UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A185 760

DTIC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/CI/NR 87-86T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Relationship Of The Basic Attributes Test To Tactical Reconnaissance Pilot Performance		5. TYPE OF REPORT & PERIOD COVERED THESIS/DISSERTATION
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Timothy G. Kinney		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS AFIT STUDENT AT: St Mary's University		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFIT/NR WPAFB OH 45433-6583		12. REPORT DATE 1987
		13. NUMBER OF PAGES 78
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC ELECTE NOV 0 4 1987 S D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

D

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE: IAW AFR 190-1

LYNN E. WOLAVER 23 Sep 87
Dean for Research and Professional Development
AFIT/NR

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ATTACHED

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A

87 10 20 164

ABSTRACT

Timothy G. Kinney, Relationship of the basic attributes test to tactical reconnaissance pilot performance, Major, United States Air Force, 1987, 82 pages, Master of Science, St. Mary's University.

This research study was conducted to contribute to the United States Air Force Pilot Selection and Classification Research and Development Program. Specifically, this was a pilot project with the purpose of providing concurrent validity information on the ability of the Basic Attributes Test (BAT) to measure the unique abilities of successful tactical reconnaissance pilots. This was accomplished by proposing criterion measures of successful reconnaissance pilot performance and comparing them to pilot BAT scores. The BAT consists of a newly developed group of tests designed to assess psychomotor skills, and psychological and cognitive attributes believed to be associated with successful pilot performance. A second purpose of the study was to explore the use of supervisor rankings as a possible future criterion measure of pilot performance. This was achieved by comparing reconnaissance pilot performance measures to the rank-order assigned to a pilot by his supervisors. The results of the study indicated that three BAT cognitive tests (Decision Making Speed, Item Recognition, and Mental Rotation) were significantly related to pilot performance. These results suggest that further studies should be conducted to determine if the BAT could be used to improve the selection of reconnaissance pilots The results also indicated that supervisory rank-ordering was not related to performance. These findings imply that personnel rating systems should be used with caution when used as a criterion of pilot performance. Keywords: personnel selection; performance (human), skills, theses

86

RELATIONSHIP OF THE BASIC ATTRIBUTES TEST TO TACTICAL RECONNAISSANCE PILOT PERFORMANCE

A

THESIS

Presented to the Faculty of the Graduate School of

St Mary's University in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

in

Psychology

By

Timothy G. Kinney, BA, MA

San Antonio, Texas

May, 1987

RELATIONSHIP OF THE BASIC ATTRIBUTES TEST TO TACTICAL RECONNAISSANCE PILOT PERFORMANCE

Timothy G. Kinney

St. Mary's University, 1987

Supervising Professor: Jeffrey E. Kantor, Ph.D

This research study was conducted to contribute to the United States Air Force Pilot Selection and Classification Research and Development Program. Specifically, this was a pilot project with the purpose of providing concurrent validity information on the ability of the Basic Attributes Test (BAT) to measure the unique abilities of successful tactical reconnaissance pilots. This was accomplished by proposing criterion measures of successful reconnaissance pilot performance and comparing them to BAT scores. The BAT consists of a newly developed group of tests designed to assess psychomotor skills, and psychological and cognitive attributes believed to be associated with successful pilot performance. A second purpose of the study was to explore the use of supervisor rankings as a possible future criterion measure of pilot performance. This was achieved by comparing reconnaissance pilot performance measures to the rank-order assigned to a pilot by his supervisors. The results of the study indicated that three BAT cognitive tests (Decision Making Speed, Item Recognition, and Mental Rotation) were significantly related to pilot performance. These results suggest that further studies should be conducted to determine if the BAT could be used to improve the selection of reconnaissance pilots. The results also indicated that supervisory rank-ordering was not related to

performance. These findings imply that personnel rating systems should be used with caution when used as a criterion of pilot performance.

PREFACE

This research was supported by the Department of the Air Force Request for Personnel Research (RPR) 78-11, Aircrew Selection and Classification.

Many individuals at the Air Force Human Resources Laboratory aided in the completion of this project. I particularly wish to thank Jeffrey E. Kantor, Ph.D and Major John C. Quebe for their assistance, encouragement, and creative suggestions which aided in the conception and completion of this project. In addition, thanks are extended to Edison F. Watkins who performed the computer analysis of Basic Attributes Test results.

Sincere appreciation is also owed to the committe members for their technical assistance, patience, and time.

