 6. For example, it was found possible to evaluate certain important features of performance only in terms of outcomes observed later in the mission.

Certain generalizations about the measurement of the performance resulted from the project's intensive experience with proficiency measures. Although too lengthy for description here, these are presented in detail in chapter 7.

RADAR BOMBING CIRCULAR ERROR

It was impossible for the Radar Project to make an analysis of radar bombing circular error until near the end of the war. Very little data were available earlier because of lack of photographic and ground radar scoring equipment. The most significant analysis of such data as did become available consisted of determining its reliability. The results, as presented in chapter 9, indicate that circular error scores have low reliability; coefficients range between 0.20 and

0.30. Correlation of average circular error with course grades are between 0.15 and 0.20. These figures suggest that, while average circular error would be of little value if used alone as a criterion of proficiency, it should be considered for use as part of a composite criterion.

Comparison of studies of radar-bombing scores with studies of visual-bombing scores suggest that the former are somewhat more reliable. A possible explanation for this difference is that there is less variability among students in visual-bombing proficiency than in radar bombing. Consequently, the differences between students in visual bombing are more easily obscured by variations in factors unrelated to bombing ability.

A number of suggestions have been made in chapter 9 regarding the improvement of reliability of radar-bombing scores. These are primarily concerned with the standardization of bombing conditions and with the problem of obtaining more complete records of bombing scores for each student. The scoring of bombing runs with the aid of ground radar installations promises much toward the elimination of error in the scoring process.

Of interest to students of measurement is a discussion in chapter 9 of the commonly used methods of determining circular error reliability. The inadequacy of methods which did not take account of day-to-day variations are pointed out. The conclusion reached is that the only adequate method of estimating circular error reliability is that of correlating scores from odd bombing missions with scores from even bombing missions, allowing bombing conditions to vary as they do under routine training conditions.

INSTRUCTOR SELECTION AND EVALUATION

Although extensive plans were made for research in instructor selection and evaluation, most of the plans were cancelled by the termination of hostilities. Following a job description and job analysis of radar-observer instruction, an experimental selection battery was chosen but never validated. Scales for the rating of instructors by students were constructed and their reliability determined. It is a fact of some significance for future research that these scales were found to have very high reliability. A possible implication of this is that the instructors in the radar observer training program differed greatly with respect to the qualities of motivation and instructional ability upon which they were rated. This conclusion agrees with the observation of Radar Project personnel who were enrolled in the course. An inference of this is that, under similar circumstances in the future, scientific instructor selection would yield large returns in improvement of instruction.

TRAINING RESEARCH

While the Radar Project made many informal contributions to wartime training procedure, time did not permit the conduct of any formal training research. However, since project personnel developed considerable familiarity with the training curriculum, the following suggestions are made for possible post war research.

A training problem frequently encountered in radar observer training centers around the importance to learning of student understanding of the operations being performed. A clear example of this arises in relation to instruction in the adjustment of controls on the radar equipment. It is possible, at one extreme, to teach use of the controls almost entirely in terms of the effect they produce on various meters and oscilloscope screens. At the other extreme, considerable explanation can be made of the reasons for the effects produced by various adjustments. In wartime instruction, opinion and practice on this point varied considerably. It is probable that guiding principles could be worked out by experimental comparisons of curricula which differ in the extent to which technical understanding of the equipment is stressed.

Another point at which the value of technical understanding is in doubt is in radar bombing instruction. A series of procedures with appropriate qualifications may be worked out to govern the radar observer's behavior during the bomb run. The effect of limiting instruction to these procedures is a matter upon which there is considerable disagreement. How much an understanding of the theory of bombing and the mechanics of the bombsight speeds learning, contributes to accuracy, improves the handling of emergency situations—all are questions calling for experimental investigation.

A second general problem of radar training has to do with the most efficient media for presentation of instruction. The use of motion pictures and other graphic aids is advocated, but the extent to which such techniques facilitate learning remains undetermined. It is possible that graphic aids such as motion pictures can aid instruction in two respects: (*a*) by increasing classroom interest and, thus, the motivation of the student to learn, and (*b*) by enlarging the scope and variety of material presented. An example of a potentially profitable application of graphic aids may be seen in the classroom teaching of radar navigation. Here the use of actual or artificial scope photographs in simulated missions might afford a closer approximation to aerial missions in the classroom than is accomplished through the problem-solving materials currently used. In addition, the effectiveness of motion pictures as an aid to briefing for aerial missions should be investigated.

A series of possible research problems relate to economy in training time. A conclusion from the extended training experiment conducted by the National Defense Research Committee is that student radar observers require approximately 110 hours of aerial practice to reach peak bombing efficiency. It is probable that in the wartime radar training program more time than this would have been needed, since the NDRC experiment was conducted under conditions specially favoring learning and atypical of the radar training program in general. This exorbitant price in aerial training time raises research questions as to whether the time devoted to other features of the course could be profitably revised in order to reduce the number of hours required for a student to reach maximum efficiency in the air.

One specific source of economy would be to attempt to govern the length of different parts of the ground school curriculum in accordance with their contributions to the achievement of peak aerial proficiency. This problem is essentially one of discovering optimum training periods in the various areas of ground school instruction. A potential study of supersonic trainer instruction may serve as an illustration. It is possible that a substantial increase in supersonic training beyond that provided for in the wartime curriculum would shorten the amount of aerial practice needed. On the other hand, it is equally possible that, beyond a certain point in supersonic trainer instruction, additional practice would interfere with the development of aerial proficiency.

Perhaps the most fundamental problem encountered in training is the proper organization of instruction in the component skills of the radar observer's task. Instruction may proceed along either of two lines—or it may follow some middle course. It is possible, on the one hand, to assume that the radar observer's task consists of basically discrete skills, and that the integration of these skills should be attempted only after each has been mastered separately. Thus, separate courses would be given in the major phases of the radar observer's job: radar navigation, scope interpretation, set operation, radar bombing, and so forth. Another view, on the other hand, looks upon the radar observer's task as primarily a complex of skills, all of which are so interrelated functionally as to make separate instruction in each unrealistic and uneconomical. In this view, efficient learning is aided most by organizing all learning around the integration of skills. Thus, aerial and trainer practice would constitute the major phases of the curriculum with instruction in component skills such as navigation, bombing, set operation, and scope interpretation taking place within the context of the total task. A research answer should be sought as to the correct compromise between these two views.

APPENDIX A

Selection Tests Validated Against Success in Radar Observer Training[1]

Three selection batteries representing 18 printed and psychomotor tests were validated against success in radar observer training. These are: the Aircrew Classification Battery (November 1943), the Radar Observer Selection Battery, and the Experimental Battery. Descriptions of tests in these batteries, together with diagrams showing the composition of the three aircrew specialty stanines, are presented in this appendix.

AIRCREW CLASSIFICATION BATTERY (NOVEMBER 1943)

Printed Tests:

Test: Biographical Data Blank, CE602D.

Description: This test contains items of personal history which are related to success in pilot and navigator training. It consists of items concerning the individual's educational history, work experience, sports and hobby participation, place of birth, parentage, etc. The items in this test were selected from a much larger number which had been tried out experimentally in previous forms of the test (CE602A and CE602B).

Number of items, 65
Time limit, 25 minutes
Scoring formula: R − W ÷ 20

[1] Compiled by Sgt. Hyman Heller with the assistance of Sgt. Alfred S. Arnett and Sgt. Albert H. Hastorf.

At several points in the report references have been made to copies of tests and performance checks which were to have been included in an appendix B. [illegible] report it was decided that these materials were not of sufficient general interest to compensate for the considerable space their presentation required. Persons interested in the copies of the tests will find them on file in the Office of the Air Surgeon, Headquarters, Army Air Forces, Washington, D. C., and Psychological Section, Office of the Surgeon, Headquarters, AAF Training Command, Barksdale Field, La. Stencils from which reproduction in quantity may be made are on file in the Psychological Section, Office of the Surgeon, Headquarters, AAF Training Command, Barksdale Field, La.

Reliability: Pilot Key, $r = 0.86$. Navigator Key, $r = 0.49$. Correlations obtained on a test-retest basis with a time interval of approximately 28 days. N = 711.[2]

Sample item: During most of your life you have lived:

A. In a large city (over 100,000).
B. In a city (10,000 to 100,000).
C. In a small town (1,000 to 10,000).
D. In a very small town (under 1000).
E. In the country.

Test: Dial Reading, CP622A.

Description: This test requires the subject to read various dials quickly and accurately. It consists of 10 pages, each page containing 7 dials. The scales on the different dials vary in size, angular range, and units of measure. The needle of each dial points to some value. Below the bank of dials are a number of five-choice items consisting of the label of an instrument (amperes, altitude, etc.) and five alternatives for the reading. The task is to find the dial, read it, and then find the correct answer among the five alternatives. This test is printed as the first part of a booklet which also contains Table Reading, CP621A.

Number of items, 57.

Time limit, 9 minutes.

Scoring formula: Combined with Table Reading, CP621A; $R - W/2$.

Reliability: $r = 0.76$. Correlation between two separately timed halves corrected for length. N = 1,167.[3]

Sample items: See Figure A.1.

Test: Table Reading, CP621A.

Description: This test is printed in a booklet with Dial Reading CP622A, and consists of two parts which are timed separately.

Part I consists of a large bivariate table. The subject is given a pair of marginal values, and he must find the entry in the body of the table which is directly below one value and horizontally in line with the other.

Part II consists of a set of four tables, each of which has a pair of entries for each of the three values of one variable, and 10 entries for a second variable. The subject must determine which of the four tables to enter, find the correct column and row in that table, and finally select the correct pair of numbers.

Number of items: Part I, 43; part II, 13.

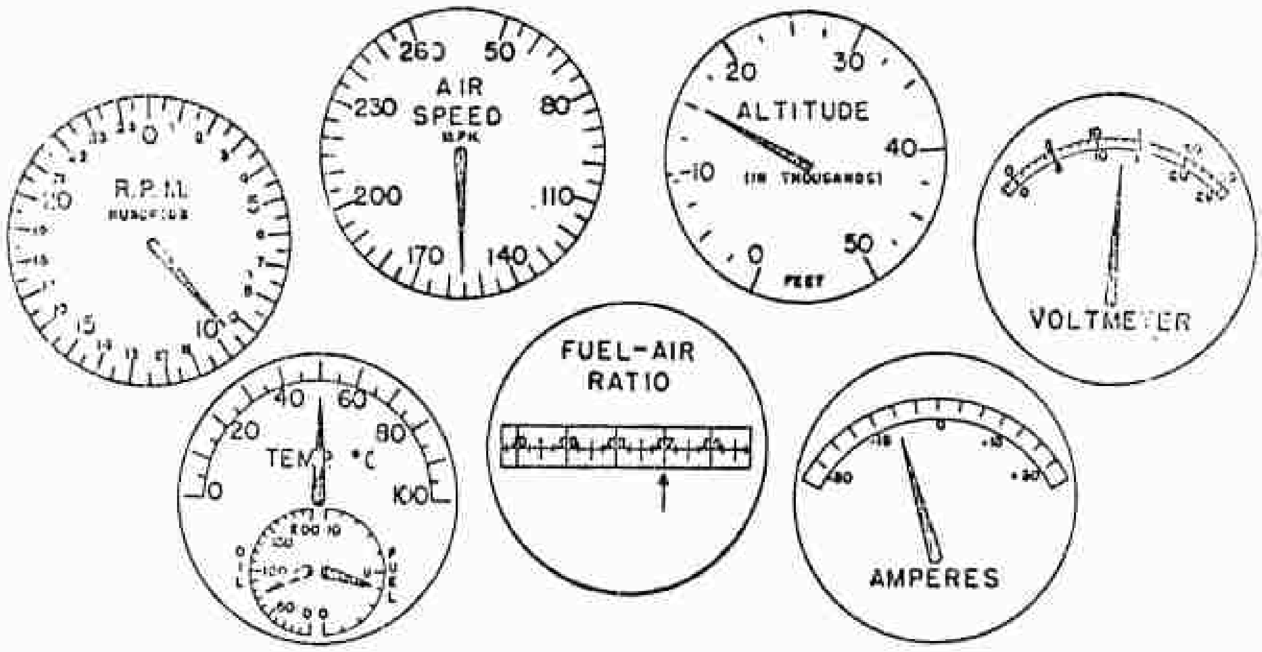
Time limit: Part I, 8 minutes; part II, 7 minutes.

Scoring formula: Combined with Dial Reading, CP622A: $R - W/2$.

---

[2] Guilford, J. P. and Lacey, J. I., eds. *Printed classification tests.* AAF aviation psychology program research reports. No. 5. Washington: Government Printing Office. 1947. Chapter 27.

[3] *Ibid.*, chapter 16.



Sample Problems:	A	B	C	D	E
I. R. P. M.	91.	89. 5	9[illegible]5	92.	105.
II. Amperes	16.	−15. 5	14.	−10.	−2.
III. Altitude	157. 5	15. 5	155	1. 5	152.

Sample Answer Sheet

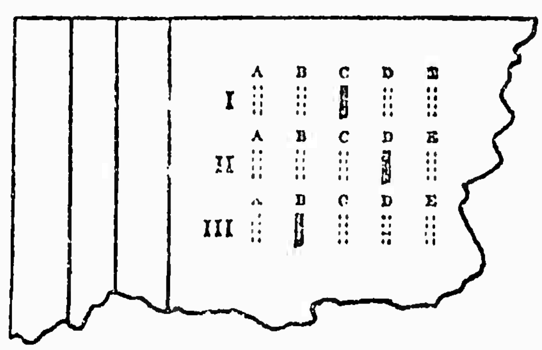


Figure A.1.—Dial Reading Test, CP622A.

Directions: This is a test to measure your ability to read tables quickly and accurately. In each problem in this part, a first value is given in the left-hand column and a second value is given in the right-hand column. You are to find in the table the entry that occurs at the intersection of the row and column corresponding to the values given. Then on your separate answer sheet blacken the space corresponding to the correct answer.

Look at sample problem I in the right-hand column on this page. Two values are given. The "First Value" (−17) is in the left-hand column. The "Second Value" (+14) is in the right-hand column. Now look at the big table on the page above. Notice the headings of the columns under FIRST VALUE across the top of the table. These run from −17 at the left-hand edge to +17 at the right-hand edge. Find the vertical column headed −17. Go down this column as far as the horizontal row that is labeled +14 at the edge of the table. At the intersection of the column headed −17 and the row labeled +14, the entry in the table is 69.

Look at the five choices in sample problem I. Choice A is 69, which is the correct answer. Therefore, space A has been blackened on the sample answer sheet below the sample problems.

Look at sample problem II. The first value for this problem is +17, so find the column in the big table headed +17. The second value is −10, so go down the column headed +17 to the row labeled −10. The entry at this intersection is 144. This is choice C in sample problem II. Therefore, choice C has been blackened on the sample answer sheet.

Next, you work through sample problem III, which is already answered on the sample answer sheet.

Now find the correct answers to practice problems 61 and 62 and blacken the appropriate spaces *on your separate answer sheet.*

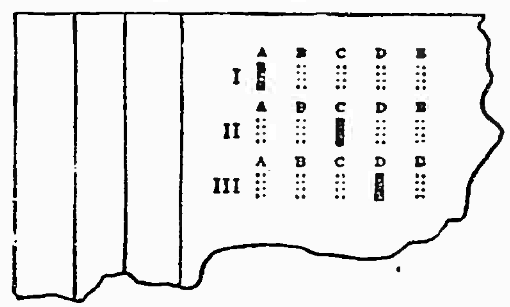
The correct answer to practice problem 61 is A; to practice problem 62 the correct answer is C. If you did not get these answers, work through the problems that you missed, and correct your errors by erasing the wrong answers completely and marking the correct ones.

Your score in this part will be the number of items marked correctly minus a fraction of the number marked incorrectly. Most people will not be able to complete all of the problems in the time allowed. Since this is a speed test, however, work as rapidly as possible without making mistakes. Stop immediately when you are told to do so. You are *not* to use a straightedge to follow down the columns. Be sure *not* to make pencil marks on the charts. You will be allowed [illegible] minutes.

SAMPLE PROBLEMS:

	First Value	Second Value	A	B	C	D	E
I	−17	+14	69	94	154	95	70
II	+17	−10	73	90	144	106	142
III	0	0	81	124	123	100	101

Sample Answer Sheet



A.

FIRST VALUE

FIRST VALUE

SECOND VALUE	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14	+15	+16	+17
+17	67	68	67	69	70	72	72	73	74	74	75	76	77	78	79	80	80	81	82	83	84	85	86	[illegible]	87	88	89	90	92	92	93	94	95	95	95
+16	68	67	69	70	71	72	73	74	74	75	76	77	78	79	80	[illegible]	81	82	83	84	85	86	87	88	88	89	90	91	92	93	94	95	96	97	98
+15	67	68	70	71	72	72	74	[illegible]	75	73	77	78	79	80	81	81	82	83	84	85	85	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	92	[illegible]	93	[illegible]	95	97	97	[illegible]	100
+14	69	70	71	72	73	74	74	75	75	77	73	79	80	81	82	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	[illegible]	[illegible]
+13	70	71	72	73	74	74	75	76	77	78	79	80	81	82	82	83	84	85	86	87	88	89	90	92	92	93	94	95	96	97	98	99	100	101	102
+12	71	72	73	74	[illegible]	75	76	77	78	79	80	81	82	83	83	84	85	86	87	88	89	[illegible]	91	92	93	94	95	96	97	98	99	100	101	[illegible]	103
+11	72	73	74	74	75	76	77	78	79	80	81	82	82	83	84	85	86	87	88	89	90	91	[illegible]	93	94	95	95	97	98	99	100	102	103	104	105
+10	73	74	74	75	76	77	78	79	80	81	82	82	83	84	85	85	87	88	89	90	91	92	93	94	95	[illegible]	98	99	100	101	102	103	104	105	106
+9	74	74	75	76	77	78	79	80	81	82	83	84	85	86	86	87	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	105	106	107	108
+8	[illegible]	75	76	77	[illegible]	79	80	81	82	83	83	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	[illegible]	[illegible]	[illegible]	[illegible]	107	[illegible]	110
+7	75	76	77	[illegible]	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	101	102	103	104	105	106	107	109	110	111
+6	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	100	101	102	103	104	105	107	108	109	110	112	113
+5	77	78	79	80	81	82	83	84	85	86	87	88	89	90	[illegible]	93	93	94	95	96	97	98	100	101	102	103	104	106	[illegible]	[illegible]	[illegible]	111	112	113	115
+4	78	79	80	81	82	82	[illegible]	84	85	87	88	89	90	91	92	93	94	95	96	[illegible]	99	100	102	102	103	105	106	107	108	110	112	112	114	115	116
+3	79	80	81	82	82	83	84	85	87	88	89	90	91	92	93	94	95	96	97	99	100	101	102	104	105	106	107	109	110	111	113	114	115	117	118
+2	80	81	81	82	83	84	85	87	88	89	90	91	92	93	94	95	96	98	99	100	101	102	104	105	106	107	109	110	111	113	114	115	117	118	120
+1	81	81	82	83	84	85	86	86	89	90	91	92	93	94	95	96	98	99	100	101	102	104	105	[illegible]	[illegible]	109	110	112	113	114	116	117	119	[illegible]	[illegible]
0	81	82	83	84	85	86	87	89	90	91	92	93	94	95	96	98	99	100	101	102	104	105	106	108	107	110	112	113	114	116	117	119	120	[illegible]	123
-1	82	83	84	85	86	87	89	90	91	92	93	94	95	96	96	97	100	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	109	[illegible]	112	113	115	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
-2	83	84	85	86	87	88	89	91	92	93	94	95	95	98	99	100	101	102	104	105	106	108	109	111	112	113	115	116	118	119	121	122	[illegible]	[illegible]	[illegible]
-3	84	85	86	87	88	89	90	[illegible]	93	94	95	96	97	[illegible]	100	101	102	103	105	106	[illegible]	[illegible]	111	112	114	115	116	118	119	121	122	124	126	[illegible]	129
-4	85	86	87	88	89	90	91	93	94	95	96	97	99	100	101	102	104	105	106	108	109	110	112	113	115	115	118	119	121	123	124	125	[illegible]	[illegible]	131
-5	[illegible]	87	88	89	90	91	92	94	95	96	97	98	100	101	102	104	105	104	106	[illegible]	111	112	114	116	117	118	[illegible]	121	123	124	[illegible]	126	[illegible]	[illegible]	[illegible]
-6	87	88	89	90	91	92	93	95	96	97	98	100	101	102	104	105	106	108	109	110	112	113	115	116	118	[illegible]	[illegible]	123	125	126	129	130	132	133	135
-7	87	[illegible]	90	91	92	93	94	96	97	[illegible]	99	101	[illegible]	103	[illegible]	[illegible]	[illegible]	109	110	112	113	115	116	118	120	[illegible]	123	125	[illegible]	128	130	[illegible]	134	[illegible]	137
-8	88	89	91	92	93	[illegible]	95	97	98	99	[illegible]	[illegible]	[illegible]	105	106	107	[illegible]	110	112	113	115	116	118	120	[illegible]	[illegible]	[illegible]	[illegible]	123	[illegible]	[illegible]	134	[illegible]	[illegible]	140
-9	89	90	92	93	94	95	96	98	99	100	102	103	104	106	107	109	[illegible]	112	113	115	116	118	120	121	[illegible]	125	[illegible]	127	[illegible]	132	134	135	[illegible]	[illegible]	142
-10	[illegible]	91	93	94	95	[illegible]	97	99	100	102	103	[illegible]	[illegible]	107	[illegible]	110	[illegible]	113	[illegible]	116	[illegible]	[illegible]	122	123	125	[illegible]	[illegible]	130	[illegible]	134	[illegible]	138	[illegible]	[illegible]	[illegible]
-11	91	92	93	95	96	97	98	100	101	[illegible]	[illegible]	105	107	108	110	112	[illegible]	115	[illegible]	[illegible]	[illegible]	[illegible]	123	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	140	[illegible]	[illegible]	147
-12	92	93	94	95	97	98	100	101	[illegible]	[illegible]	105	[illegible]	[illegible]	110	111	113	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	125	126	[illegible]	[illegible]	132	134	[illegible]	[illegible]	[illegible]	142	[illegible]	[illegible]	[illegible]
-13	93	[illegible]	95	[illegible]	[illegible]	99	101	102	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
-14	94	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
-15	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	101	102	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
-16	95	97	98	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	109	110	[illegible]	[illegible]	115	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
-17	96	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	110	111	113	115	116	[illegible]	[illegible]	122	[illegible]	[illegible]	[illegible]	[illegible]	131	133	135	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

FIRST VALUE

[illegible]

Fig. A2 – Table Reading Test, [illegible], part I.

Directions: In this test you will use the four tables [illegible], each one of which is used for a different air speed. These tables are used for finding either a drift correction or the ground speed of a plane when the air speed, the wind velocity, and the wind angle are known. Notice that each table contains three sections under the heading "Wind Velocity." Each of these three sections is, in turn, divided into two columns headed "Dri. Cor." (for Drift Correction) and "Gro. Spd." (for Ground Speed). At the left-hand edge of each of the four tables you will find wind angles from 0 to 90.

Your task is to use the four tables to find the "Dri. Cor." or the "Gro. Spd." from the information given in each item in three columns; in the left-hand column is the Air Speed, in the middle column is the Wind Velocity and in the right-hand column is the Wind Angle. At the right of these three columns is the statement of what you are to find—either the "Dri. Cor." or the "Gro. Spd." Look at sample problem I in the next column. Next, look at the table headed AIR SPEED 100 MILES PER HOUR. Then find the section in this table headed WIND VELOCITY 20 M. P. H. In this section, look at the column headed "Dri. Cor." Follow this column down to the row labeled WIND ANGLE 90. The entry in the table at this intersection is 12, and this is the correct answer. Of the five choices in sample problem I, choice D is 12. Therefore, on the sample answer sheet, space D has been blackened.

Now look at sample problem II. To find the correct answer, look at the table headed AIR SPEED 180 MILES PER HOUR. Then locate the section of the table headed WIND VELOCITY 10 M. P. H. In this section, look at the column headed "Gro. Spd." Follow this column down to the row labeled WIND ANGLE 50. The entry in the table at this intersection is 173, and this is the correct answer. Of the five choices in sample problem II, choice A is 173. Therefore, on the sample answer sheet, space A has been blackened.

Next, find the correct answers to practice problems 106 and 107 and blacken the appropriate spaces *on your separate answer sheet*.

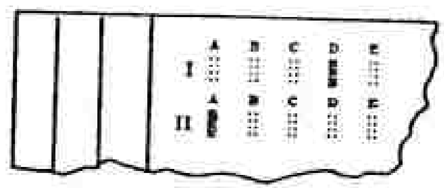
The correct answer to practice problem 106 is choice E; for practice problem 107 the correct answer is choice C. If you did not get these answers, work through the problems that you missed, and correct your errors by erasing the wrong answers completely and marking the correct ones.

Your score will be the number of [illegible] a fraction of the number of items marked incorrectly. Most people will not be able to complete all of the items in the time allowed. Since this is a speed test, however, work as rapidly as possible without making mistakes. Stop immediately when you are told to do so. You will be allowed 7 minutes.

SAMPLE PROBLEMS:

	AIR SPEED	WIND VELOC.	WIND ANGLE	What is the	A	B	C	D	E
I	100	20	90	Dri. Cor.	6	9	5	12	8
II	180	10	50	Gro. Spd.	173	133	176	175	170

Sample Answer Sheet



A.

AIR SPEED 100 MILES PER HOUR

WIND ANGLE	WIND VELOCITY 10 M.P.H. Dri. Cor.	10 M.P.H. Gro. Spd.	15 M.P.H. Dri. Cor.	15 M.P.H. Gro. Spd.	20 M.P.H. Dri. Cor.	20 M.P.H. Gro. Spd.
0	0	90	0	85	0	80
10	1	90	1	85	2	80
20	2	90	3	86	4	81
30	3	91	4	87	6	82
40	4	92	6	88	7	84
50	4	93	7	90	9	86
60	5	94	7	92	10	89
70	5	96	8	94	11	92
80	6	98	8	96	11	95
90	6	100	9	99	12	98

AIR SPEED 120 MILES PER HOUR

WIND ANGLE	WIND VELOCITY 10 M.P.H. Dri. Cor.	10 M.P.H. Gro. Spd.	15 M.P.H. Dri. Cor.	15 M.P.H. Gro. Spd.	20 M.P.H. Dri. Cor.	20 M.P.H. Gro. Spd.
0	0	110	0	105	0	100
10	1	110	1	105	2	100
20	2	110	2	106	3	101
30	2	111	4	107	5	102
40	3	112	5	108	6	104
50	4	113	6	110	7	106
60	4	[illegible]	6	112	8	109
70	5	116	7	114	9	112
80	5	118	7	116	9	115
90	5	120	7	119	10	118

AIR SPEED 140 MILES PER HOUR

WIND ANGLE	WIND VELOCITY 10 M.P.H. Dri. Cor.	10 M.P.H. Gro. Spd.	15 M.P.H. Dri. Cor.	15 M.P.H. Gro. Spd.	20 M.P.H. Dri. Cor.	20 M.P.H. Gro. Spd.
0	0	130	[illegible]	125	0	120
10	1	130	1	125	1	120
20	1	130	2	126	3	121
30	2	131	3	127	4	123
40	3	132	4	128	5	124
50	3	133	5	130	6	126
60	3	135	5	132	7	129
70	4	136	6	134	8	132
80	4	138	6	136	8	135
[illegible]	4	140	6	139	8	139

AIR SPEED 180 MILES PER HOUR

WIND ANGLE	WIND VELOCITY 10 M.P.H. Dri. Cor.	10 M.P.H. Gro. Spd.	15 M.P.H. Dri. Cor.	15 M.P.H. Gro. Spd.	20 M.P.H. Dri. Cor.	20 M.P.H. Gro. Spd.
0	0	170	0	165	0	160
10	1	170	1	165	1	160
20	1	170	2	166	3	158
30	2	171	2	167	3	163
40	2	173	3	168	4	[illegible]
50	2	173	4	170	5	[illegible]
60	3	175	4	172	6	169
70	3	[illegible]	5	[illegible]	6	172
80	3	[illegible]	5	177	6	176
[illegible]	3	[illegible]	5	[illegible]	6	[illegible]

TABLE READING TEST PART II

FIG. A 8.—Table Reading Test, CP621A, part II.

Reliability: $r = 0.84$. Correlation between two separately timed halves corrected for length. N = 1167.[4]

Sample items: See Figure A.2.

Test: General Information, CE505E.

Description: This test contains items based on knowledge of aircraft, aviation technique, automobile driving, mechanics, and sports and hobbies.

Number of items: Part I, 25; part II, 32; part III, 43.

Time limit: Part I, 10 minutes; part II, 12 minutes; part III, 14 minutes.

Scoring formula: R − W/4.

Reliability: $r = 0.87$. Correlation between odd and even items corrected for length. N = 1,000.

Sample item: Which one of the following is most commonly used to train pilots on the ground?

A. The Waco Trainer

B. The Ryan Trainer

C. The Fairchild Trainer

D. The White Trainer

E. The Link Trainer

Test: Instrument Comprehension I, CI615B.

Description: This test requires the subject to interpret readings of six aircraft instruments and to relate the instrument readings to verbal descriptions of an aircraft's performance. The test explains the function of the six instruments to the subject: Altimeter, artificial horizon, compass, rate of climb indicator, air speed meter, and turn bank indicator.

In each item the subject is presented with a drawing of the six instruments with the pointers showing readings and five verbal statements describing the action of the plane, such as "Flying level at 200 m. p. h., straight and unbanked, headed due south, gaining altitude at 9,800 feet." The subject must choose the one of five descriptions which fits the readings of the six instruments.

No. of items; 15.

Time limit; 12 minutes.

Scoring formula: 20 − (R − W/4).

Reliability: $r = 0.81$. Correlation between odd and even items corrected for length. N = 500.[5]

Sample Item:

A. Flying level at 200 m. p. h., straight and unbanked, headed due south, gliding altitude at 9,800 feet.

B. Flying level at 200 m. p. h., straight and unbanked, headed due south, losing altitude at 5,000 feet.

C. Flying level at 200 m. p. h., straight and unbanked, headed due south, maintaining altitude at 4,000 feet.

---

[4] *Ibid.* [5] *Ibid.*, chapter 19.

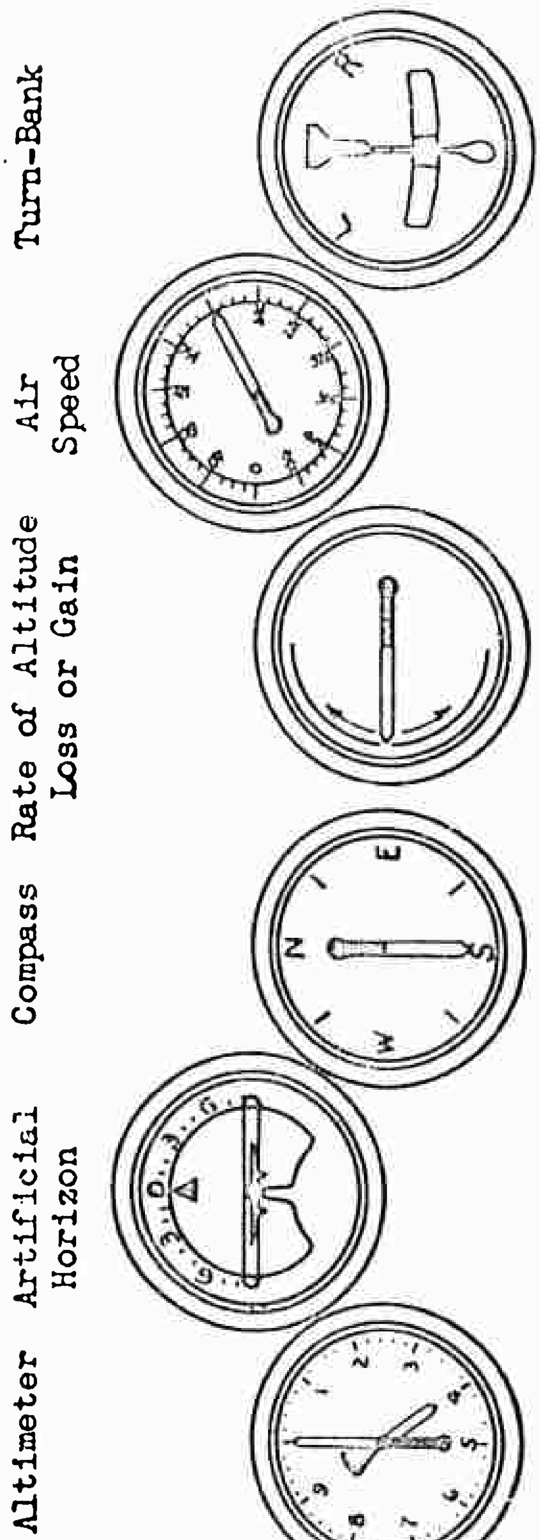


FIGURE A.4.—General Information Test, CE505E

D. Flying level at 200 m. p. h., straight and banked to left, headed due north, maintaining altitude at 4,000 feet.

E. Flying level at 200 m. p. h., turning properly to left, with 30° bank, maintaining altitude, headed due north at 4,000 feet.

Description C is the correct answer. Examine these dials again and check them carefully while description C is read to you.

Test: Instrument Comprehension II, CI616B.

Description: Each item consists of a representation of a compass and artifical horizon and five pictures of an aircraft in different attitudes of flight. The subject must indicate which picture corresponds to the readings on the compass and artificial horizon. The picture is considered to have been taken from the south and on a level with the aircraft, so that an aircraft heading south and flying straight and level is seen exactly head on.

Number of items, 60.

Time limit, 15 minutes.

Scoring formula: R – W/4.

Reliability: $r = 0.95$. Correlation between odd and even items corrected for length. N=500.[6]

Sample item: See Figure A.5.

Directions: In each of the problems in part II you will be given a picture of a single plane in five different positions. At the left of the picture you will be shown two dials, an artificial horizon and a compass. You are to choose the position of the plane which agrees with the readings on these dials. In reading the dials, remember that you are at the controls of the plane, looking forward.

Looking at the picture, a plane heading away from you is going north. Planes flying south will be coming directly toward you, while those going to your right will be headed east, and those to your left, headed west.

Problems A and B are samples. Examine the readings on the dials at the left of the pictures in Problem A. Now look at the five positions of the plane and select the position which is correct according to the readings on these dials.

According to the dials, the plane is flying level and unbanked and is headed due west. Note that the plane at position D is the only one which is correct for all these readings. Notice that position B is correct in every respect except that the plane is flying south. Position C would also be correct except that the plane is flying north. Remember that every reading on the dials must be checked in order to determine the correct position.

Test: Mathematics A, CI702F (General mathematics).

Description: This test is designed to measure knowledge of algebra and elementary trigonometry. The subject is required to solve equations and trigonometric problems.

Number of items, 35.

Time limit, 25 minutes.

Scoring formula: 2R – W/2.

Reliability: $r = 0.93$. Correlation between odd and even items corrected for length. N=1,000.[7]

[6] Ibid.

[7] Research Bulletin 44-18, Psychological Section, Office of the Surgeon, Headquarters, AAF Training Command.

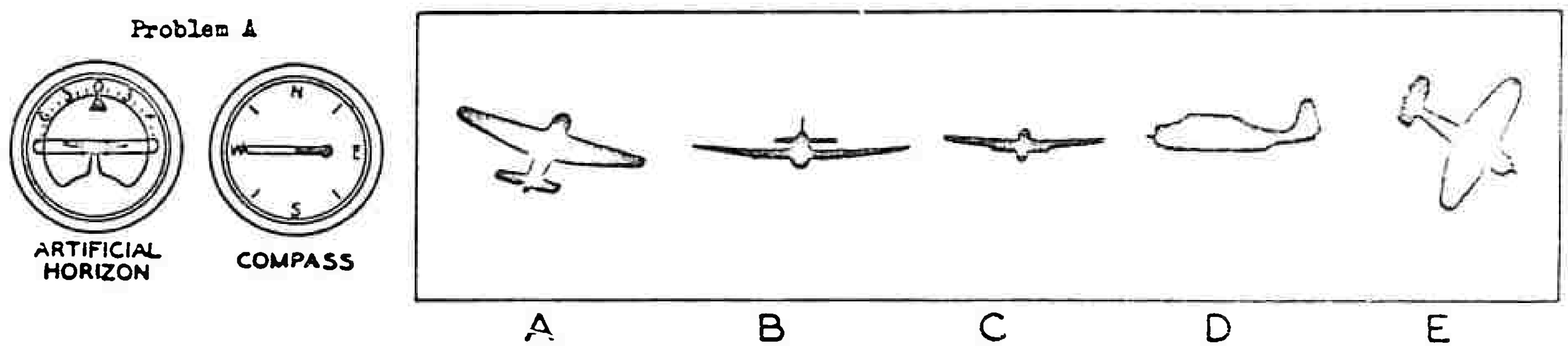


Figure A.5—Instrument Comprehension Test II, C1616B.

Sample item: If a = 3 and b = 1, the numerical value of $a^4 - b^4 \div (a - b)^3 =$

A. 98.
B. 78.
C. 66.
D. 48.
E. 34.

Test: Mathematics B, CI206C (arithmetic reasoning).

Description: This test requires the solution of arithmetic problems expressed in verbal form.

Number of items, 30.

Time limit, 35 minutes.

Scoring formula: 2R−W/2.

Reliability: $r = 0.84$. Correlation between odd and even items corrected for length. N = 1,000.[8]

Sample item: If plane A can fly 150 miles while plane B is flying 100 miles, how many miles can plane A fly while plane B is flying 250 miles?

A. 275 miles.
B. 300 miles.
C. 330 miles.
D. 350 miles.
E. 375 miles.

Test: Mechanical Principles, CI903B.

Description: This is a test of the ability to understand mechanical forces and movements. It includes items covering gear systems, mechanical movements, the principles underlying physical phenomena, and a number of items concerning levers, propellers, pulleys, etc.

Number of items, 40.

Time limit, 20 minutes.

Scoring formula: R−W/2.

Reliability: $r = 0.83$. Correlation between odd and even items corrected for length. N = 1,000.

Sample item: See Figure A.6.

Test: Reading Comprehension, CI614H.

Description: This test is made up of 8 paragraphs, each of which consists of 250 to 300 words in length. The paragraphs deal with technical topics including compass variation, dark adaptation, compass compensation, Mercator projection, the air speed meter, compass bearing and Bridgman's operational concepts. Each paragraph is followed by several five-choice questions.

---

[8] Guilford, J. P. and Lacey, J. I. *eds.*, *op. cit.*, chapter 7.

Number of items, 36.
Time limit, 30 minutes.
Scoring formula: 2R - W/2.
Reliability: r = 0.85. Correlation between odd and even items corrected for length. N = 1,000.

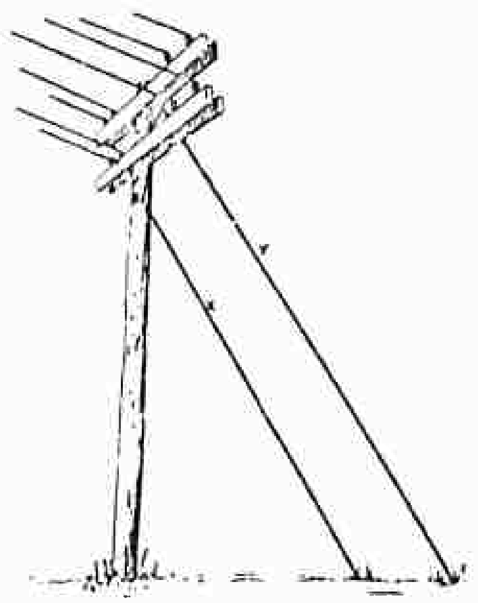


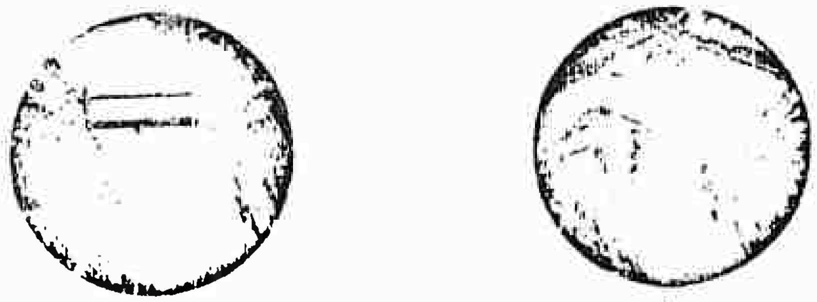
FIGURE A.6.—Mechanical Principles Test, CP303B.