Presented to the committee on March 26, 1987.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 1

INTRODUCTION

The Aircrew Selection and Classification Research Function of the Air Force Human Resources Laboratory (AFHRL) has been tasked with refining the Air Force's pilot selection process by identifying those attributes and abilities which an individual must possess to become a successful military pilot. The pilot selection and classication research and development (R&D) program was intiated by the AFHRL in response to an Air Force Air Training Command (ATC) request to capitalize on state-of-the-art technologies to improve the way the Air Force selects and classifies people for pilot training.

To date, most research by the AFHRL has been conducted on the selection of pilot candidates for initial entry into Undergraduate Pilot Training (UPT). Particular interest has been in exploring the potential of a recently developed battery of psychomotor and information processing tests named the Basic Attributes Tests (BAT). These studies have concentrated on the use of the BAT in predicting an individual's ability to become a pilot as measured by his completion of, or failure to complete pilot training (Kantor, & Bordelon, 1985). However, the scope of the research has recently been expanded as the classification of pilots into one of two mission specialties upon completion of UPT has been investigated (Kantor, Carretta, & Quebe, 1986).

This study was intended to contribute to the classification phase of the USAF Pilot Selection and Classification R&D Program. Specifically, this was a pilot project with the purpose of providing concurrent validity information on the capability of the BAT to measure the unique abilities of successful tactical reconnaissance pilots. This was accomplished by proposing criterion measures of successful reconnaissance pilot performance and comparing them to pilot BAT scores. The goal of such a validation study was to provide ATC and Air Force Military Personnel Center (AFMPC) with an additional objective measure to utilize in placing pilots in tactical reconnaissance units following completion of UPT. A refined pilot classification technique could be beneficial by reducing upgrade training and retraining costs, and by reducing training time, especially continuation training time which requires the use of critical operational squadron instructor pilot resources. A second purpose of the study was to explore the use of supervisor rankings as a possible future criterion measure of pilot performance.

The History of Pilot Selection

Research into pilot selection and performance measurement is by no means new. Cursory attempts to select the individual most likely to succeed in flight training and then succeed as an operational pilot began as far back as World War I. However, it was the events of World War II which spawned the need to efficiently select individuals most likely to succeed in pilot training (Super & Crites, 1962). As a result the first scientific tests for pilot selection were developed. Initially developed by the Civilian Pilot Training Program of the Civil Aeronautics Administration, these tests were expanded upon by the U. S. Navy and the Army Air Force's Aviation Psychology Program.

The Army Air Force's test, the Aviation Cadet Classification Battery, developed under Flanagan initially in 1942 contained the largest variety of tests and was subjected to the most extensive validation procedures (Super & Crites, 1962). These tests were developed from job analysis procedures and attempted to measure unitary traits and to duplicate the job situation. The objective of the tests was to measure aptitudes and abilities which were thought to relate to pilot success. Also interesting was the fact that as the variety of tests increased the number of factors which were believed to be related to pilot success also increased. Shartle named 11 in 1945 and in 1947 Guilford named 28 abilities. Some of the abilities which Guilford identified included spatial relations, visualization, perceptual speed, paired associate memory, visual memory, picture-word memory, several forms of reasoning and integration, psychomotor speed, coordination and precision, and judgment (Super & Crites, 1962).

Based on these identified abilities of successful pilot performance a final battery of tests were developed. The subtests included in the battery were; general information, instrument comprehension, mechanical principles, dial and table reading, biographical data, stanine (battery score), aviation cadet qualifying, army general classification, education, flying adaptability rating, discrimination-reaction time, complex coordination (stick-rudder), and two-hand coordination (Super & Crites, 1962).

A most notable result of these tests was the validation studies which were conducted on them. The validation study was conducted with a "group of 1143 candidates for aviation cadet training who were sent to pilot training regardless of their scores on psychological tests" (Super & Crites, 1962, p. 366). As a result, the validity coefficients provided probably the best estimates of the validity of the tests used since the sample was minimally affected by attenuation in range, a problem which has affected most follow-on studies. The results of analysis of these tests revealed that; a) a behavioral sample was present in most of the tests, b) the three most valuable tests were paper-and-pencil tests, c) the most valid tests were custom-built, d) the battery of tests had more predictive value than any one test, e) and, objective tests had more predictive value than psychiatric judgment (Super & Crites, 1962).