Sample item: As the eyes become dark-adapted, the relative brightness of the different colors in the spectrum changes. The point of maximum brightness in an intense prismatic spectrum is in the yellow for the light-adapted eye. As the brightness of the total spectrum is lessened and the eye becomes dark-adapted, this point gradually shifts into the green. That is, the shift in brightness as intensity decreases is from the long wave lengths in the red end of the spectrum toward the short wave lengths in the blue end of the spectrum. If the eye is thoroughly dark-adapted and the intensity of the spectrum is further diminished, the spectrum becomes colorless just before it becomes invisible. In the colorless condition the spectrum still differs in brightness at different points.

16. The strain on the guy wire is—
   A. greater if attached in position X.
   B. greater if attached in position Y.
   C. the same if attached in either position X or Y.

Which of the following would have the shortest wave length in an intense spectrum?

A. Red.
B. Yellow.
C. Yellow-green.
D. Green.
E. Blue.



0 1

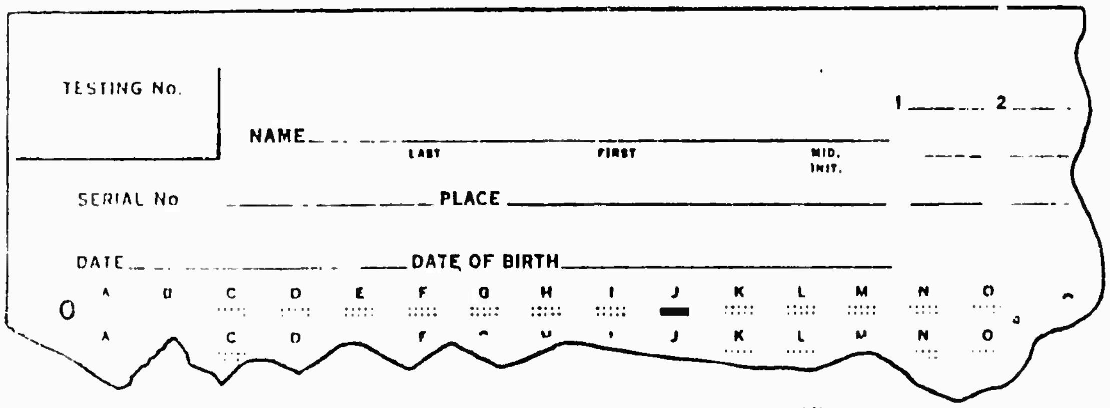
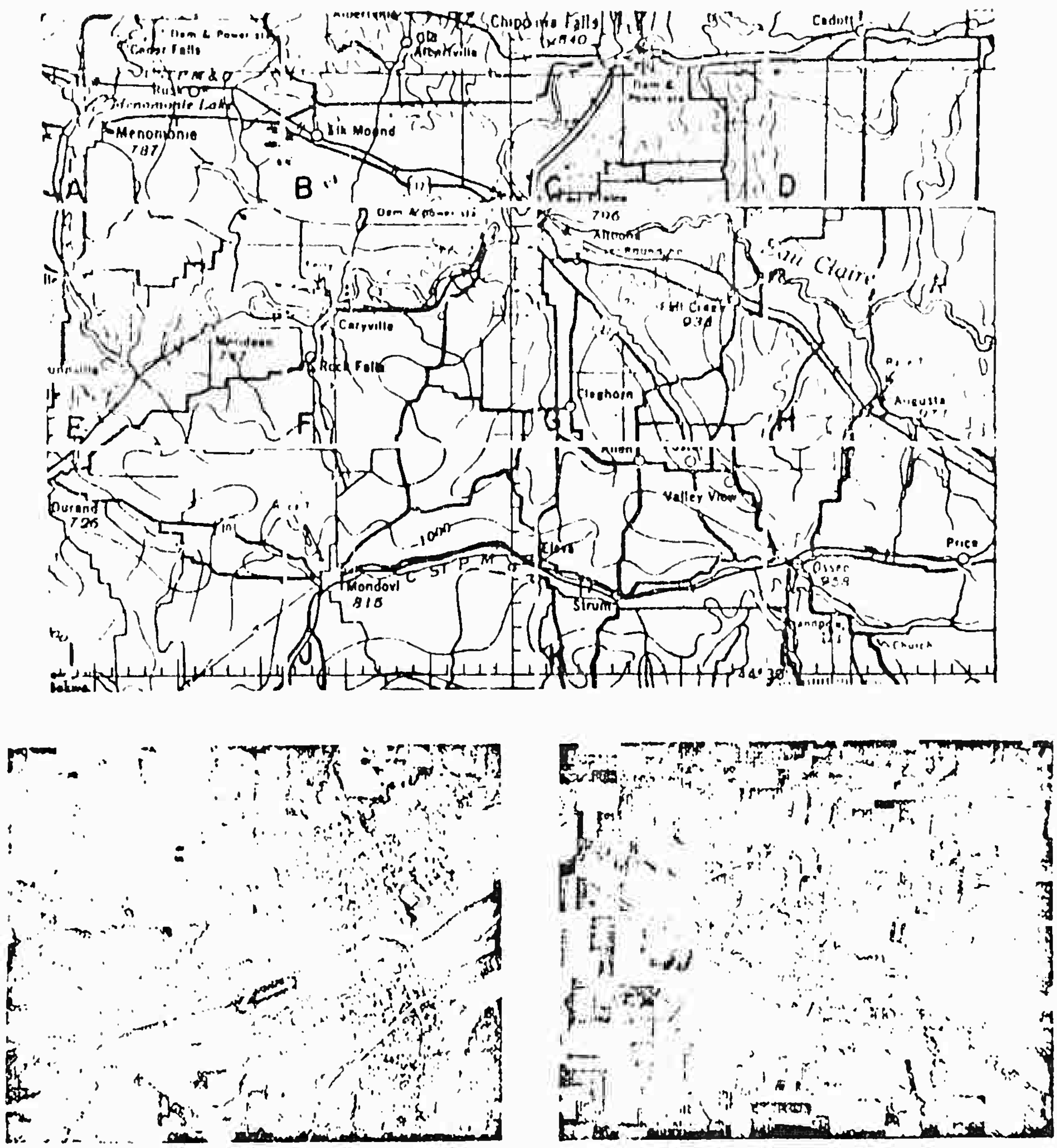


Figure A.7.—Spatial Orientation Test I, CP501B.

703327 47 (Faces p. [illegible]) No. 1



First Sample Item.

50327 – 47 (Faces p. 204) No. 2 Fig. A.8.—Spatial Orientation Test II, CP503B

Test: Spatial Orientation I, CP501B.

Description: This is a test of the ability to locate small sections of an aerial photograph within a larger picture. The test consists of eight large aerial photographs; six excerpts are to be located in each.

Number of items, 48.

Time limit, 5 minutes.

Scoring formula: R − W/5 in Classification Battery; total right in Radar Observer Selection Battery.

Reliability: r = 0.97. Correlation between odd and even items corrected for length. N=1,000.

Sample item: See Figure A.7.

Test: Spatial Orientation II, CP503B.

Description: This is a test of the ability to locate an area on a map corresponding to a section of an aerial photograph. The test consists of 12 pages, each of which contains a part of a standard aviation map in color, and aerial photos of four small areas within the area covered by the map. The subject must determine in which section of the map lies the area covered by the photo.

Number of items, 48.

Time limit, 3 minutes per part.

Scoring formula: R − W/5.

Reliability: r = 0.80. Correlation between odd and even items corrected for length. N=1,000.

Sample items: See Figure A.8.

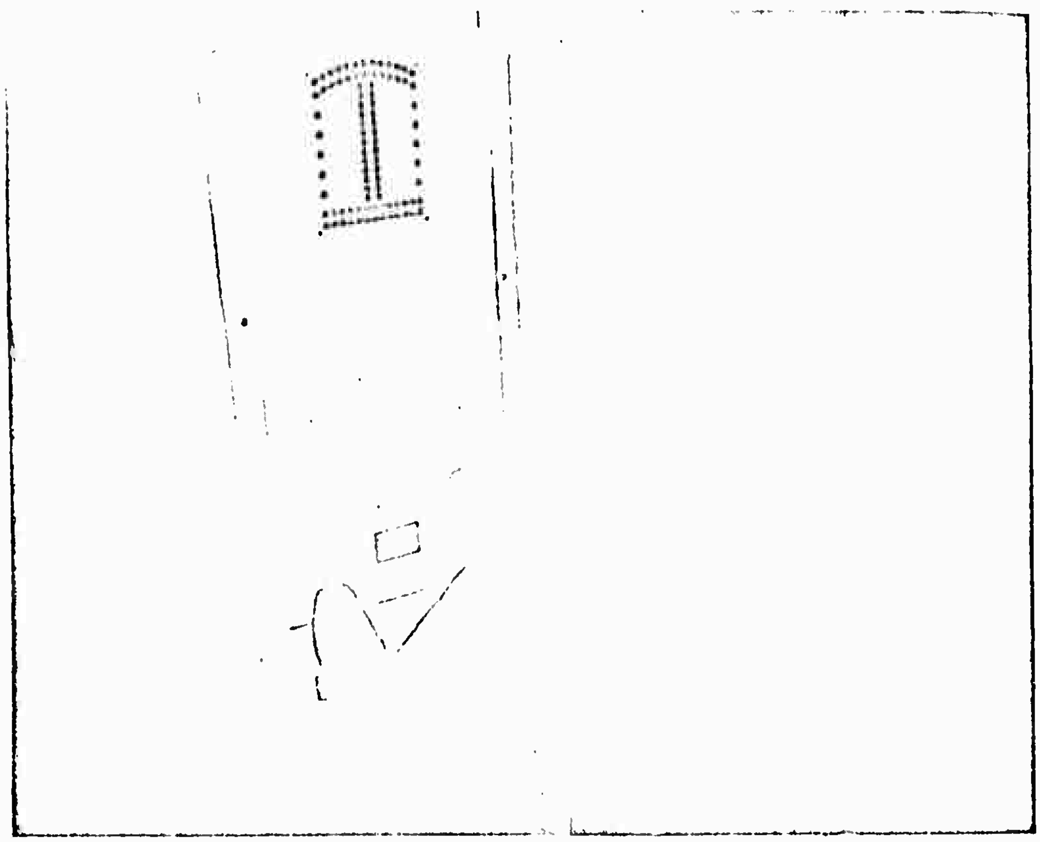
Psychomotor Tests

Test: Complex Coordination, CM701A.

Description: In this test, the subject operates controls similar to those used in an aircraft in flight. A stick, as in an aircraft, can be moved forward, backward, and laterally. The feet operate pedals similar to the rudder controls in a plane.

In front of the subject is a stimulus panel on which there are three rows of red lights and three corresponding rows of green lights. A pattern of red lights, one in each row, is presented to the subject. His task is to move his controls so as to turn on the green light corresponding to each of the red lights. By moving his stick from left to right, he can control the green lights in the top horizontal row; by moving his stick backward or forward, he can control the lights in the vertical row; by moving the pedals he can control the lights in the bottom horizontal row. As soon as each of the three red lights is matched by the corresponding green light, a new set of red lights is presented. The subject's task is to match as many sets of lights as possible within a specified period of time.

Number of items: One test period.
Time limit, 8 minutes.
Scoring formula: Number of patterns correctly matched during test period.
Reliability: $r = 0.91$. Correlation between odd and even trials corrected for length. N = 125.
Apparatus:



Second Sample Item.

FIG. A.9. Complex Coordination Test, CM701A.

Test: Discrimination Reaction Time, CP611D.
Description: In this test, the subject must react to the relative position of a red and green light on a stimulus panel. Four lights are arranged in the corners of a square. The upper left light and the lower right light are red. The upper right light and the lower left light are green. A white light is used as a warning signal; then a pair of lights, one red and one green, are illuminated. One of four toggle switches must be pushed, depending upon whether the red light appears above, below, to the right, or to the left of the green light. A time clock records the time until the correct toggle switch is pushed; the time is added for all trials.
Number of items, 80 trials.
Time limit: None.

Scoring formula: Total time necessary to complete 80 trials.

Reliability: $r=0.88$. Correlation between odd and even trials corrected for length. N = 125.

Apparatus:

FIGURE A.10.—Discrimination reaction time test, CP611D.

Test: Finger Dexterity, CM116A.

Description: The test consists of a form board (8 x 21 x 1 inches) containing 48 square holes ($^9/_{16}$-inch) spaced $1\frac{1}{2}$ inches apart, and 48 pegs $\frac{1}{2}$ inch square with round tops $\frac{3}{4}$ inch in diameter. The subject's task is to remove each peg in turn with his right hand, rotate it clockwise a half turn, and reinsert it in the hole. The lower and upper halves of each peg top are painted uniformly different to aid in determining the number correctly turned.

Number of items, five trials.

Time limit, 35 seconds per trial.

Scoring formula: Total number of pegs turned.

Reliability: $r=0.93$. Correlation between odd and even trials corrected for length. N = 125.

Apparatus: See Figure A.11.

Test: Rotary Pursuit with Divided Attention, CP410B.

Description: The subject's task is to keep the point of a stylus in contact with a metal target which is set in a revolving turntable, and

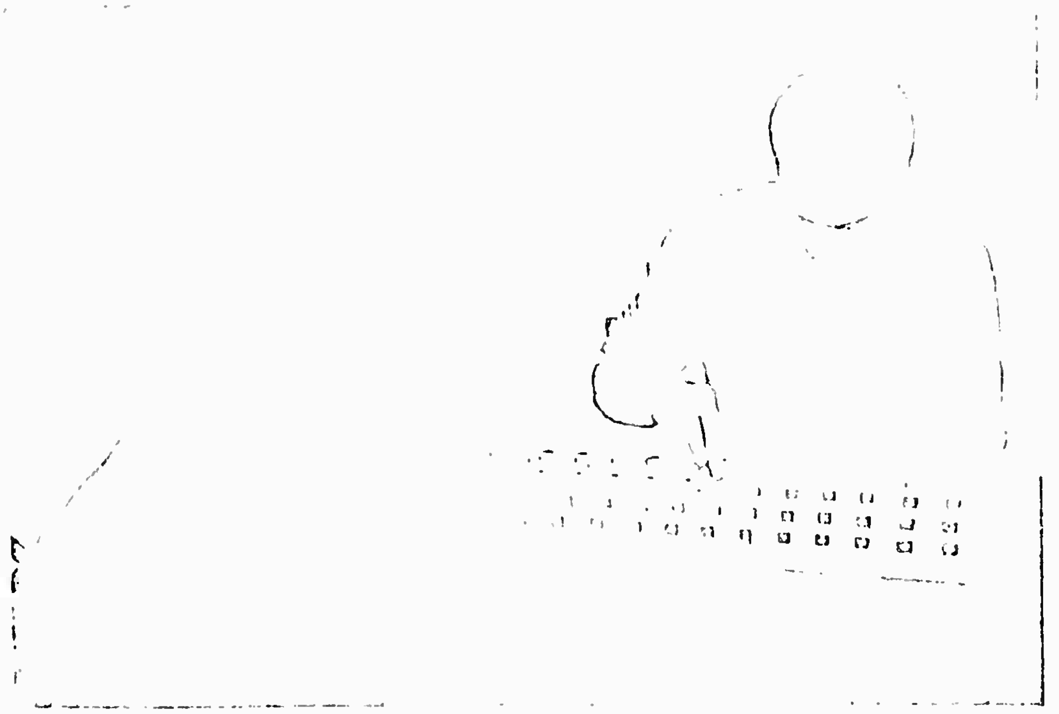


FIGURE A.11. Finger dexterity test, CM116A.

with his left hand to press down at all times either one of two buttons on a box to the left of the target. The turntable is rotated at a speed of 60 r. p. m. The button to be kept down depends upon whether a red or a green light is illuminated. The divided attention aspect of the test is the requirement that the subject must shift from one button to the other each time the signal light changes. He receives a score only when the stylus is in contact with the rotating target and the correct button is pressed down.

Number of items: Five trials without divided attention, 10 trials with divided attention.

Time limit: 20 seconds per trial.

Scoring formula: Total "contact" time.

Reliability: $r = 0.94$. Correlation between odd and even trials corrected for length. N = 125.

Apparatus: See Figure A.12.

Test: Rudder Control Test, CM120B.

Description: The subject is seated in a simulated cockpit which swings on a pivot and is mounted on a heavy base. The cockpit can swing to the right or left and is controlled by pedals similar to rudder controls in an aircraft. By varying the pressure on the pedals, the subject can keep the cockpit balanced in a central position. Pushing the right pedal forward turns the cockpit to the right and pushing the left pedal turns it to the left. The subject's task is to keep the cockpit centered so that a sighting bar mounted on the front of the cockpit is pointed at a target.

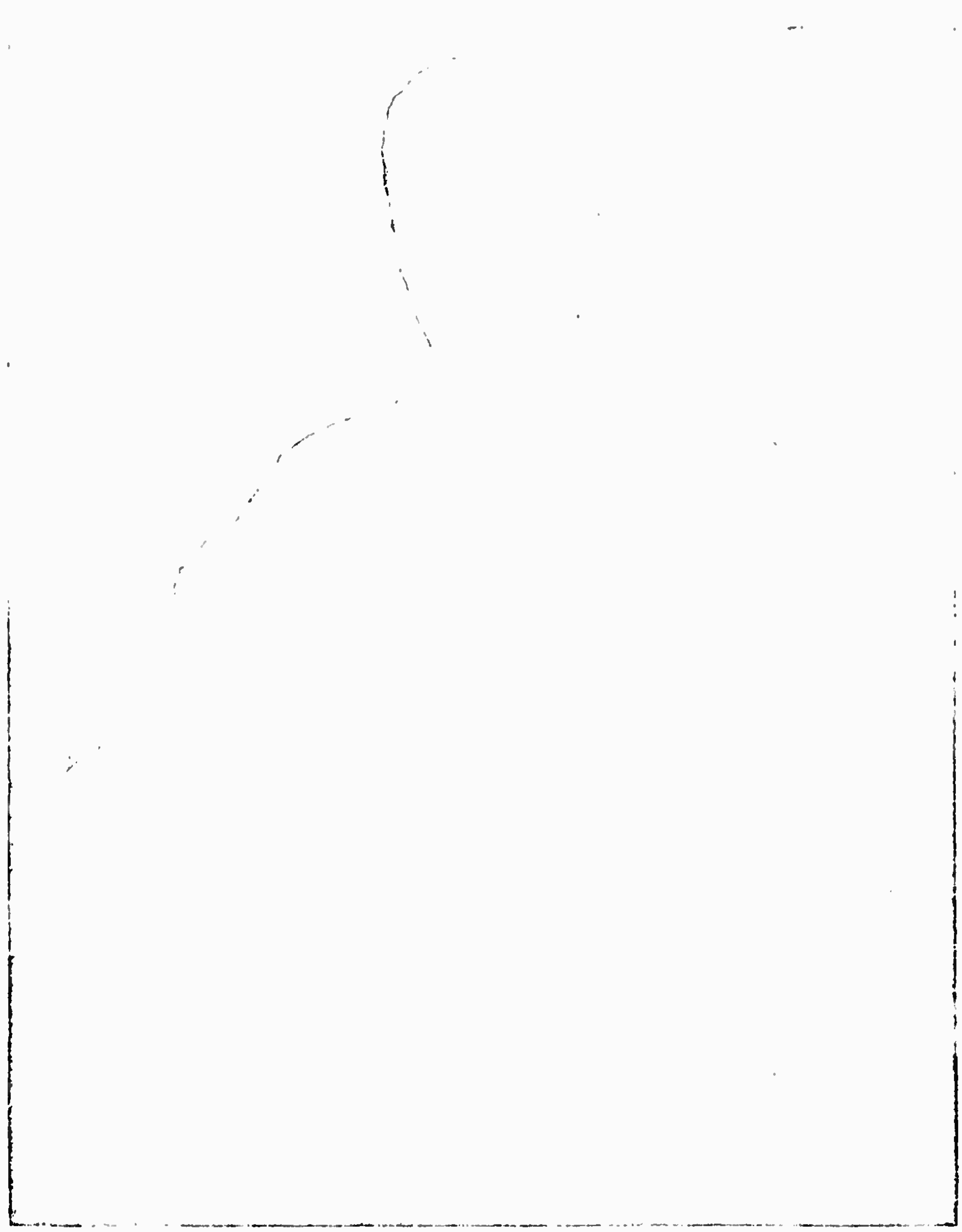


FIGURE A.12.—Rotary pursuit test, with divided attention, CP110B.