This test battery was utilized in the selection of individuals for pilot training until 1955. When first implemented the test battery was administered and maintained in one centralized location. In the early 1950's the test battery, including the apparatus-based testing units, was decentralized so that testing could be accomplished more efficiently and economically. This decentralization resulted in difficulty in maintaining the calibration of the electro-mechanical apparatus testing devices. As a result, the consistency of the apparatus was reduced which in turn acted to lower the reliability of the tests (Carretta, in press). Consequently, even though this battery of tests had good promise in selecting successful pilot candidates it was dropped in 1955 and replaced with a new battery of paper-and-pencil tests.

While still in use today, these paper-and-pencil tests have not adequately improved the pilot selection process. Attrition rates have typically ranged between 22-25% (Kantor et al., 1986). Given rising costs of training this loss rate has become increasingly unacceptable. This consequent need to improve the pilot selection process along with recent advances in computer technology, especially in the area of the table top personal computer, has renewed interest in the use of apparatus-based testing.

To determine the nature of the tests to develop, a review of recents studies was conducted to identify abilities which may predict successful piloting skills and performance. Fleishman and Ornstein (1971) suggested that the individual factors of control precision, spatial orientation, multilimb coordination, response orientation, rate control, and kinetic discrimination provide a means of discriminating among individuals in complex tasks. Gerathewohl suggested a list of psychological and psychophysiological variables that were important in determining pilot performance. The list included; perception, attention, reaction, orientation, sensorimotor, stamina, cognition/mentation, interpersonal relations, decision-making, experience, learning, personality, mechanical ability, and motivation (Gerathewohl, 1978). Most recently, the Advanced Research Resources Organization, under contract with the AFHRL, conducted a review of available cognitive tests aimed at identifying unique attributes contributing to successful pilot performance (Imhoff & Levine, 1981). The interesting analogy between the new and old lists of pilot abilities was that while the terminology had changed somewhat the factors which were expectedto be related to successful pilot performance remained basically the same. At least to some degree, then, the suspected predictive relationship between these abilities and successful pilot performance has withstood the test of time.

The Basic Attributes Test

As a result of the Imhoff and Levine review, "a set of experimental tests were developed to assess an individual's ability to handle large amounts of information, share attention across simultaneous tasks, and make decisions in a high information-content environment" (Kantor & Bordelon, 1985, p.259). These resultant tests compose the BAT battery. The BAT subtests seem to adequately assesses all of those factors identified by Fleishman and Ornstein, most of those psychophysiological factors identified by Gerathewohl, plus additional cognitive abilities identified by the Advanced Research Resources Organization. In addition, several of the tests are similar and based upon the same concepts as the original Aviation Cadet Classification Battery. Of this original set of tests, two tests, complex coordination and two-hand coordination, have demonstrated the most promise in predicting an individual's success in completing pilot training.

The BAT psychomotor tests when combined with other existing selection criteria, including the Air Force Officer Qualification Test (AFOQT) score, and Flight Screening Program (FSP) grades have been shown to have promise in reducing the student pilot attrition rate (Kantor & Bordelon, 1985). These predictors have been integrated into an approach named the Pilot Candidate Selection Method (PCSM) which is currently in Operational Test and Evaluation (OT&E). More recently, three BAT cognitive subtests have been shown to have promise in discriminating among pilots who received Fighter-Attack-Reconnaissance (FAR) ratings versus those who receive Tanker-Transport-Bomber (TTB) ratings at the completion of UPT (Kantor et al., 1986).

The main purpose of the present study was to further explore the predictive capabilities of the BAT. That is, it was intended to determine if BAT scores were also related to successful performance as an operational reconnaissance pilot. To accomplish this, relevant measures of operational pilot performance had to be identified.

Measures of Reconnaissance Pilot Performance

While initial interest was directed at pilot selection, the major emphasis of research during the1950's and into the 1970's was directed toward the establishment of a criterion measure of successful pilot performance. When attempting to determine a criterion meaure one must be aware that as with most social science problems an exact, objective measure of actual or true pilot performance may not yet exist. As deLeon pointed out "the Air Force has no consistent and objective set of guidelines with which to evaluate the proficiency of its combat pilots" (deLeon, 1977, p.24). In addition, Smith and Flexman (1972) indicated that:

> subjective techniques for measuring pilot performance have been used for many years in pilot training and have been demonstrated to be satisfactory for routine instructional usage. However, such methods are usually inadequate for use in pilot training research (p. 70).