Items, 12 trials.

Time limit, 30 seconds per trial.

Scoring formula: Total time on target.

Reliability: $r=0.92$. Correlation between odd and even trials corrected for length. N=675.

Apparatus: See Figure A.13.

Test: Two-Hand Coordination, CM101A.

Description: The subject must control the movement of a carriage and maintain contact with a moving target disk. On the carriage an electric contact point is controlled by two lathe-type cranks. Each

703327—47——20

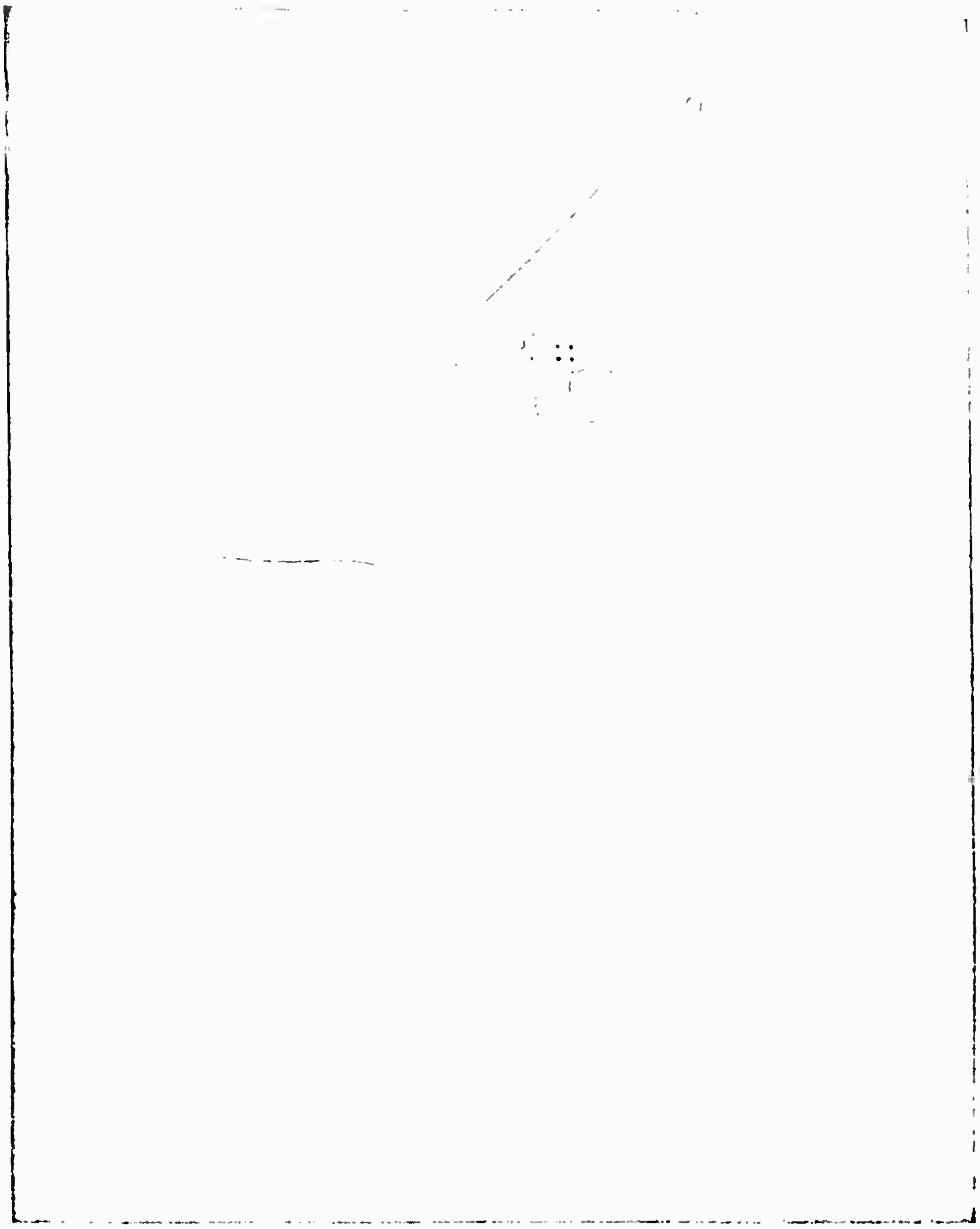


FIG. A.13.—Rudder Control Test, CM120B.

crank controls movement in one direction and by turning both cranks at once the movement of the contact point can be controlled simultaneously in both directions. The target is moved in an irregular path by a system of cams driven by an electric motor. The target makes one complete revolution a minute. The path of the target disk follows an irregular pattern, which is repeated every fourth trial.

Items, 8 trials.

Time limit, 1 minute per trial.

Scoring formula: Length of time contact is made.

Reliability: $r = 0.80$. Correlation between odd and even trials corrected for length. N = 125.

Apparatus:

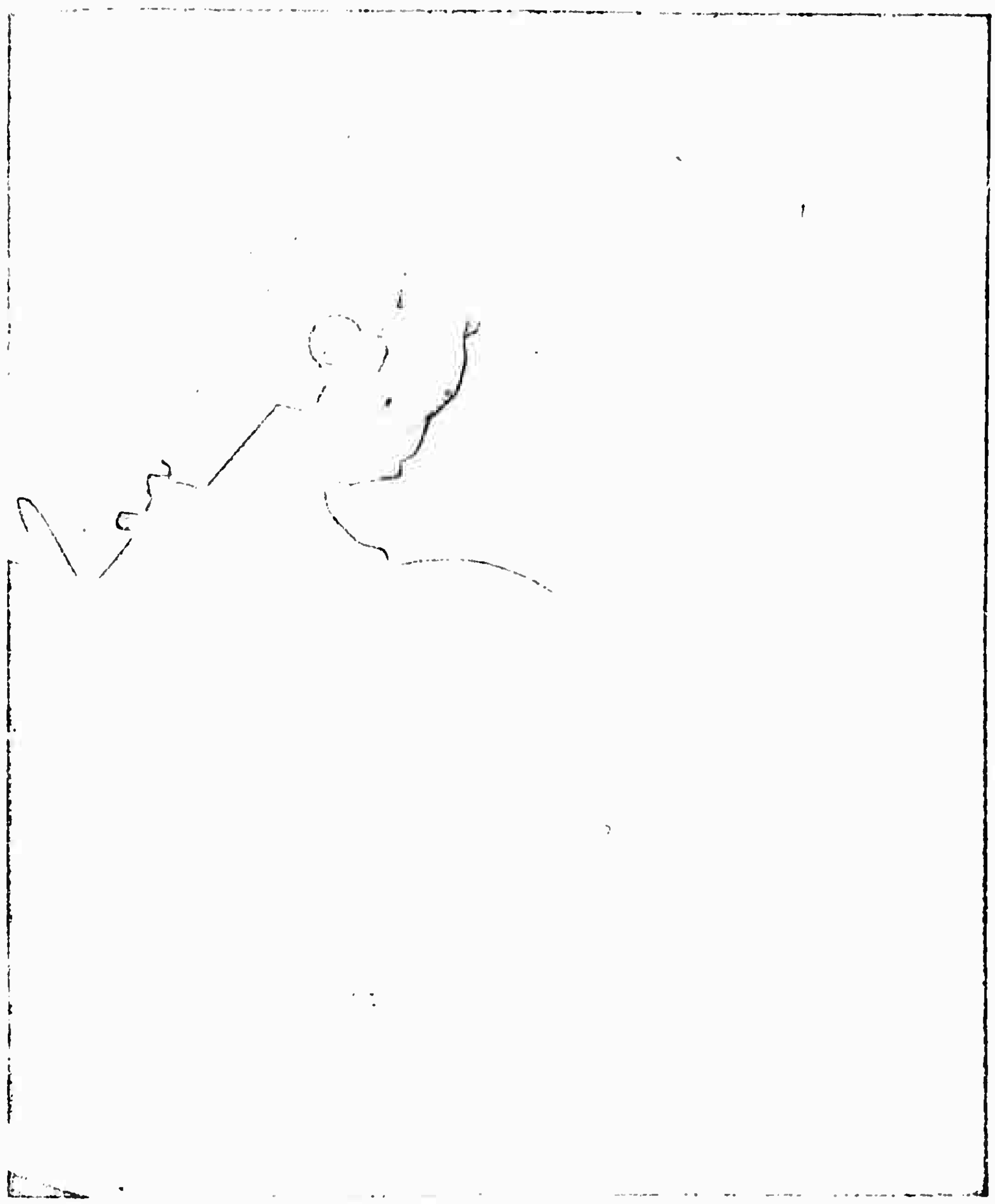


FIG. A.14.—Two-Hand Coordination Test, CM101A.

RADAR OBSERVER SELECTION BATTERY

Test: Coordinate Reading, CP224B.

Description: This test consists of a circular graph simulating an oscilloscope screen. The circle is graduated in degrees from 0° to 360°. A scale graduated in miles runs from the center to the edge and concentric circles appear at 10-mile intervals from the center. Located within the circle are dashes representing target returns on the oscilloscope screen. The task for each item is to determine the bearing and ranges of a dash line from the center of the circle. Of the five choices for bearing and range readings, only the last digits are used.

Number of items, 85.

Time limit, 20 minutes.

Scoring formula: Total right.

Reliability: $r = 0.90$. Correlations obtained by split-half technique corrected for length. $N = 100$.[9]

Sample items: See Figure A.15.

Test: Oscilloscope Interpretation, CP817A.

Description: This test approximates the recognition of radar oscilloscope signals through interference. Three standard types of signals are used: irregular forms, various sized curved lines, and various sized blips. The test consists of a series of circles and rectangles, each of which contains a number of one of the three signals. The task is to count the number of signals which are hidden by interfering lines. An answer key is presented to enable the answer to be entered on a standard IBM A–O answer sheet.

Number of items, 90.

Time limit, 20 minutes.

Scoring formula: Total right.

Reliability: $r = 0.90$. Correlation obtained by split-half technique corrected for length. $N = 152$.[10]

Test: Pattern Orientation, CP816A.

Description: In each item, a pattern of circles is shown in a square on the left side of the page. In a large circle on the right side of the page the same pattern is rotated and shown again along with other circles. The task is to identify the pattern. In order to do this, a cross is presented in the first pattern, while in the second pattern lettered crosses are presented, only one of which corresponds to the cross in the first pattern. The task is to determine which of the five lettered crosses corresponds to the cross in the first pattern.

Number of items:

Part I: 24.

Part II: 24.

Time limit:

Part I: 10 minutes.

Part II: 10 minutes.

Scoring formula: Total right.

Reliability: $r = 0.71$. Correlation between part I and part II corrected for length. $N = 170$.[11]

---

[9] OSRD Report No. 1813, Formal Memorandum No. 4, Preliminary Report of Results from Oscilloscope Operator Tests

[10] NDRC Informal Memorandum No. 22, 20 March 1945, Oscilloscope Interpretation Test for the Selection of Radar Operators.

[11] Progress Report for Pathfinder Project of AERD No. 1, Sept. 1944, p. 9.

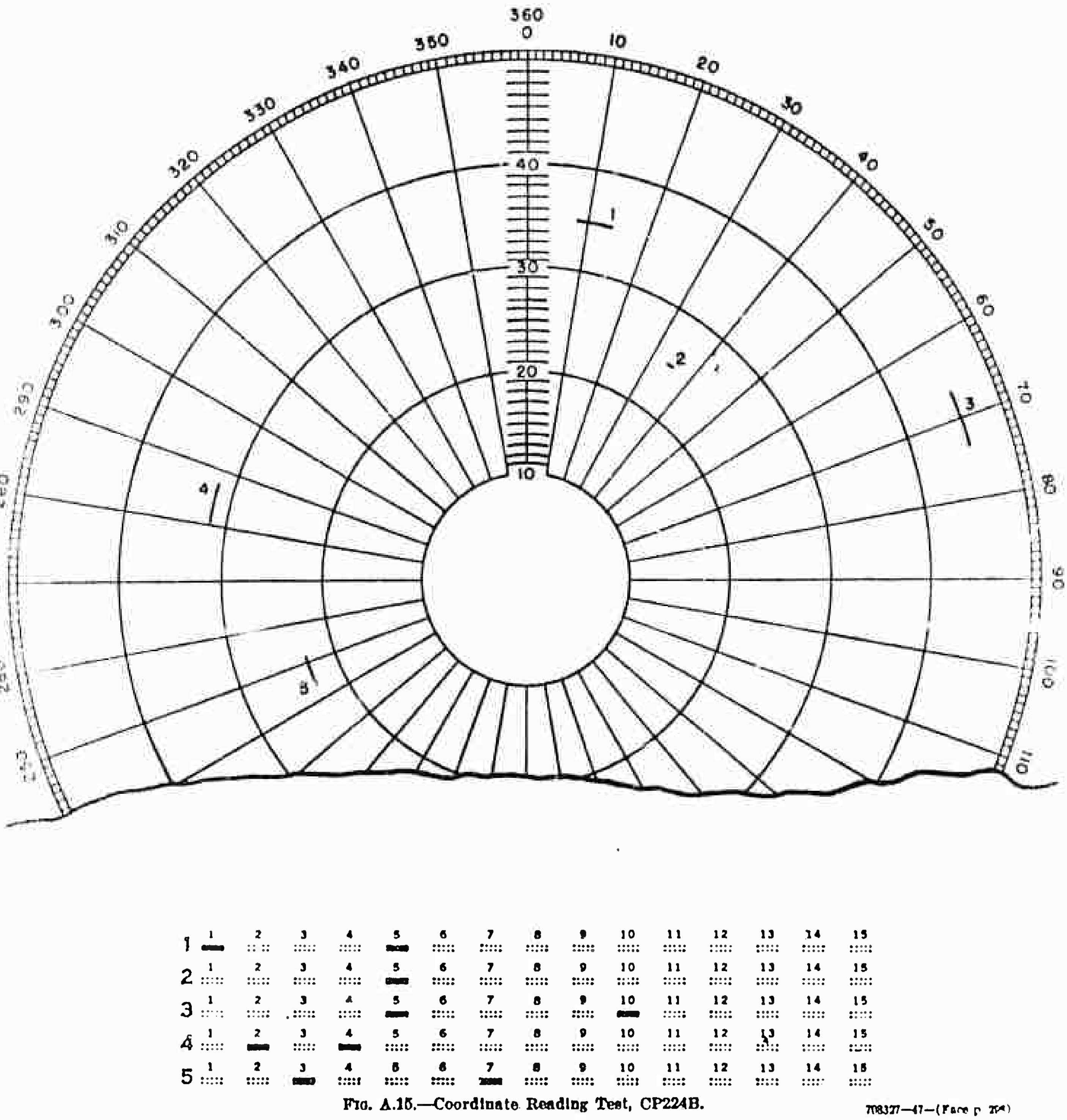


Fig. A.15.—Coordinate Reading Test, CP224B.

Sample items:

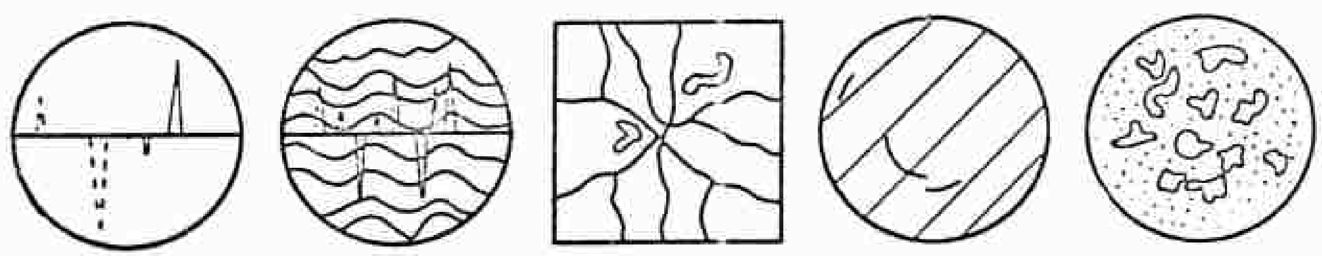
The three types of signals which will be seen in the test are: sharp blips like this , which may vary in size; small curved lines like this , which may vary in size; and irregular figures like this . The irregular figures will always be the same size. The curved line signals always follow the curve of the circle in which they appear. In the test items other lines will be drawn to make it difficult to see the blips or figures.

The practice test below will show you how to work the test. Count the number of blips in the circle P below. This number should be 4. Now look at the Answer Key which follows. The letters A through O refer to the spaces on your special answer sheet.

ANSWER KEY:

If you count:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14 or more
Then your answer is:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O

Since you counted 4 blips in the circle P, your answer would be E. Circle Q has 8 blips; therefore your answer for this item is I. Do not make any marks on your answer sheet, until you begin the actual test. Work problems R, S and T now, and find the correct letter by using the answer key above.

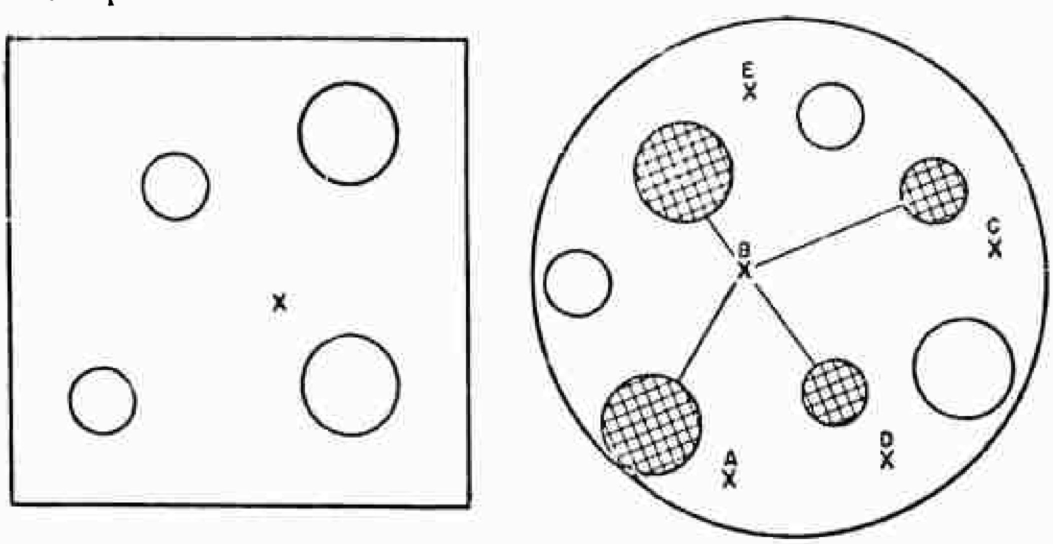


P. Q. R. S. T.

The correct answer to item R is C; item S is D; item T is L.

FIGURE A.16.—Oscilloscope Interpretation, CP817A.

Sample item:



In the square, note the pattern of two small and two large circles. On the right, note that the four cross-hatched circles form the same pattern. The pattern has been rotated so that now the two small circles are to the right rather than to the left of the two large circles.

In the figure at the right there are five crosses lettered A, B, C, D, and E. Look at the cross marked B. It bears the same relationship to the four circles as does the cross X in the square on the left. Therefore, B is the answer to problem 1. Indicate that B is the answer to problem 1 by blackening the appropriate space on your answer sheet.

FIGURE A.17.—Pattern Orientation Test, CP816A.

Test: Spatial Orientation I, CP501B.

Description: This test is described in the preceding section.

Test: Radar Preference I.

Description: See chapter 10, page 224.