Consequently, a review of past studies was conducted to identify an adequate criterion from which to measure pilot performance.

Research has primarily been directed toward measuring performance on specific tasks performed during air-to-air combat and air-to-ground bombing missions. A review of the literature, including United States Air Force (USAF), United States Navy (USN), and Federal Aviation Association (FAA) research projects revealed no previous attempts to identify a criterion of successful reconnaissance pilot performance. As a result, past studies of performance measurement in the air-to-air and air-to-ground arenas were reviewed in hopes of obtaining some insight into possible reconnaissance task performance measurement techniques.

de Leon (1977) suggested that the true measure of a pilot's performance was his success in target coverage and survivability in actual combat. Of course, such a measure was impossible to obtain in the peacetime training environment which has existed since 1973. In light of this, several prior studies suggested that bombing accuracy could be used as an accurate measure of task performance. Shartle and

Hemphill indicated that "of all the observations made of crew performance during Combat Crew Standardization School evaluations, those concerning accuracy of bombing could be considered to have greatest relevance to the ultimate criterion of combat effectiveness" (Shartle & Hemphill, 1950, p.9). Additionally, Pierce, De Maio, and Eddowes (1979) used bombing scores of student F-4 pilots to validate a proposed rating system of elements of the bomb delivery maneuver. These studies suggest, then, that while not a measure of total or true pilot performance, the use of final mission results could serve as an adequate measure of one aspect of successful military pilot performance.

In addition to actual target coverage, ratings of performance on flight maneuvers or elements of particular maneuvers had been suggested by many previous researchers during their study of pilot performance. Shannon (1980) found that inflight maneuver ratings provided valid and reliable measures of performance. Erickson and Burge (1978) identified a set of factors which might be used in developing a target acquisition and survivability index thus providing a measure of performance. Shannon and Waag (1973) showed a significant number of relationships between ratings of pilots by squadron commanders and past performance. Pierce et al. (1979) developed a subjective pilot rating form which called for the instructor pilot (IP) to assess performance on the critical stages of specific maneuvers in air-to-ground bombing. Finally, in all flight training programs it is common practice for the IP to rate a pilot's ability to perform flight maneuvers following a training sortie.

These past studies have relied almost exclusively on bomb scores, ratings of critical mission elements, and ratings by supervisors as criteria of performance. Of these, adaptations of bomb score measures and supervisor ratings seemed to possess the most potential as criterion measures of successful reconnaissance pilot performance. In this study, the ultimate criteria of pilot performance was a measure of final mission outcome. That is, just as bomb scores have served as a criterion of bomber and fighter pilot performance, so to may a reconnaissance pilot's ability to optimally acquire his assigned targets on camera film serve as the most relevant measure of his performance.

Although frequently used, rating scales presented the problem of having an IP fly with the pilot and rate performance at the conclusion of the mission. Rating scales, then, presented two undesirable outcomes. First, it limited measuring performance to only a few of the total flights flown. Second, the affect that the IP's presence in the aircraft had on mission results was not determinable. The findings of Shannon and Waag (1973), however, indicated that squadron supervisors were able to rate a pilot's ability based on his past performance. This suggested that a second criterion could possibly be obtained by using squadron supervisors as raters. This presented the opportunity to test the ability of supervisors to rank their unit's pilots based on their knowledge of each pilot's past performance or demonstrated ability.

Ranking of a unit's pilots did not seem to have been formally researched. However, ranking seemed to be practiced both formally and informally within most Air Force wings and squadrons. (Although many of the individuals involved did not attach the term ranking to the process they utilized.) Rankings seemed to be used by squadron supervisors to identify aircrew members for such events as special upgrade training programs (eg. instructor pilot upgrade), special mission qualifications, and for participation in flying exercises. From a research standpoint, ratings are supported by Shannon and Waag (1978) who suggested that those pilots most proficient and knowledgable in the mission (in this case squadron supervisors) are the best equipped to identify those pilots who best perform their mission.