COMPOSITION OF AIR-CREW SPECIALTY STANINES

Bombardier stanine:

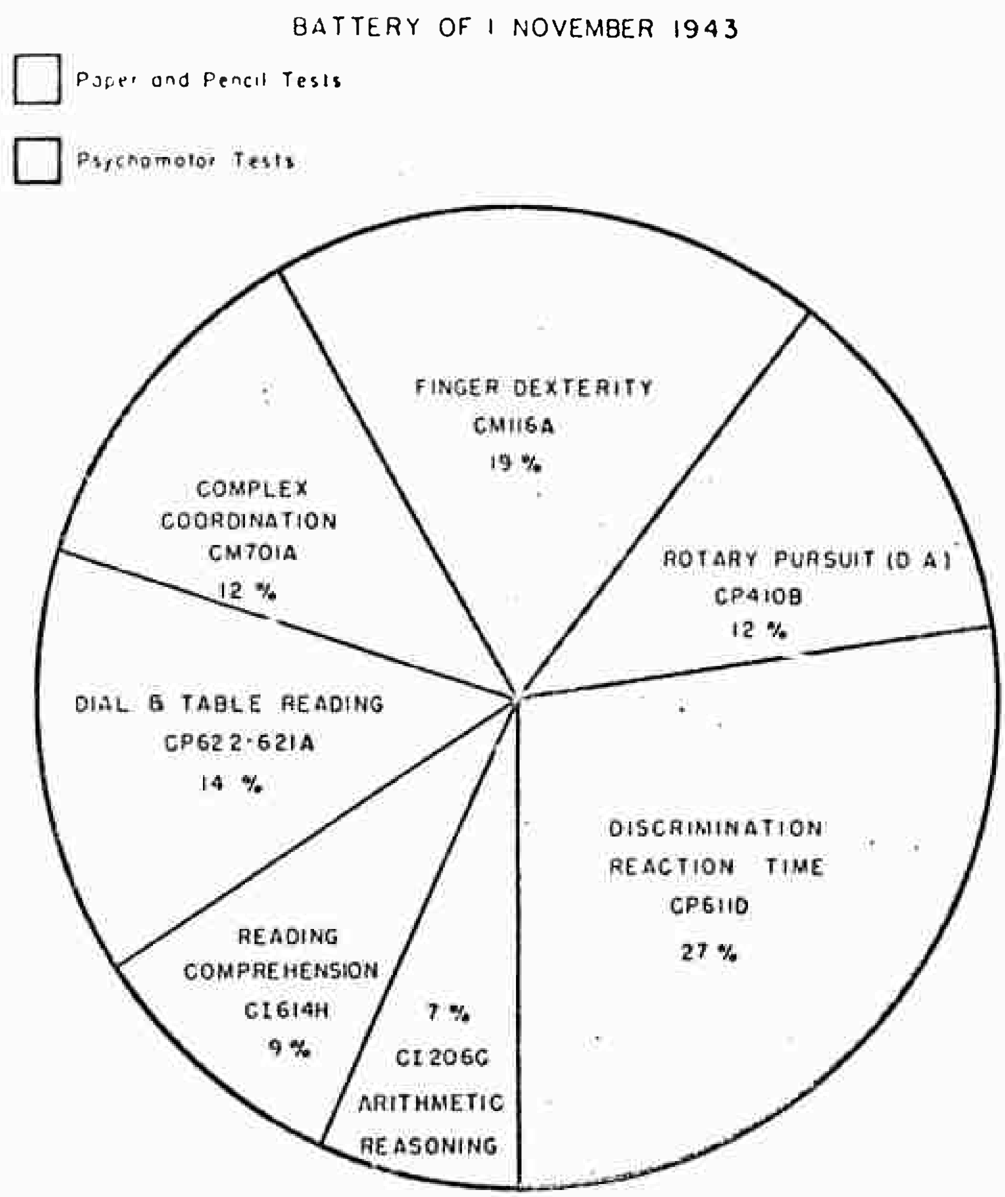


FIGURE A. 18.—Composition of bombardier stanine. Percent contributed by each test to the composite bombardier aptitude rating. Battery of 1 November 1943.

Navigator stanine:

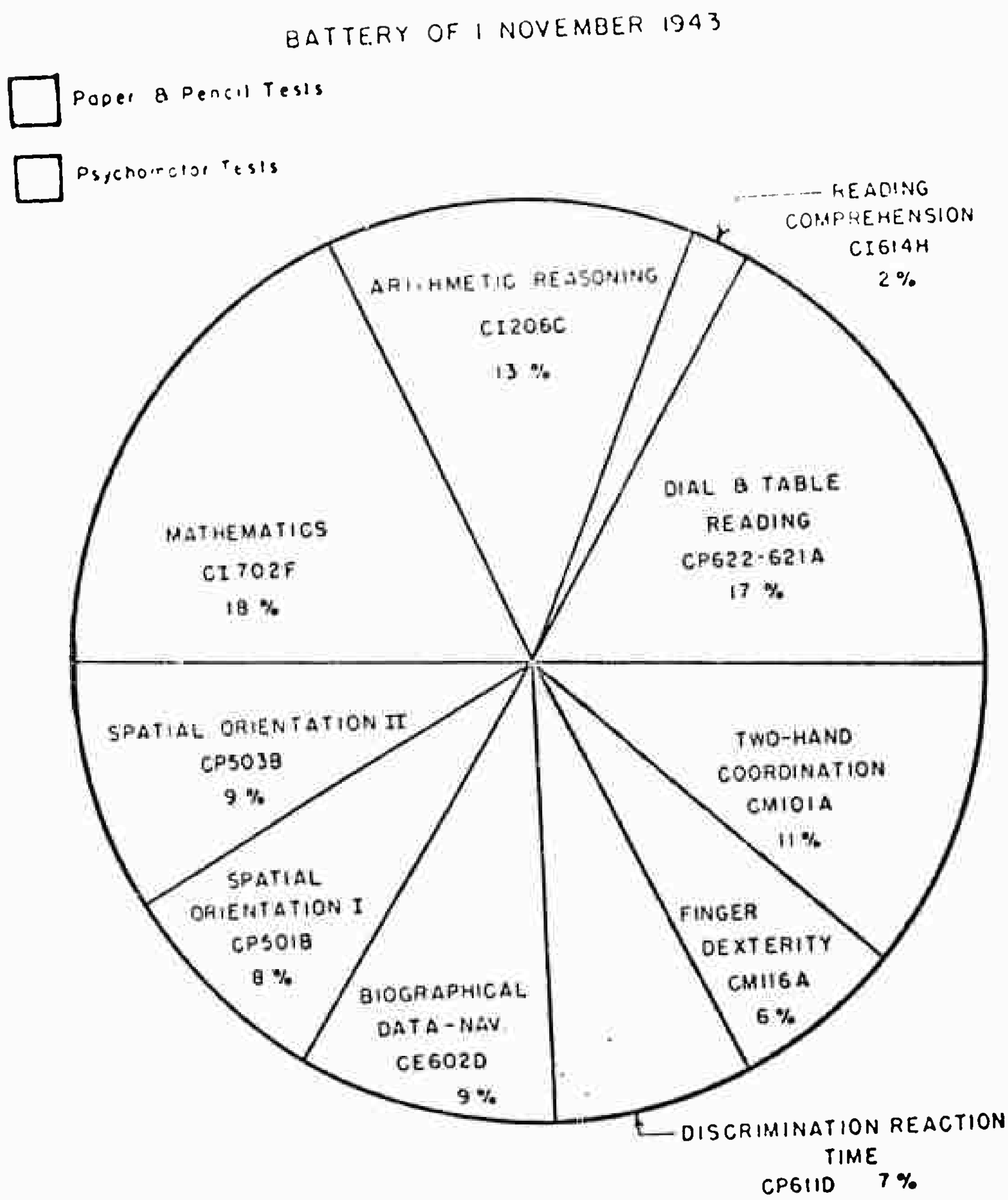


FIGURE A.19.—Composition of navigator stanine. Percent contributed by each test to the composite navigator aptitude rating. Battery of 1 November 1943.

Pilot stanine:

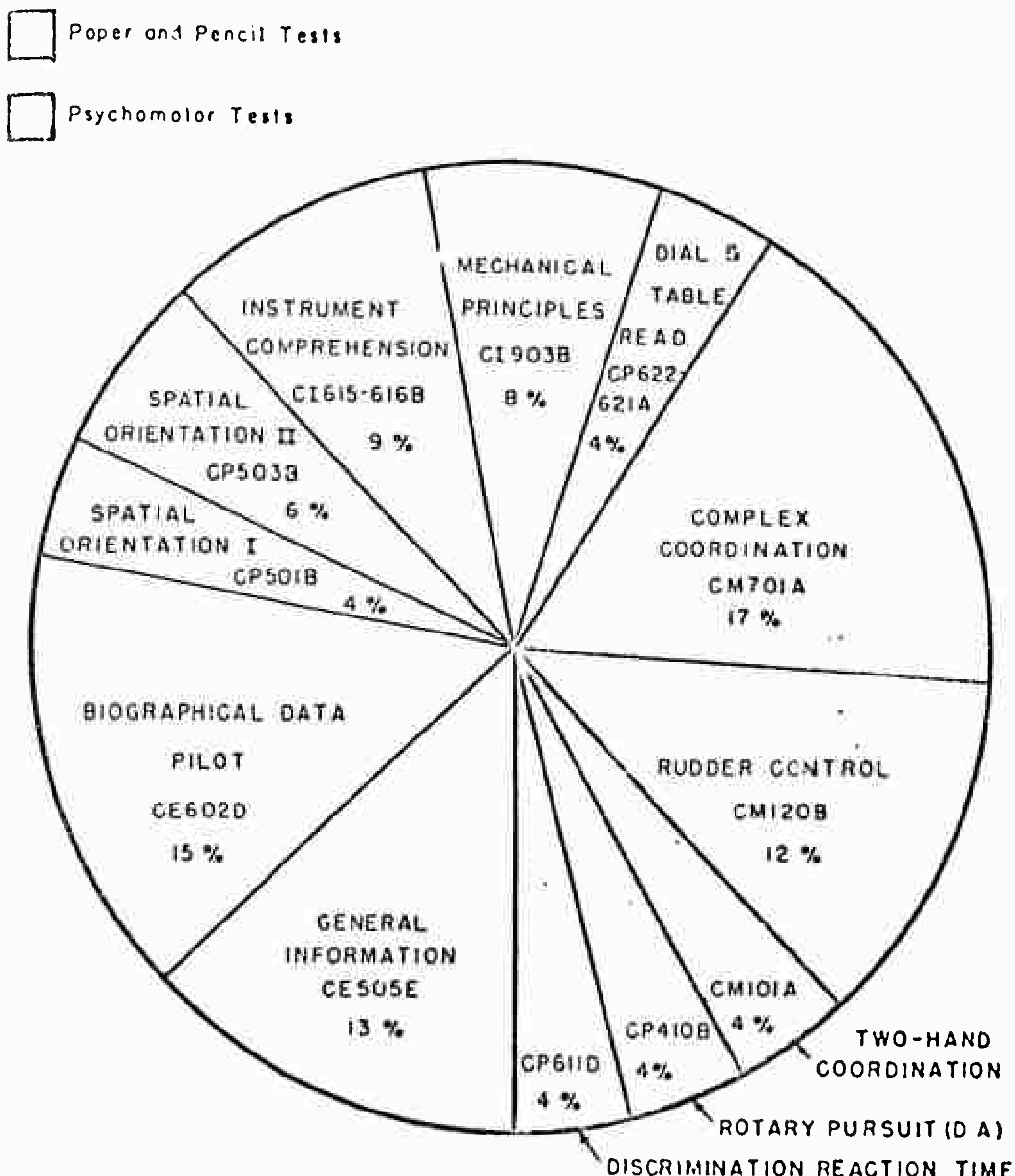


Fig. A.20

FIGURE A. 20.—Composition of pilot stanine. Percent contributed by each test to the composite pilot aptitude rating. Battery of 1 November 1943.

Radar observer stanine: The composition of this stanine is described in chapter 10, page 224, as derived from the Radar Observer Selection Battery. Data necessary for the preparation of a diagram are not available.

EXPERIMENTAL BATTERY

Printed Tests

Test: Aerial Orientation, CP520A.

Description: Each item consists of a cockpit view of the terrain over which an aircraft is flying. To the right of the view are five pictures, each showing a plane in a different position over the terrain. The task is to match the cockpit view with the correct plane position.

Number of items:

Part I: 30.

Part II: 28.

Time limit:

Part I: 10 minutes.

Part II: 8 minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r = 0.89$. Correlation between part I and part II corrected for length. N=443.

Sample item:

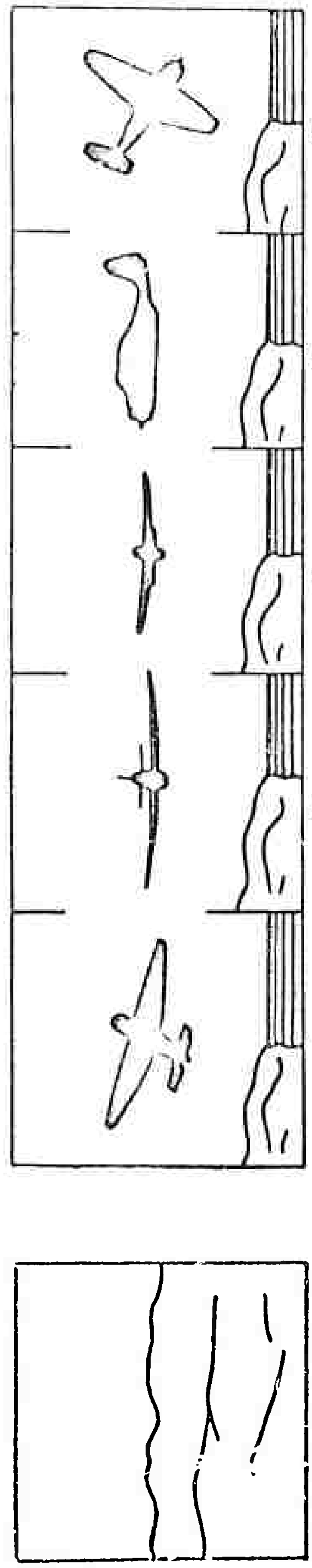


FIGURE A.21.—Aerial Orientation Test, CI520A.

Test: Area Visualization, CP815A.

Description: Each item consists of two geometric drawings which must be combined to form one of three geometric drawings labeled A, B, and C. The task is to determine which of the labeled drawings is the result of fitting the first drawing together properly.

Number of items:

Part I: 30.

Part II: 30.

Time limit:

Part I: 7 minutes.

Part II: 7 minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: No data available.

Sample item:

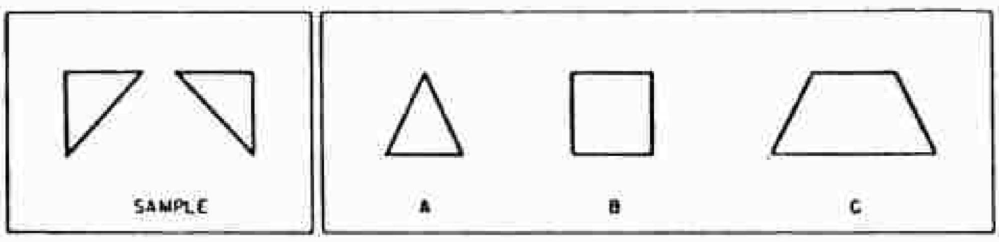


FIGURE A.22.—Area Visualization Test, CP815A.

If the two triangles are rotated, they would fit together to form a square. Therefore, the answer to the sample problem is B.

The picture at the left shows a cockpit view or the view seen by the pilot as he looks out over the nose of his plane. Each of the five pictures at the right shows a plane in a different position over the coast line. Your task is to match the cockpit view with the correct plane position. Notice in each picture that the coast line runs directly away from you as far as the eye can see. Notice also that ocean is on your right and land is on your left.

D is the correct answer to problem 1. You can tell from the cockpit view at the left that the plane is flying level, unbanked, and is headed directly toward the mountains. A more detailed explanation will follow. Blacken space D after item No. 1 now.

In order to select the plane position from the cockpit view you must consider bank, climb or dive, and direction of flight.

Test: Air Corps Vocabulary, 1942.

Description: This is a speeded test. Each item is a word followed by five choices of synonyms.

Number of items, 150.

Time limit, 15 minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: No data available.

Sample item: Carnival:

A. Slaughter.
B. Banquet.
C. Funeral.
D. Sensual person.
E. Festival.

Test: Compass Orientation, C1660A.

Description: The subject assumes that he is in an aircraft. He is told in what direction he is flying, and that he makes a turn, right or left. His task is to determine his new direction. The test is highly speeded since 5 minutes are allowed to do the 150 items.

Number of items, 150.

Time limit, 5 minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: No data available.

Sample items:	You are flying—	and turn	New direction
	1. North	left	West (answer).
	2. West	right	
	3. North	right	

Test: Estimation of length, CP631A.

Description: Five bars of standard lengths arranged in order from A to E are shown. Part I of the test consists of bars of different lengths. The task is to match each bar with the correct standard bar. In part II of the test the same standard bars are used. The item bars, however, are double length. The task is to determine which standard bar has been doubled to form each item bar.

Number of items:

Part I: 75.
Part II: 75.

Time limit:

Part I: 4 minutes.
Part II: 5 minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r=0.65$ for rights, 0.72 for wrongs. Correlation between part I and part II corrected for length. $N=586$.

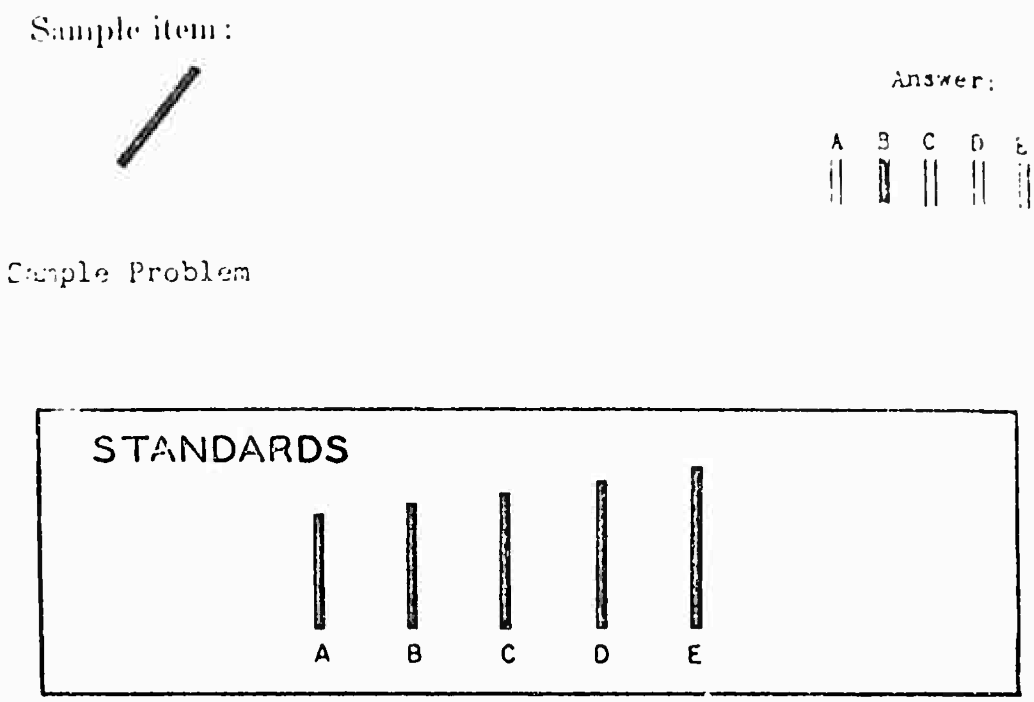


FIGURE A.23.—Estimation of Length Test, CP631A.

This sample bar is seen to be of the same length as standard B. The correct answer to this item, then, is B. Consequently the space under B in the sample answer has been blackened.

Test: Flight Orientation, CP528A.

Description: In this test each item consists of two pictures. The picture at the left shows a cockpit view of the terrain over which an aircraft is flying. The picture at the right shows the same cockpit view as it appears after the plane has performed a single maneuver. The task is to determine which one of six possible maneuvers the plane has performed: left or right turn, left or right roll, climb up or down.

Number of items:

Part I: 47.

Part II: 50.

Time limit:

Part I: 8 minutes.

Part II: 14 minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r=0.78$. Correlation between part I and part II corrected for length. N=502.

Test: Mechanical Information, CI905B.

Description: This test contains verbal items concerning automobile mechanics and the use of tools.

Number of Items: 30.

Time Limit: 12 minutes.

Scoring Formula: Two scores were obtained, total right and total wrong.

Look at Sample Problem 1 below. Two pictures are shown. The picture at the ft shows a cockpit view. The picture at the right shows the cockpit view as appears after a single maneuver. Your task is to determine the maneuver. he maneuver will be one of the following: Left or Right Turn, Left or Right oll, Climb Up or Down.

What is the maneuver in Sample 1?