Rank-ordering is a form of personnel rating, specifically termed a personnel-comparison system (PCS) (McCormick & Ilgen,1985). The rank-order system was preferred over paired-comparisons, and forced distributions since ranks were the simplest and least time consuming of the three to accomplish. Additionally, rank-ordering forced the rater to make a comparison among individuals and assign a specific rank position to each. Conversely, rating scales are often reduced in effectiveness due to the raters tendency to pile up the ratings at one end of the scale. Consequently, rating scales often do not differentiate adequately among individuals. Rank-order systems avoid this problem since individuals are rated relative to each other (McCormick & Ilgen, 1985). Ranks, then, are an ipsative measure and as such have a built in systematic restraint which specifies that the rater has a set number of ranks and must use each one of them once and only once and that he/she must use all of them (Kerlinger, 1973). In conclusion, ranks were of interest in this study since they forced the raters to spread out their ratings, seemed to have moderate reliability, allowed for relative judgment on the part of the rater, and because they were amenable to reflecting order of merit in one ability, which in this study was the individual's ability to fly day low level reconnaissance. Finally, the intent was to determine if rank-ordering could be used in future studies as a criterion of pilot performance.

Statement of the Problem

The purpose of this study was to address several problems identified in the review of the literature. First, was a pilot's ability to photographically acquire assigned targets a consistent measure of task performance? Second, were supervisor's rank-ordering of a group of pilots' ability to successfully accomplish a day low level reconnaissance sortie a reliable and valid predictor of actual task performance? This second problem presented two specific questions: (a) Was there a high amount of agreement among raters and groups of raters? (b) Was there a significant relationship between a pilot's actual flying performance (his target acquisition) and his assigned rank-order? Finally, do measures obtained from BAT tests contain an indicator or taxonomy of indicators that are unique to the attributes possessed by the successful reconnaissance pilot? In other words, which BAT test or tests are related to successful target acquisition?

Independent and Dependent Variables

The first set of independent variables were from the subject's BAT scores. The BAT-Version 4 test battery was modified to include only those tests which have shown the most promise of being significantly related to pilot performance or that appeared to include job aspects of the tactical reconnaissance mission. A summary of the BAT test battery is at Appendix A. The two-hand coordination and complex coordination tests had been shown to be related to final UPT outcome, an individual's ability to pass or not pass UPT (Kantor & Bordelon, 1985). The item recognition test had been shown to be related to final UPT outcome as well as being related to Advanced Training Recommedation Board (ATRB) ratings (FAR vs TTB) (Carretta, in press) (Kantor et al.,1986). The decision making speed test had also been shown to be related to ATRB ratings. The embedded figures test and the mental rotation test both seemed to consist of measures (field dependence/independence and mental-spatial transformation and classification) which were required for a reconnaissance pilot to quickly and accurately locate targets on the ground. In summary, tests included in the reconnaissance BAT configuration were (a) two-hand coordination, (b) complex coordination (stick and rudder), (c) embedded figures, (d) mental rotation, (e) decision making speed, and (f) item recognition.

The rank-ordering system was the second set of independent variables. Rank-ordering was tested to determine its adequacy as a possible future criterion. Squadron supervisors and standardization evaluation crew members served as judges who rank-ordered operational squadron assigned pilots. (Standardization evaluation personnel are tasked with administering evaluation flights to aircrew members and were selected to rank-order pilots due to their assigned role of assessing a pilot's performance.) Rank-ordering was a qualitative process in which judges ranked pilots in hierarchial order. The rank-order arrived at was based on the judge's relative judgment of each pilot's ability to accomplish a day low level reconnaissance sortie.

The dependent variable was a measure of a pilot's final mission outcome. Final mission outcome was defined as the pilot's ability to optimally acquire his assigned targets on camera film. This procedure was similar to that used in the joint military reconnaissance competition named Reconnaissance Air Meet (RAM) 86 which was sponsored by Tactical Air Command (TAC) in October 1986. In this competition, post-mission photographic analysis consisted of comparing actual target location on the sensor film against optimum placement.

To support the conceptual validity of film scores as a measure of task performance the hypothesis is proposed that if photographic coverage of assigned targets is a consistent measure of task performance, then performance among sorties will be consistent. Furthermore, the performance of high scoring pilots will be more consistent than that of lower scoring pilots. To test this hypothesis film scores were operationally defined. Optimum placement was defined as the location on the film which maximized photographic interpretation of the target for intelligence gathering purposes. Comparison of actual to optimum target location resulted in a value assigned to each target. The values for all targets attempted during a sortie were averaged. This average yielded a film score for the sortie. The average of the film scores for all sorties flown was named Target Acquisition Score (TAS). This procedure was similar to that used to assess task performance in air-to-ground bombing in that the pilot's responsibility was to identify and acquire a ground target, the success of which produced an objective measure and served as a unidimensional indicator of actual performance.

Extraneous Variables

The possibility existed for many extraneous variables to influence the expected outcome of the study. Due to the unobstrusive nature of the study, however, only a few of these variables were controlled. Variables which had the potential to most profoundly affect the performance measures were (a) crew effect (pilot-Weapon Systems Officer (WSO) interaction) on mission results, (b) pilot experience, (c) mission difficulty, (d) weather, and (e) aircraft and reconnaissance sensor reliability.

The most complicated issue was controlling the affect of the WSO on mission accomplishment. In the extreme, the WSO could have affected mission results in one of two ways: (a) the advantageous situation in which the WSO enhanced pilot performance through outstanding performance himself, or (b) the less desirable situation in which the WSO acted to reduce pilot performance due to inadequate support thus increasing the workload of the pilot to an extent which resulted in mission degradation. Obermeyer (1974) also observed that the performance contributions of, and the pilot's interaction with, other crew members must be dealt with when attempting to determine the level of performance of the pilot. Consequently, the affect the WSO may have had on mission results required an attempt to eliminate, control, or randomize the affect of his presence. In this study, the affect the WSO had on mission results was controlled by randomly assigning a WSO from the squadron WSO population to fly with the subject on each sortie. Other methods of control such as analyzing crew interaction by observing the type, quantity, and quality of intercockpit communication (Obermeyer, 1974), or statistically blocking the WSO affect based on some measurement of WSO experience would have possibly provided the additional benefit of measuring the contributions of the WSO. However, neither of these controls were possible due to either the nonavailability of cockpit communication recording equipment in all aircraft or the absence of a sufficiently large sample of WSO's with a particular experience level.

Flying experience was generally accepted as affecting a pilot's level of performance. Lovell and Steward observed that pilot performance tended to improve as a function of total experience but to decline as a function of pilot age. However, experience tended to offset the affects of age (Lovell & Steward, 1970). The Air Force supports this view by allowing each Major Command (MAJCOM) to establish semi-annual and annual flying requirements based on experience. For example, TAC's use of the Graduated Combat Capability (GCC) sortie requirements is based on experience level (Tactical Air Command Manual 51-50, 1985).

Other experiences may have also affected the level of pilot performance. These included the effects of requalification training as well as breaks in flying experience, such as assignment to non-flying duties. These experiences were, therefore, expected to confound the measurement of the attributes and qualities which the pilot must initially possess in order to demonstrate successful performance in the reconnaissance mission. The following criteria were established for controlling the experience factor: 1. The individual must have been on flying status for six years or less. 2. The individual must have been on active flying status continually since graduation from UPT. 3. The individual must not have attended formal requalification training.

A difference in the difficulty of the targets among sorties could have affected final mission results. One sortie may have been quite easy requiring the pilot to accomplish little more than basic aircraft control whereas the following sortie may have been so difficult that the pilot became task saturated and, consequently, missed targets. Shartle and Hemphill (1950) identified this variable in their attempt to measure strategic bomber crew effectiveness. To control for this variable they developed route profiles for each mission flown. As in the Shartle and Hemphill study, a separate mission profile was developed for each test sortie. To the extent possible targets were selected which were subjectively determined to be of equal difficulty. All subjects flew the same targets when flying a specific sortie.

Weather may have affected the pilot's ability to maintain course enroute to the target or within the target area and/or affect his ability to visually acquire the target. The affect of weather on mission results was beyond control in this study.

Aircraft malfunctions could have affected performance in that a malfunctioning system or systems required a greater amount of attention than was normally allotted to that system. The severity of the malfunction was a crucial issue. If the malfunction was severe the pilot aborted the mission and returned to base. In this situation the pilot was rescheduled to fly the mission at a later date. Less severe aircraft malfunctions do not always clearly mandate that the mission be aborted. In these situations the pilot was required to decide on whether or not to continue the mission.

Reconnaissance sensor malfunctions could have affected the ability to obtain accurate target acquisition scores. The system of obtaining a target acquisition score required that a light sensitive sensor be operative aboard the aircraft. Therefore, a sensor malfunction or degraded film quality rendered the film uninterpretable which resulted in loss of data but not a degradation in performance. To recover the data the pilot was scheduled to refly the sortie on an alternate route. This procedure prevented the pilot from flying against the same targets twice which would have contaminated or spoiled that test.