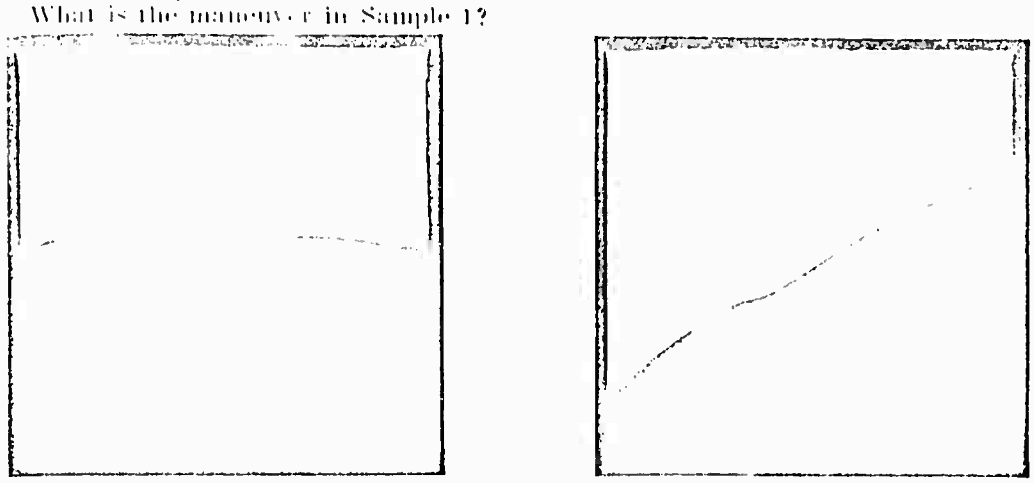


FIGURE A.24.—Flight Orientation Test CI528A.

Roll right is correct. The first picture shows the cockpit view as it appears while flying straight and level toward a mountain range. The second picture shows the view as it appears in a right bank. To move from the first position to the second the pilot has rolled the plane to the right. The answer to Item No. 1 should be marked as shown on your answer sheet.

Reliability: No data available.

Sample Item: A main bearing supports a—

A. camshaft.
B. universal.
C. driveshaft.
D. crankshaft.

Test: Memory for Landmarks, CI510AX2.

Description: A page containing 15 landmarks (e. g. rivers, lakes, etc.), each of which is named, is studied for 4 minutes. The page is then turned and the same landmarks are shown without the names. The task is to select, from a list of 15 names, the correct name for each landmark.

Number of Items: Three parts with 12 items per part.

Time Limit: 4 minutes study period per part, and 4 minutes to answer 12 items per part.

Scoring Formula: Two scores were obtained, total right and total wrong.

Reliability: r=0.82. Correlation between part I and part II of form CI510AX1 corrected for length. N=238.

Sample Items: Landmarks to be studied.

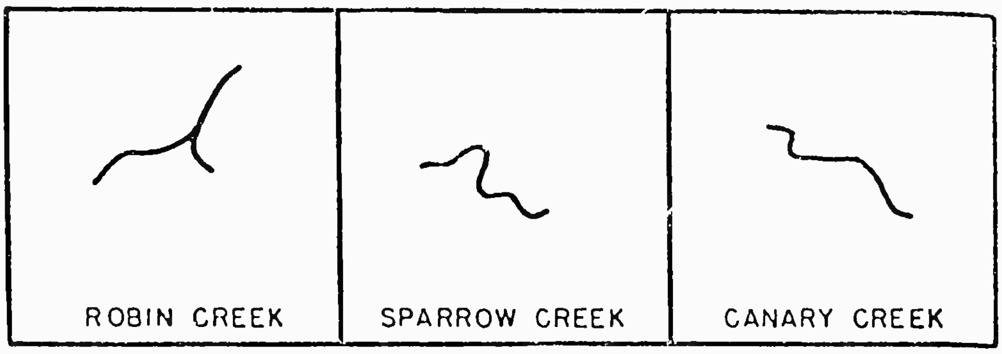
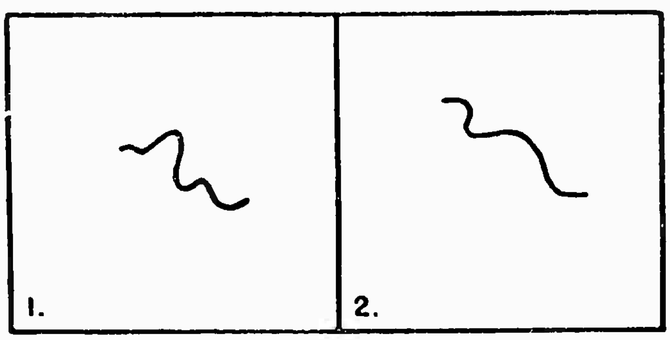


FIGURE A.25.—Memory for landmarks test CI510AX2 (samples)

Two sample problems.



A ROBIN CREEK
B. SPARROW CREEK
C. CANARY CREEK

FIGURE A.26.—Memory for landmarks test CI510AX2 (problems)

Test: Numerical Operation, CI702BX1.

Description: This test contains 65 problems in addition, multiplication, subtraction, and division. Whole numbers, fractions, decimals, and percentages are used. Ten items require approximate answers to more complicated problems.

Number of Items: 75.

Time Limit: 10 minutes.

Scoring Formula: Two scores were obtained, total right and total wrong.

Reliability: Correlation between the first and second half of form CI702B is 0.68. This is the lower limit of the reliability coefficient as the two parts are not comparable in content. N = 1774.

Sample Item: Perform the following numerical computations:

Add: 5 + 8 + 44 = (A) 54 (B) 55 (C) 56 (D) 57 (E) 58.

Approximate: $\frac{6{,}125 \times 8}{30 \times 15} =$ (A) 108.89. (B) 136.7. (C) 141.2. (D) 204.1. (E) 312.2.

Test: Pattern Comprehension, CP803A.

Description: This test requires the subject to visualize the relationship between a pattern drawing and the object it represents. Each item consists of two drawings. The drawing on the left represents a three-dimensional object. On the right is a pattern drawing of the object. The edges of the object are numbered; the edges on the pattern are lettered. The task is to match each numbered edge with a lettered edge. Two edges in both drawings, labeled X and O are always given for reference purposes.

Number of Items: 30.

Time Limit: 15 minutes.

Scoring Formula: Two scores were obtained, total right and total wrong.

Reliability: No data available.

Sample Items:

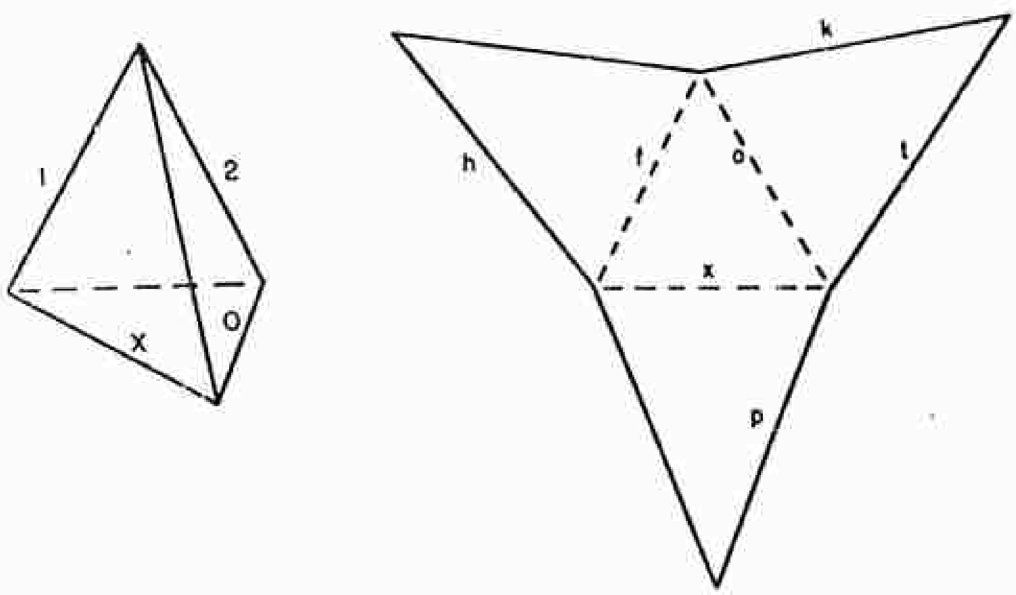


FIGURE A.27.—Pattern Comprehension Test, CP803A.

1 corresponds to (A) h (B) p (C) f (D) t (E) k

2 corresponds to (A) t (B) f (C) h (D) k (E) p

Test: Pattern Identification, CP820A.

Description: Each item consists of two groups of small circles. The group on the left consists of four circles which form a definite pattern by reason of their size and relationship to each other. In this group there is also a small cross. On the right, in a larger circular area, the pattern found on the left is repeated. There are, however, additional circles which make more complex the identification of the pattern. In addition, the basic pattern of four circles is at times placed to one side of the area so that a part or all of some of the circles in the pattern is not shown. Five lettered crosses, one of which corresponds to the cross in the pattern on the left are shown in the complex pattern on the right. The subject indicates his choice by choosing the correct lettered cross.

Number of Items:

Part I: 24.

Part II: 24.

Time Limit:

Part I: 9 minutes.

Part II: 7 minutes.

Scoring Formula: Two scores were obtained, total right and total wrong.

Reliability: r = 0.81. Correlation between part I and part II corrected for length. N = 168.[12]

Sample item:

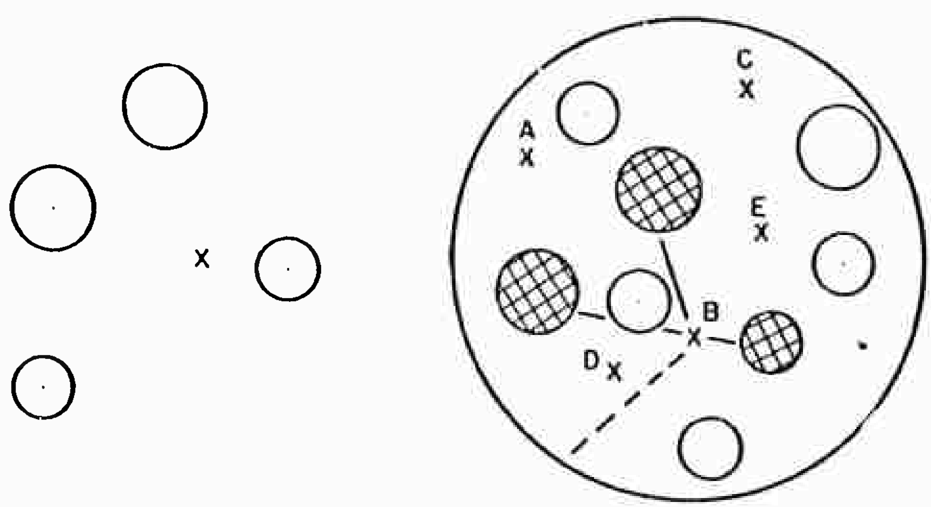


FIGURE A.28. Pattern Identification Test, CP820A.

For this sample the correct circles in the pattern have been filled in. The cross B is the correct answer for this problem.

Test: Polar Grid Coordinate, CP819B.

Description: The subject is given two numbers representing the bearing and range of a point from the center of a circle. He must plot the bearing of this point in the circle which is calibrated in degrees azimuth, and must plot range by concentric circles within it indicating units of distance from the center. After plotting the point, he must determine its coordinates on the X and Y scales of a square grid circumscribing the circle.

Number of items:

Part I, 26.

Part II, 24.

Time limit:

Part I, 8 minutes.

Part II, 7 minutes.

[12] Progress Report for Pathfinder Project of AERD No. 1, 1 Sept. 1944, p. 8.

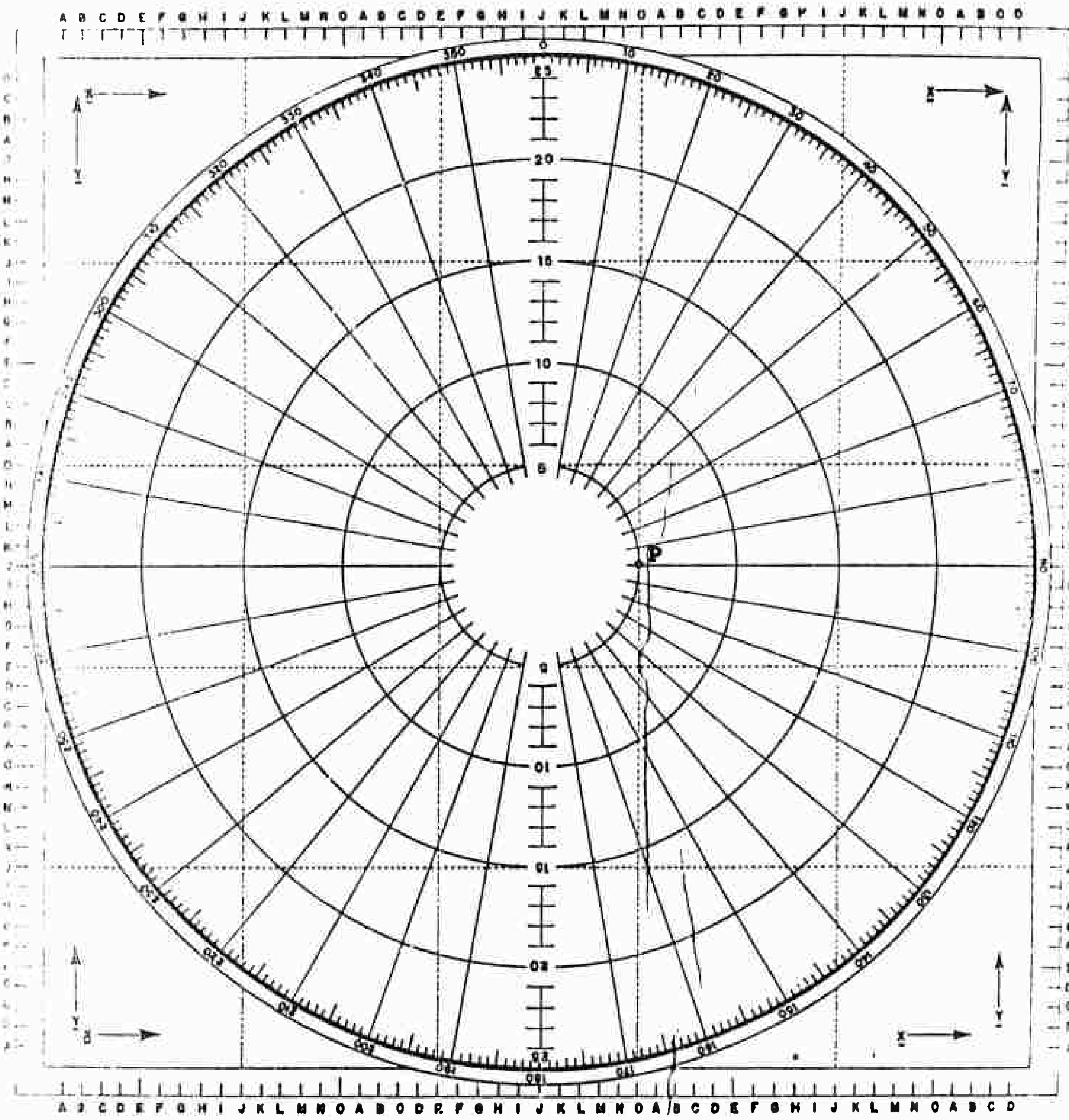


FIGURE A.29.—Polar Grid Coordinate Test, OP819D.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r = 0.90$. Correlation obtained by split-half technique corrected for length. $N = 100$.[13]

Sample item: Circle location numbers, 177–15. See Figure A.29.

Answer to Problem: K J.

Test: Position Orientation, CP526A.

Description: Parts I and II of this test consist of drawings of hands. The task is to determine whether each hand is a left hand or right hand. Parts III and IV contain drawings of hands, arms, feet, legs, and eyes. The task again is to determine whether each is the left or right member. A special answer sheet is required for this test.

Number of items:

- Part I, 26.
- Part II, 30.
- Part III, 26.
- Part IV, 30.

Time limit:

- Parts I and II, 7 minutes per part.
- Parts III and IV, 7½ minutes per part.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r = 0.83$. Correlation between part I and part II corrected for length. $N = 500$.

Sample Items:

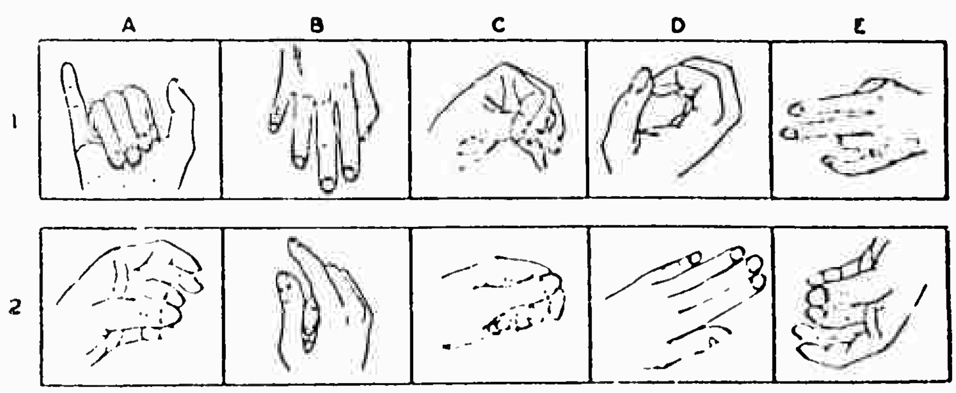


FIGURE A.30.—Position Orientation Test, CP526A.

Test: Ratio Estimation CP225A.

Description: The subject is presented with pairs of lines of different lengths. The task is to determine the proportion of the shorter line to the longer line.

---

[13] OSRD Report No. 1813, Formal Memorandum No. 4, Preliminary Report of Results from Oscilloscope Operator Tests.

Number of items:

Part I, 40.

Part II, 40.

Time limit:

Part I, 4 minutes 15 seconds.

Part II, 3 minutes 15 seconds.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r = 0.96$. Correlation obtained by split-half technique corrected for length. N = 100.[14]

Sample items:

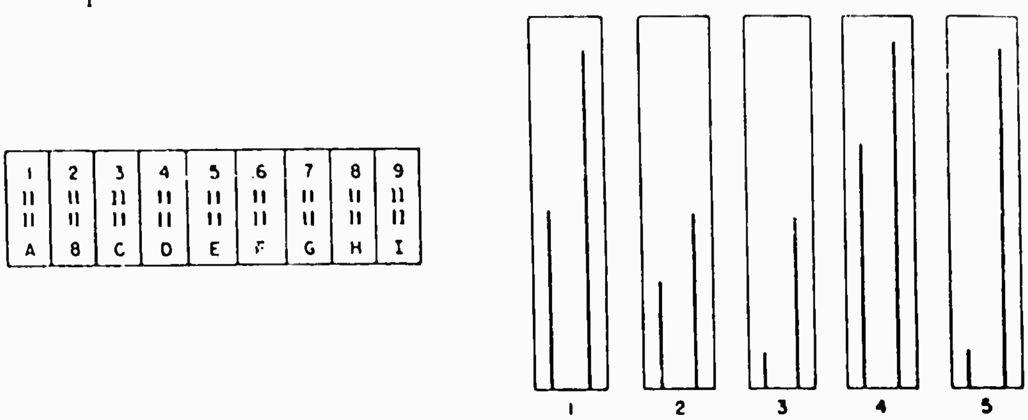


FIGURE A.31.—Ratio Estimation Test, CP225A.

Test: Scale Reading, CP637A.

Description: Different types of scales of varied complexity are presented. The subject must read one or more points on each scale.

Number of items, 70.

Time limit, 15 minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r = 0.81$. Correlation obtained by split-half technique for length. N = 100.[15]

Sample items:

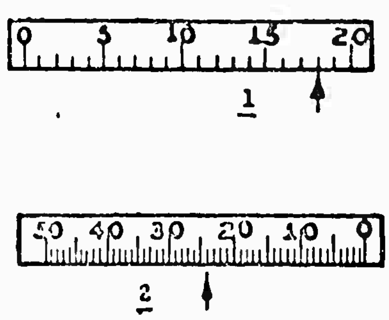


FIG. A.32.—Scale Reading Test, CP637. Alternatives: 1. (a) 15.5; (b) 18; (c) 22; (d) 17; (e) 20. 2. (a) 24; (b) 28; (c) 27; (d) 30; (e) 25.

---

[14] OSRD Report No. 183, Formal Memorandum No. 4, Preliminary Report of Results from Oscilloscope Operator Tests.

[15] *Ibid.*

Test: Spatial Reasoning, CI211BX2.