Hypotheses

The experimental hypotheses proposed that (a) the rank-order assigned to a pilot by his supervisors would be related to his level of task performance, and (b) BAT tests would be related to level of task performance. Two specific hypotheses were formulated. The first hypothesis stated that if rank-ordering is a valid measure of task performance, then a pilot's rank-order will be related to his target acquisition score. This also implied that if rank-ordering is a reliable measure of performance, then the rank-order assigned to pilots will be consistent among raters. The second hypothesis stated that, if the BAT test is a valid measure of operational task performance, then at least one of the six BAT subtests will be related to target acquisition scores.

CHAPTER II

METHOD

Subjects Twenty two RF-4C pilots assigned to the 67 Tactical Reconnaissance Wing (TRW) at Bergstrom AFB, Texas served as subjects. Within each of two operational squadrons participants included twelve subjects from the 12 Tactical Reconnaissance Squadron (TRS) and ten subjects from the 91 TRS. Twenty four pilots participated at the start of the study. However, during the course of the research six subjects were reassigned outside the wing. Four of these subjects were replaced by newly acquired pilots. The remaining two losses did not have replacements. This resulted in only twenty two subjects completing all requirements.

Subjects were selected for participation based on their meeting the following criteria: 1. The individual must have been on flying status for six years or less. 2. The individual must have completed initial RF-4C mission qualification training and must have been declared mission ready (MR) by his assigned squadron. 3. The individual must have been on active flying status continually since graduation from UPT. 4. The individual must not have attended formal requalification training.

To further control for differences in experience an attempt was made to limit selection of subjects to those who had followed a direct pipeline from UPT to an operational tactical reconnaissance squadron. However, due to pilot availability this goal was not obtained. Subjects included two pilots who had been assigned as T-37 First Assignment Instructor Pilots (FAIPs) and one pilot who had been assigned as a T-33 tow target pilot prior to being assigned to a tactical reconnaissance unit.

Fourteen squadron supervisors participated as judges. Their task was to rank-order subjects who were assigned to their particular squadron. Two groups of seven judges each were selected from the two operational squadrons. Those serving as judges consisted of the squadron commanders, operations officers, assistant operations officers, and eight flight commanders.

Finally, eight 67 TRW Standardization Evaluation Flight Examiners (SEFE) were selected as judges. Four teams consisting of one pilot and one WSO were formed and tasked to rank-order all twenty- two subjects. Justification for this pairing was based on the standard procedures used to evaluate a pilot's proficiency in flying day low level reconnaissance. Typically, day low level standardization evaluation checkrides are administered by SEFEs, one pilot and one WSO, who follow the crew being evaluated in a chase aircraft during a low level sortie. In this way they evaluate the pilot's ability to fly the aircraft, and both the pilot's and WSO's ability to accomplish the day low level reconnaissance mission. In addition, creating teams increased the likelihood that at least one of the judges had personal, first hand knowledge of the pilot's ability.

Materials

Ranking. Each subject's name was printed on a three-by-five card. To accomplish the ranking task within each squadron, the cards were sorted into piles identifying the squadron to which each subject was assigned. The names were then stacked in alphabetical order. For the ranking task accomplished by the 67 TRW standardization evaluation personnel the names of all subjects from both squadrons were arranged in alphabetical order. To ensure consistency of the factors considered when rank-ordering subjects, each judge was presented with the following instructions.

You have been selected to participate in a research study designed to improve the Air Force pilot selection and classification process. In particular, the intent of the present study is to determine the effectiveness of the selection process in placing pilots in the tactical reconnaissance system following completion of Undergraduate Pilot Training.

Each pilot within your wing/squadron performs, at least to some degree, differently. That is, each pilot has his own strengths and weaknesses. Some are usually better at one phase or phases of a mission than other pilots in the wing/squadron. Some are better overall pilots.

The cards provided you contain the names of selected pilots in your wing/squadron who have six years or less flying experience. Your task is to rank-order these pilots based on their overall day low level reconnaissance performance. Rankings should not be based on rank or grade, extra duties assigned or accomplished, flight qualifications, or by reference to the squadron/flight to which the individual pilot is assigned. Order these cards so that you place the best pilot first, on the top of the stack, second best pilot second, sequentially such that the pilot ranked last is placed last in the stack of cards. While ranking these pilots remember that no ties are permitted. Each pilot must be assigned only one specific rank.

When completed with the task, number the cards first to last, secure the stack with a rubber band, place the stack in the folder, and return the folder to the researcher.