Description: For each item, a particular rule determines the position of symbols in a series of dashes and gaps. In some instances the rule governing the position of the symbols is based on their location in relation to gaps, others in relation to the dashes, and still others in relation to both gaps and dashes. The task is to discover and apply the rule for each item.

Number of items, 70.

Time limit, 25 minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r = 0.85$ for form CI211BX1. Correlation between part I and part II corrected for length. $N = 224$.[16]

Sample items:

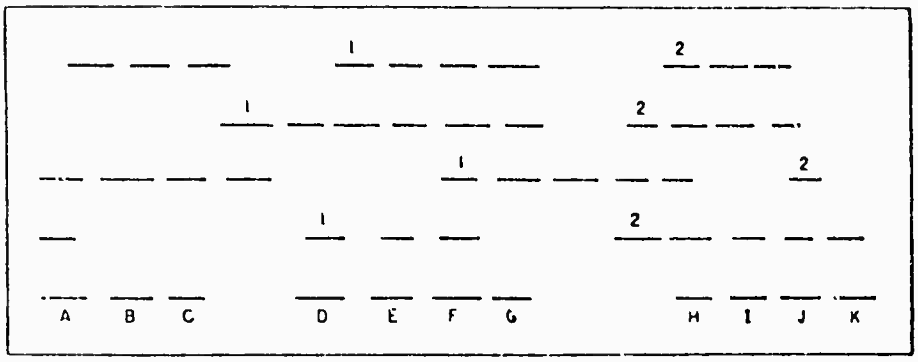


FIG. A.33.—Spatial Reasoning Test, CI211BX2.

In the above sample, notice that the numerals 1 and 2 are both placed just to the right of the gaps in the first four rows. The problem is to determine where the numerals would occur on the last row. In this case, the rule is: "Place the numerals just right of the gaps." The numeral 1 would thus be placed on the line above D and the numeral 2 would be placed on the line above H. D and H are therefore the correct answers for problems 1 and 2.

Test: Speed of Identification, CP610C.

Description: Items are in groups of four, each group consisting of four numbered designs and five lettered designs. Four of the lettered designs are the same as the four numbered ones. The task is to match each lettered design with the identical numbered design.

Number of items, 96.

Time limit, 5½ minutes.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r = 0.76$ for form CP610A. Correlation between separately timed halves. $N = 1{,}090$.[17]

---

[16] Guilford, J. P. and Lacey, J. I., *eds. op. cit.*, chapter 7.

[17] *Ibid.*, chapter 16.

Sample items:

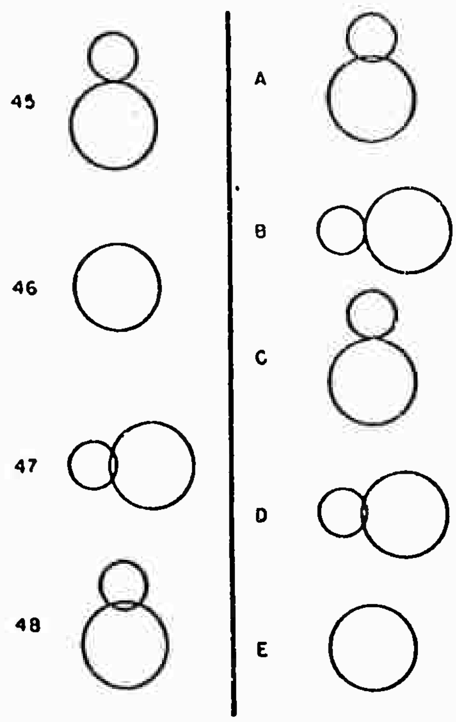


FIG. A. 34.—Speed of Identification Test, CP610C.

The correct answers are 1, C; 2, E; 3, D; 4, A.

Test: Spot Location, CP818A.

Description: The subject is presented with a large circle which is divided into lettered areas. Surrounding the large circle are smaller circles in which are numbered dots. The task is to determine in which lettered area of the large circle the dots in the smaller circles would fall if the smaller circles were the same size as the large circle.

Number of items:

Part I, 40.

Part II, 40.

Time limit:

Part I, 4 minutes 15 seconds.

Part II, 3 minutes 15 seconds.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r=0.98$. Correlation obtained by split-half technique corrected for length. N=100.[18]

[18] OSRD Report No. 183, Formal Memorandum No. 4, Preliminary Report of Results from Oscilloscope Operator Tests.

Sample items:

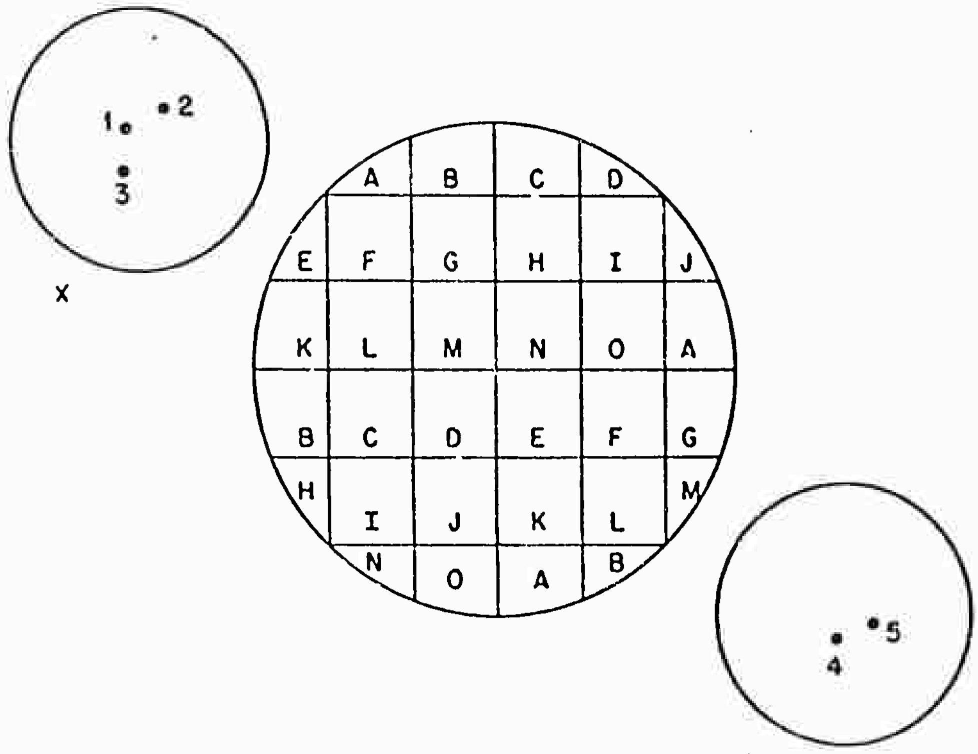


FIGURE A.35.—Spot Location Test, CP818A.

The correct answers are: 1, M; 2, N; 3, D; 4, D; 5, E.

Test: Visual Memory, CI514A.

Description: Each of the five parts of this test consists of a large aerial photograph which is studied for 1 minute. The page is then turned and 24 small aerial photographs are presented to the subject. The task is to determine which of the small photographs are represented in the large photograph and which are not.

Number of items: Five parts of 24 items per part.

Time limit: One minute study period per part; 2 minutes to answer 24 items per part.

Scoring formula: Two scores were obtained, total right and total wrong.

Reliability: $r=0.87$. Kuder-Richardson method. N=624.[19]

Sample item: Photograph to be studied.

[19] Guilford, J. P. and Lacey, J. I., *eds. op. cit.*, chapter 11.

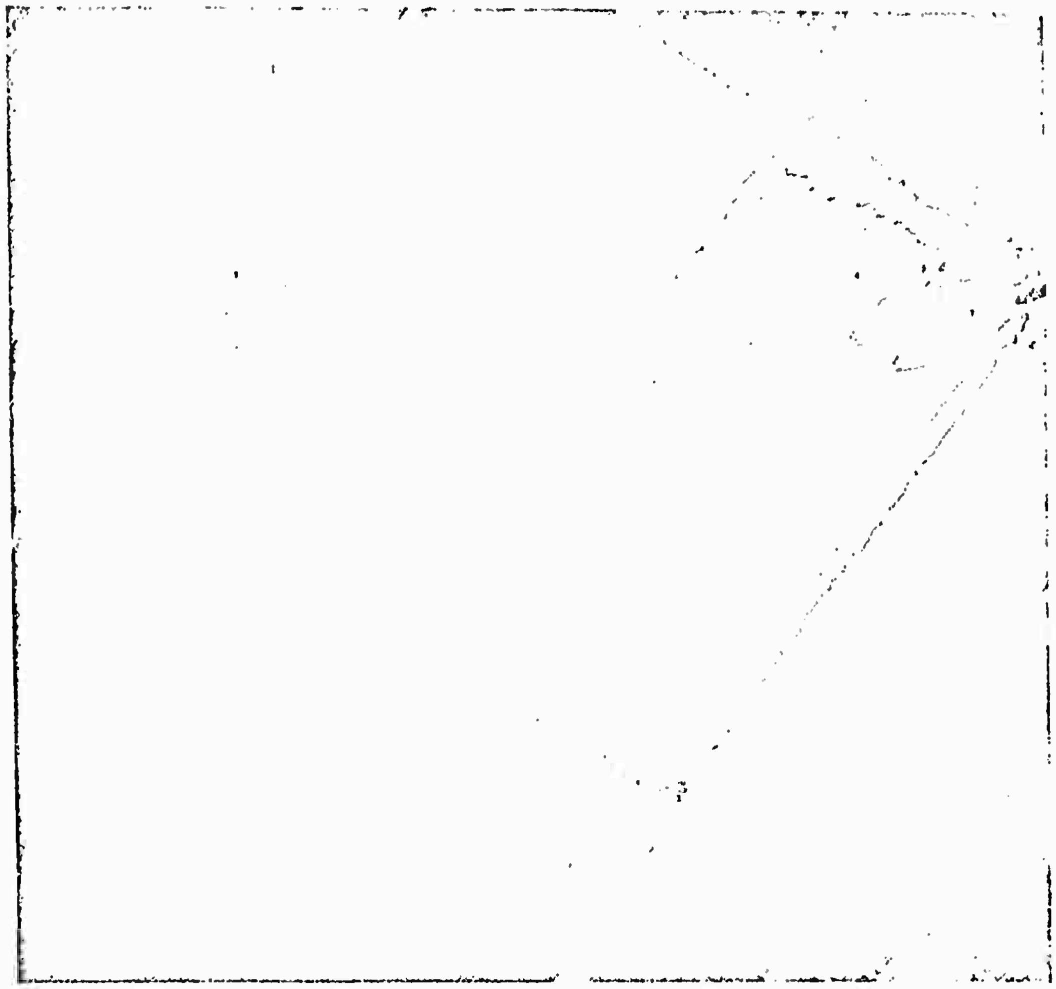


Figure A.36.—Visual Memory Test, C1514A (Study).

Now look at the sample items below. Is item 1 a section of the sample plate? Remember, you cannot refer back to the previous page.

Item 1 is a section of the sample plate, so blacken space A opposite number 1 on your answer sheet now. Item 2 shows a road, but it is not the road on the photograph just studied. So blacken space B opposite number 2 on your answer sheet.

Two sample problems:

Figure A.37.—Visual Memory Test C1514A (problem).

Test: Radar Student Information Blank.

Description: Radar observer students completed this information blank at the time the experimental battery was administered. The blank contains questions covering where the students had been administered the Radar Observer Selection Battery and the Air-Crew Classification Battery. The students also selected the one of four statements that best described the strength of their desire to take radar observer training.

Psychomotor Tests

Test: Check List Dial Setting (Model A) (no code number).

Description: This test consists of four dials, each calibrated in discrete steps from 1 to 11. The subject is given a list of settings for the four dials. He must set the dials according to the list and throw a toggle switch. If the settings are correct, a light flashes and he goes on to the next group of settings.

Number of items, 2 trials.

Time limit, 5 minutes per trial.

Scoring formula: Total number of correct settings.

Reliability: $r=0.79$. Odd-even correlation corrected for length. $N=381$.[20]

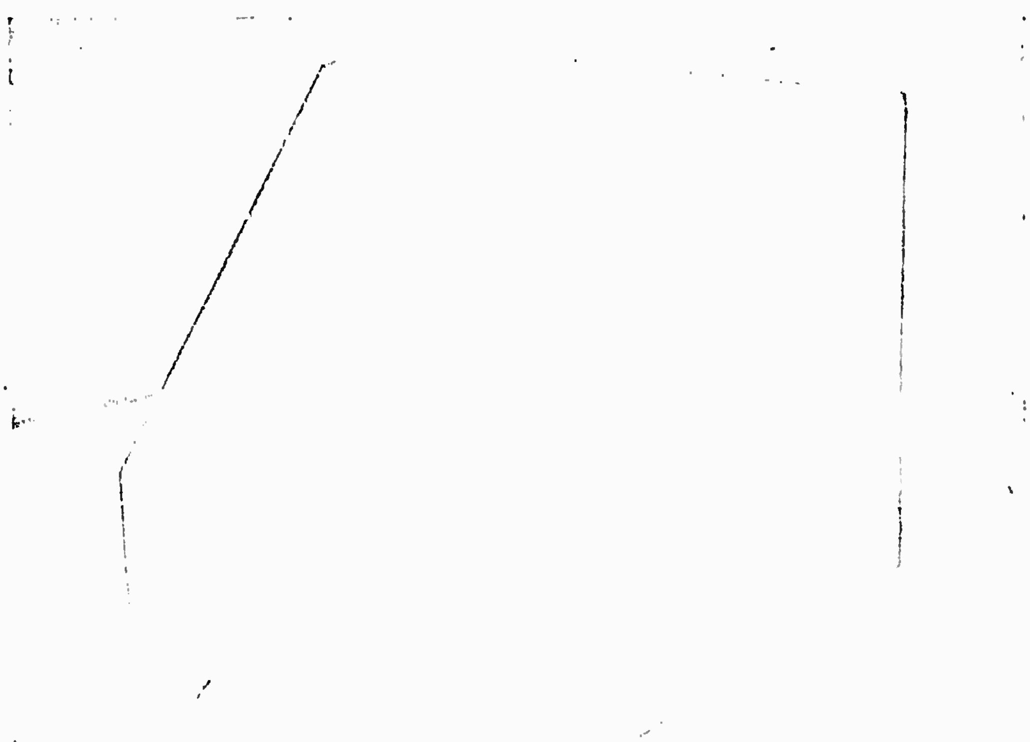


FIGURE A.38.—Check List Dial Setting Test (model A).

Test: Complex Coordination, CM701E.

Description: This is the same as the model used in the Air-Crew Classification Battery with two exceptions: First, the contacts for

[20] Melton, A. W., *ed. Apparatus tests.* AAF aviation psychology program research reports, no. 4. Washington: Government Printing Office, 1947.

the lights are much smaller. Thus, more exact movement of the control is necessary to keep the lights on. This calls for finer motor adjustments on the part of the subject. Second, the space between contacts is correspondingly larger. Thus, the time between one light going off and the next one coming on is greater.

Number of items: One test period.

Time limit, 8 minutes.

Scoring formula: Total number of correct settings.

Reliability: $r = 0.85$. Odd-even correlation corrected fo length. N=381.[21]

Apparatus:

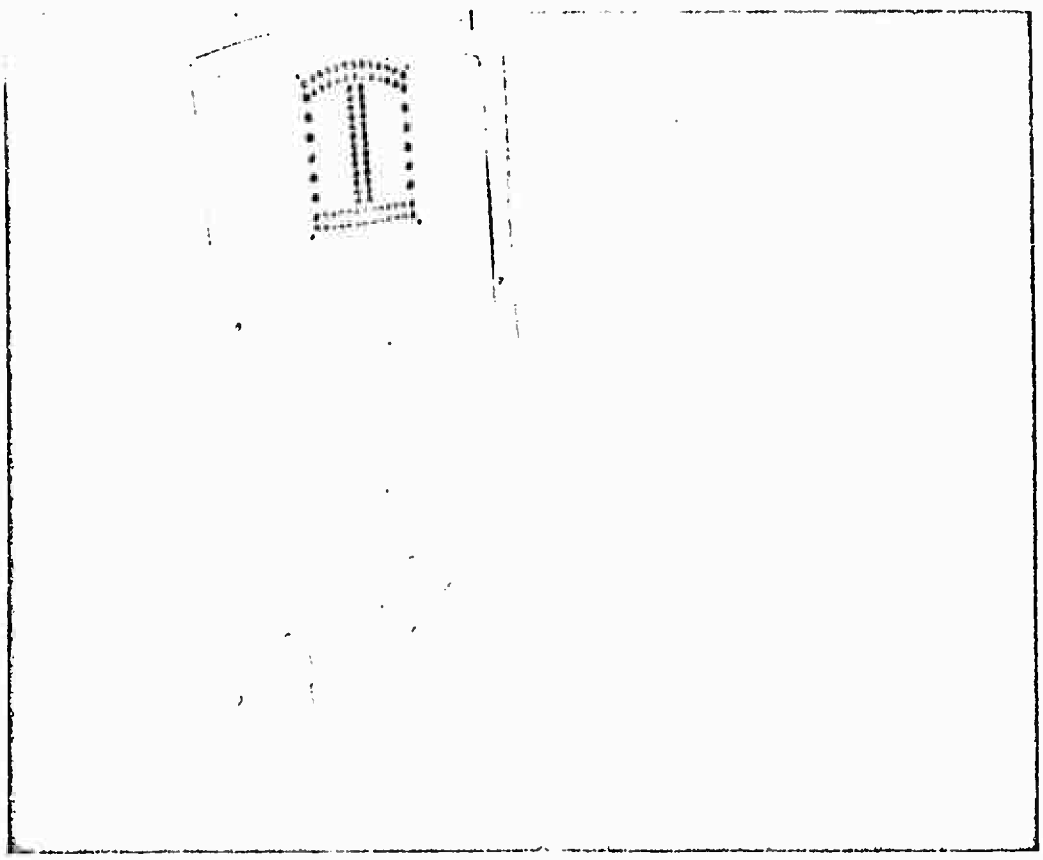


FIGURE A.9.—Complex coordination test, see M101A.

Test: Rate control, CM825A.

Description: The task is to keep a pointer on a black line which moves back and forth across a curved window at a varying rate of speed. The rate of movement of the pointer is controlled by turning a knob to the right to move the pointer to the right, and turning it to the left to move the pointer to the left. The further the knob is turned in either direction, the faster the pointer moves in that direction.

Time limit, 8-minute trial.

Scoring formula: Total time pointer is kept on black line.

[21] *Ibid.*

Reliability: $r=0.81$. Odd-even correlation corrected for length. N=381.[22]

Apparatus:

FIG. A.39.—Rate Control Test, CM825A.

Test: Self-Pacing Discrimination Reaction Time (CP611E modified).

Description: This test is much the same as the corresponding test in the Air-Crew Classification Battery with two main differences. First, instead of lights of different color, this test uses lights of different intensity. The relationship of the dimmer to the brighter lights determines which switch to throw. Second, instead of the pattern of the lights changing periodically and the score being the time required by the subject to throw the correct switch for each pattern, this test is self-paced and the pattern changes only when the subject throws the correct switch. When he makes an error, the lights go off and he must wait for the pattern to be repeated.

Time limit, 8-minute trial.

Scoring formula: Total number of correct responses.

Reliability: $r=0.90$. Odd-even correlation corrected for length. N=381.[23]

[22] *Ibid.*
[23] *Ibid.*

Apparatus:

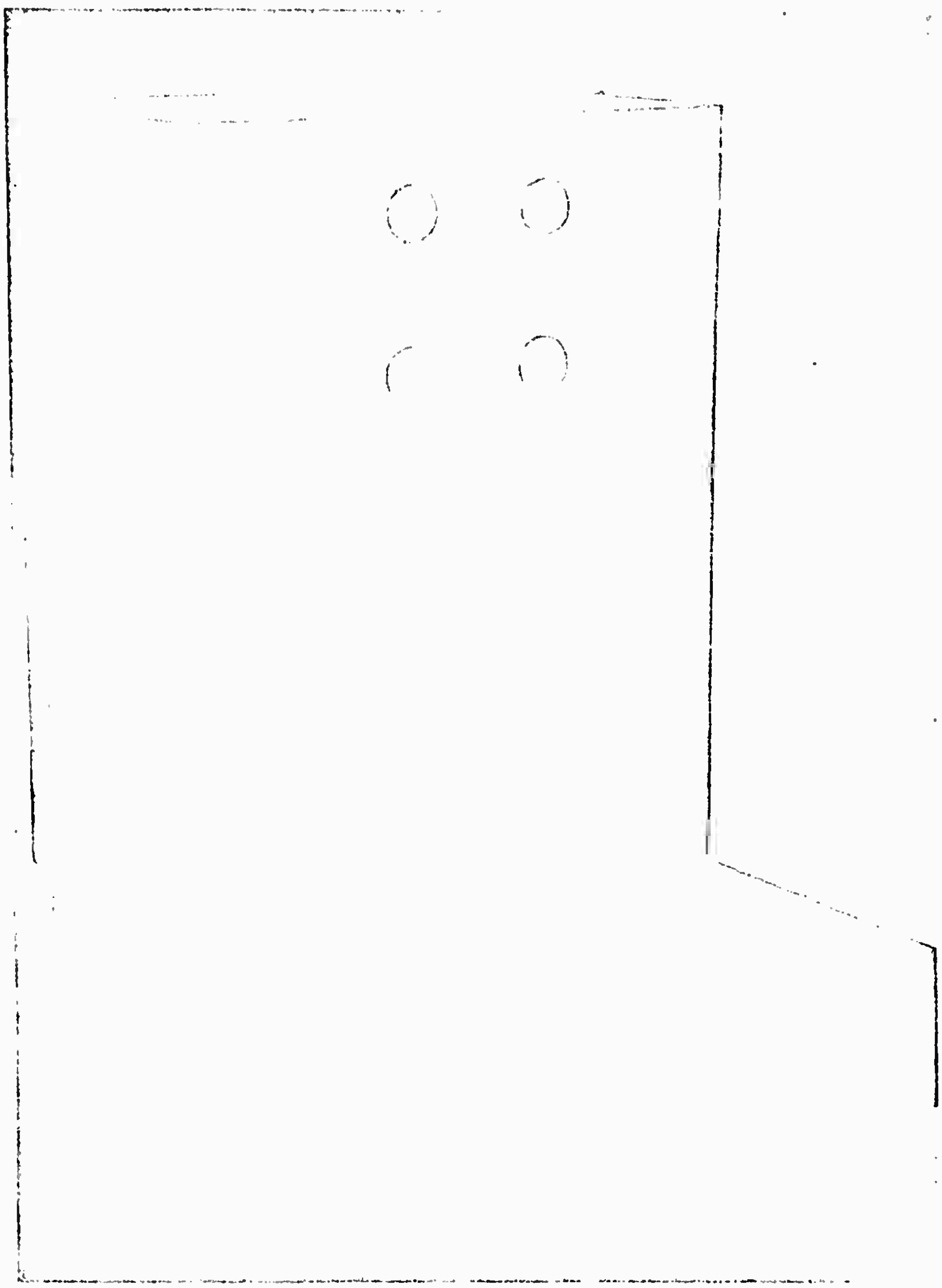


FIG. A.10.—Self-Pacing Discrimination Reaction Time Test, CP611E (modified).

Test: Two-Hand Pursuit (Thurstone), CM810A.

Description: This test is similiar to the Two-Hand Coordination Test in the Air-Crew Classification Battery. In this form, the lathe controls move a surface which is rotating eccentrically. The task is to keep a contact spot on the moving table under a stationary contact button.

Time limit, 8-minute trial.

Scoring formula: Total time that contact spot on moving table is kept under stationary contact button.

Reliability: $r=0.87$. Odd-even correlation corrected for length. $N=381$.[24]

Apparatus:

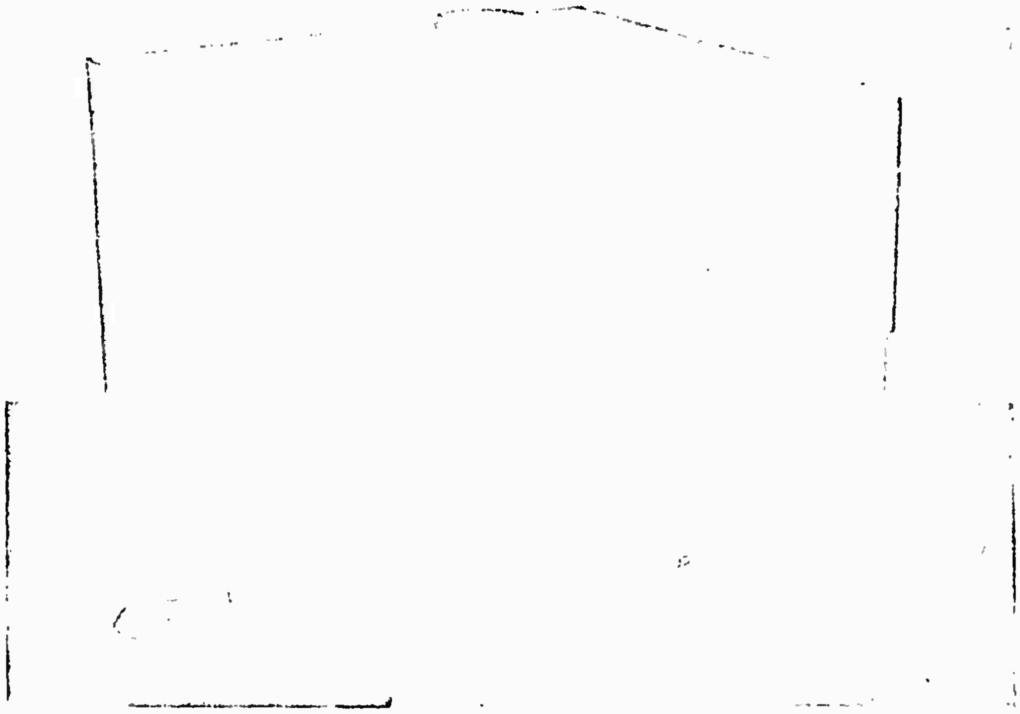


FIG. A.41.—Two-Hand Pursuit Test, CM810A (Thurstone).

Test: Visual Coincidence Reaction Test (CP613B3 modified).

Description: Two stationary lines of light on the same horizontal plane appear at a window. A moment later a moving line of light appears at the top of the window and moves down between the stationary lights. The task is to throw a toggle switch at the moment the moving light is in line with the stationary lines. In successive trials the stationary lines appear at different places along the window and the moving line travels at different rates.

Number of items, 156 trials.

Time limit, none.

Scoring formula: Total number of correct responses.

Reliability: $r=0.75$. Odd-even correlation corrected for length. $N=381$.[25]

---

[24] *Ibid.*
[25] *Ibid.*

Apparatus:

FIG. A.42.—Visual Coincidence Reaction Test, CP613133 (modified).

Glossary of Technical Terms

ABSOLUTE ALTITUDE.—Height of an aircraft above the terrain.

AFC.—Automatic frequency control. The circuit or its control which, once properly adjusted, maintains the radar receiver in tune with the transmitter.

AIMING POINT.—The landmark used as a reference for establishing the bombing run. It is usually near the center of the target area, except for off-set bombing.

AIR PLOT.—A navigation procedure for determining and recording air position.

AIR PLOT WIND.—The wind direction and velocity in terms of hourly units, obtained by comparing ground position and air position.

AIR POSITION.—Also called no wind position. A theoretical position of the aircraft computed from true air speed and heading only, assuming no wind effect. It coincides with ground position when there is no wind.

AIR SPEED.—The velocity of an aircraft with reference only to the air through which it is flying and without reference to the ground.

ALTITUDE DELAY.—The electronic time delay, or its control, which eliminates the altitude hole or blank area in the center of the scope, thus reducing distortion. The altitude hole results from the absence of reflecting objects between the aircraft and the earth.

AN/APQ, AN/APS.—Designation for air-borne radar equipment. The following are the sets referred to in the report.

AN/APQ-5.—Essentially an electronic bombsight, used in conjunction with a radar search set.

AN/APQ-7.—A radar search set known as the Eagle, characterized by unusually high definition; particularly successful as an aid to bombing. It does not have a 360° sweep, the area presented on the scope being limited to 30° to the left and 30° to the right of the heading of the aircraft. The antenna is fixed. The set is without azimuth stabilization.

AN/APQ-13.—A radar search set used as an aid to both bombing and navigation. It has a 360° sweep and is equipped with azimuth stabilization. Antenna tilt control allows for improved scope presentation. This set was used almost exclusively in the Pacific theater.

AN/APQ-23.—The most improved radar search set in production at the close of the war, characterized by simplified and automatic controls and computing devices, including the electronic solution of offset bombing problems.

AN/APS-15.—A radar search set similar to the AN/APQ-13, but equipped with an A-scope to facilitate tuning and calibration. It has fewer separate units than the AN/APQ-13 and therefore defective parts are not as easily repaired or replaced. This set was used almost exclusively in the European theater.

AN/APS-15A.—An improved model of the AN/APS-15 set, having a different computer and different procedures for adjusting range marks and for calibrating altitude and range.

A-Scope.—A cathode ray tube on which range data are presented as vertical pips on a horizontal scale of distance. It is used to facilitate tuning and calibration on the AN/APQ-7, the AN/APS-15, and AN/APS-15A sets. On the AN/APQ-7 it is also used to facilitate setting in absolute altitude.

Automatic Pilot.—Also called C-1, because of the popularity of that particular model, and AFCE, automatic flight control equipment. A gyro-stabilized electrical-mechanical device, for maintaining the aircraft at a desired altitude. Used extensively in all theaters and in training, particularly during bombing runs and formation flying.

Azimuth.—Angular distance measured in degrees clockwise from true north.

Azimuth Stabilization.—An electronic device incorporated in the radar set, maintaining north at the top of the scope to facilitate scope-map interpretation. It is operated from the flux-gate compass.

BEACON.—Also radar beacon or Racon. A fixed ground radar signal generator which, upon being activated by certain radar signals, transmits a coded signal that can be identified on the airborne scope and plotted as any other return. Used particularly for homing.

BLIP.—See Pip.

BOMB RUN.—The final approach to the target during which final corrections are made in course and in computing the release point.

BOMBSIGHT.—Usually refers to the Norden sight. An optical computing instrument used in visual bombing to determine and direct the aircraft to the bomb release point. Its solution of the bombing problem involves the handling of such factors as true air speed, absolute altitude, and drift.

BTO.—Bombing through overcast. An Army designation for airborne radar operation. See RO (B).

CALIBRATION.—The systematic adjustment of the receiver-indicator, the range unit, and the computer so that accurate distances are indicated for altitude and slant range. The set is calibrated on the ground by a mechanic and checked in the air by the radar observer.

CAMERA BOMBING.—A simulated bombing run in which, instead of bombs being released, the accuracy of release is measured photographically. At least two pictures are taken of the ground, one at "Bombs away" and one at the theoretical time of impact. Accuracy of measurement varies up to roughly 300 feet.

CATHODE RAY TUBE (CRT).—Also called the scope. A vacuum tube in which an electron beam is made visible by being focused upon a fluorescent screen on the flat end of the tube where it is converted into light energy. See also **Radar.**

CHECK POINT.—A landmark identifiable either visually or on the radar scope.

CIRCULAR ERROR (CE).—The distance between the point of impact of a bomb and the center of the target.

COORDINATED BOMBING.—A bombing procedure in which the radar observer furnishes data, particularly speed of closure in the form of sighting angles, to the bombardier who can then set up and operate his sight and release the bombs without seeing the target.

DEAD RECKONING NAVIGATION (DR).—Inferring the position of the aircraft by applying to the last known position the estimated track made good, computed from a previously determined wind, heading, and true air speed.

DEFLECTION ERROR.—The distance the bomb falls right or left of a line extended through the center of the target parallel to the aircraft's track.

Direct Bombing.—The bombing procedure in which the radar observer directs the run and releases the bombs independent of the bombardier. Considered to be less accurate than coordinated bombing.

Dividers.—The familiar compass-like instrument used to measure and plot distances on maps and charts.

Drift.—The angular difference between the aircraft's heading, or direction in which it is pointed, and its track, or the path it makes over the earth's surface. Drift is zero when the aircraft is headed directly into or away from the wind, and is maximum when the wind direction is perpendicular to the aircraft's heading.

Driftmeter.—Also called the B-3. A simple optical device for reading drift.

ETA.—Estimated time of arrival.

E-6B Computer.—A combination circular slide rule and transparent-slide vector plotter. It is used to make conversions, as from statute miles to nautical miles, and to compute such variables as air speed, ground speed, wind velocity, wind force, and ETA.

Final Point.—In bombing, the last sighting angle given by the radar operator to the bombardier who is synchronizing the bombsight. After the final point, the radar observer may set up the set for direct bombing, in the event that the bombardier is unable to make the release.

Fix.—The location of an aircraft from terrain features appearing both on the scope and on a map. A fix may consist of simultaneous bearings on two or more features, or the bearing and range of a single feature.

Gain.—See Receiver Gain and Video Gain.

Ground Position.—The point on the ground over which the aircraft is at a particular moment. May be expressed as coordinates of latitude and longitude.

Ground Range.—Distance from the ground directly under the aircraft to the object.

Ground Speed.—The velocity of the aircraft with reference to the earth's surface.

HAB.—High-altitude bombing. See RO (B).

Heading.—The direction in which the aircraft is pointed. True heading, in respect to true north, is obtained from a compass and deviation.

H2S.—See RO (B).

H2X.—See RO (B).

Initial Point.—The point on the ground over which the bombing run is started.

Knot (K).—Nautical miles per hour. One nautical mile equals 6,080 feet.

LAB.—Low-altitude bombing.

LOG.—In navigation, the systematic chronological record of an aircraft's flight.

LORAN.—Long-range aid to navigation. Two ground stations transmit synchronized radio signals. The air-borne Loran set measures the time difference in receiving signals. Fixes are taken by plotting equipment readings on specially prepared charts.

LUBBERLINE.—An illuminated radius of the scope indicating the headings of the aircraft. It points to the top of the scope if azimuth stabilization is off and to true heading if azimuth stabilization is on.

MAIN SCOPE.—See PPI. The PPI scope located on the radar set.

METRO WIND.—The forecasted wind as reported in briefing and used in the preflight planning of the mission.

MICKEY.—See RO (B).

MULTIPLE DRIFT PROCEDURE.—In radar bombing, a systematic method for estimating the correction in heading to compensate for drift on the bombing run.

OFFSET BOMBING.—A bombing procedure in which the release point is computed with reference to an aiming point outside of the target area in order to bomb a target which is not visible or is poorly visible.

PILOTAGE.—Locating and navigating the aircraft by constant reference to the ground, either directly or as represented in the scope. Contrasted to DR navigation, in which the terrain is not constantly observed.

PIP.—Also called blip. The presentation on the scope of a relative increase in current. On the PPI scope, it is a point of increased illumination on the sweep due to a target, a range mark, or the bomb release mark. On the A-scope, it is a sharp vertical peak.

PRECISION TURN.—Also procedure turn. The controlled turn of an aircraft or of a formation at the predetermined rate, usually 45 degrees or more per minute.

PPI.—Plan position indicator. A radar cathode ray tube on the screen on which the terrain under the aircraft is presented by means of a rotating sweep of varying brightness.

PULSE.—A momentarily increased current or voltage. Radar pulses are timed electronically so that distances are known from the time differences between the transmitted and received pulses.

RADAR.—Radio detection and ranging. Radar includes all electronic pulse-echo equipment. An air-borne radar search set is the basic unit of equipment used for navigation and bombing. It transmits a narrow beam of high frequency radio pulses and receives the same pulses, in weakened form, as echoes reflected back from the earth's surface. The echoes are received as an electron beam or sweep which is made visible by being focused upon a fluorescent screen on the flat end of a cathode ray tube where it is converted into light energy. Different terrain features reflect different amounts of energy. Water reflects almost no energy, flat country reflects little, while cities reflect large amounts of energy. The electron beam constantly sweeps the screen, in a circular motion for search sets, of which the AN/APS-15 is an example. The result is an illuminated picture on the screen which is, in effect, a circular map of the terrain under the aircraft. The point on the ground directly beneath the aircraft is the center of the map. Navigation and bombing data may be ascertained accurately with the use of auxiliary circuits and devices. An area within a maximum radius of 100 or more miles is represented, regardless of darkness or most weather conditions such as undercast.

RADAR BEACON.—See Beacon.

RADAR OBSERVER (BOMBARDMENT).—See RO(B).

RANGE.—Distance, measured in radar by the time required for a pulse to leave the transmitter, be reflected from a target, and return as an echo or received pulse.

RANGE MARKS.—Calibrated pips on the PPI sweep which present equidistant concentric circles for measuring the distance from aircraft to target. A manual control places range marks on the scope at either 1 or 5 nautical miles apart.

RECEIVER GAIN.—The manual control regulating the sensitivity of the set affecting the intensity of the returns only.

RO(B).—Radar Observer (Bombardment). The Army designation for the occupational specialty 0142: commissioned air-crew members subsequently trained as radar observers. RO(B) is also used to refer to radar observer equipment, training, and operations. RO(B), a more recent term, is loosely synonymous with BTO, H2X, H2S, HAS, Mickey, and Stinky, earlier terms which arose largely because all aspects of radar, including the term radar, were classified as secret.

SCOPE.—See Cathode Ray Tube (CRT), and Radar.

SCR-584.—An electronic device for measuring, from the ground, the accuracy of a simulated bomb release.

SCR-717-A.—An air-borne radar search set, now obsolete. Target returns are presented on a "B" scope, a vertical sweep on a rectangular screen. Resolution is poor, there is no azimuth stabilization, and no computer to allow for bombing. The set was used in sea search and as an aid to navigation.

SCR-717-B.—An improved model of the SCR-717-A, also obsolete. Target returns are presented on a PPI scope. Resolution is poor, there is no azimuth stabilization and no computer for bombing.

SCR–718.—A very accurate electronic absolute altimeter operating on the radar pulse-echo principle. Operating range is from 25 to 40,000 feet.

SLANT RANGE.—The shortest straight-line distance from the aircraft to the object. For distances exceeding about 18 miles, slant range is usually converted to ground range. For longer distances, slant range and ground range are considered equal.

SPINNER.—The radar antenna, so called because it constantly rotates. The rate is about 26 revolutions per minute. The antennae of radar search sets mentioned in this volume are of the spinner type except the AN/APQ–7 which has a fixed wind-shaped antenna.

STANINE.—In the AAF psychology program, a composite aptitude score ranging from 1, low, to 9, high. It is a combination of the words standard and nine. Statistically, it is a function of the standard deviation of combined differentially weighted raw scores on a battery of selection tests. Each stanine equals one-half of the sigma of the distribution. The mean and the median are thus a stanine of 5, and the sigma of the distribution of stanine scores is 2 stanines. The percentage distribution of stanines is as follows: 1, 4 percent; 2, 7 percent; 3, 12 percent; 4, 17 percent; 5, 20 percent; 6, 17 percent; 7, 12 percent; 8, 7 percent; 9, 4 percent.

STINKEY.—See RO(B).

SWEEP.—The electronic beam or its motion as it moves across the screen of the cathode ray tube. On the PPI scope it appears as an illuminated radius rotating with the spinner leaving pips on the screen indicating terrain features.

SWEEP DELAY.—A device or its control for extending the area of terrain represented on the scope. For example, with 20 miles of sweep delay set in, the area represented at the center of the scope is 20 miles from the aircraft.

TARGET.—The point or area to be bombed. Loosely, any aspect of the terrain giving a return on the scope.

TARGET TIMING.—A procedure for estimating average track and ground speed, using a stop watch and the E–6B computer. Successive fixes on a target, whether or not it is identified, are plotted on the transparent slide of the E–6B computer to give average track. The distance and time interval between the first fix and the last fix give ground speed. The procedure has the disadvantage of requiring the target to remain on the scope for about nine minutes and some targets move off the scope before this time. The procedure has the advantages of representing an average, and of allowing track to be estimated without an exact knowledge of terrain.

TILT.—The control which adjusts the facing of the antenna to the ground, bringing up desired features of the terrain.

TRACKING.—Adjusting heading so that the aircraft passes directly over a target. Also refers to following a target return as it moves across the scope face.

TRUE HEADING.—See Heading.

TURNING POINT.—A radar return or pair of coordinates chosen for the aircraft or formation to turn upon. It may be at any point on the mission, for example, a rendezvous or the last turn before the initial point (IP).

VIDEO GAIN.—The manual control regulating the intensity of all tracings on the scope, including ground returns but not affecting sensitivity. Controls the general level of illumination.

Index