THESE RANKINGS WILL BE USED FOR RESEARCH PURPOSES ONLY. THEY WILL NOT BE USED FOR PERSONNEL DECISION-MAKING, NOR WILL THEY BE PLACED IN ANY INDIVIDUAL'S PERSONNEL FOLDER. PLEASE DO NOT REVEAL

YOUR RANKINGS TO ANY PERSONNEL OTHER THAN THE RESEARCHER.

THANK YOU FOR YOUR COOPERATION !

<u>Flight profiles</u>. Seven day low level reconnaissance sorties were constructed. The first three low level sorties were designated as the primary ones to be used in the study. The remaining four sorties were designated as alternate sorties. Three of these alternate sorties were used in the event that the weather precluded flying the primary sortie or if the primary low level route was unavailable. The last alternate sortie was developed to be flown in the event the pilot was scheduled for Low Altitude Awareness Training (LOWAT).

Sorties were flown within approved low level corridors controlled by the 67 TRW. Each sortie was assigned four targets which were described on an Air Reconnaissance Request/Task Message (ATO), Bergstrom Form 1. Appendix B consists of a sample ATO. Targets were selected from the 67 TRW master target bank. Targets selected for each sortie were subjectively determined to be of equal difficulty by the researcher and a standardization evaluation assigned WSO. Both primary and alternate targets for each sortie were matched as closely as possible. In addition, only targets which the pilot would be able to visually acquire during target run-in were selected. This procedure provided a means for the pilot to align the aircraft with the target to gain optimum photography of the target area. All targets were assigned as oblique targets, either side oblique (SOB) or forward oblique (FOB). Oblique photography was used since it required the most precise placement of the aircraft, by the pilot, in relation to the target in order to acquire optimum photographic coverage.

A mission scenario was developed for the sorties. The mission scenario consisted of a mock wartime situation briefing presented by squadron intelligence personnel. Included in the briefing was the current battle situation, forward edge of battle area (FEBA), enemy troop location, enemy aerial defenses, target area briefing, safe crossing procedures, and escape and evasion procedures. After receiving their target assignments and intelligence briefing each crew planned their own sortie using 1:250,000 scale charts.

Basic Attributes Test. The BAT is a computer-based testing system which consists of microcomputer administered tests. The purpose of this test battery is to assess a candidate's characteristics in the two areas of psychomotor skill and cognitive ability. The version of the BAT used in this study consisted of a biographical data section and six subtests.

The BAT battery first measures the subject's psychomotor skills. The psychomotor tests consists of the two hand coordination and complex coordination subtests. The two hand coordination test (2HN)

> is a variation of an old rotary pursuit task in which a target box traverses a circular path on a CRT at a rate of 20 cycles per minute. The rate of movement of the target box within each cycle varies in a fixed sinusoidal pattern. The subject controls the vertical and horizontal movement of a small cross (zero order dynamics) using a left and right joystick, respectively. While the original psychomotor device version of this test uses two dual axis joysticks (isotonic), the BAT version uses a left hand single axis control device and a right dual axis device (both spring centered). Direction of control and the fact that each control device is restricted to a single axis effect (left-vertical, right-horizontal) remain the same. The subject receives instructions followed by a 3 minute practice and 5 minute test run. Both horizontal and vertical tracking error score are recorded as are respective axis stick movement rate scores.
> The second subtest, complex coordination (PS2), involves the use of a dual axis joystick (right hand, first order dynamics) to control the horizontal and vertical movement of a small cross. The original task's rudder pedals are replaced by the BAT single axis left hand joystick to control the left-right movement of a vertical "rudder bar" of light at the base of the CRT (also, first order relationships). The subject's task is to maintain the cross (against a constant horizontal and vertical rate bias) centered on a large cross fixed at the center of the CRT while, at the same time, centering the rudder bar at the base

of the CRT also against a constant bias. Instructions, practice, testing, and scoring are as in the first subtask (BAT-Version 4, 1985).

The psychological factors measured by the first subtest include low to moderate order tracking and time sharing ability in pursuit tracking. The second subtest measures compensatory tracking tasks involving multi-axis continuous events (BAT-Version 4, 1985).

The remaining tests measure different cognitive capabilities. A brief description of each test follows.

EMBEDDED FIGURES (EMB) - The subject is presented with a simple geometric figure and two complex geometric figures. His task is to decide which of the two complex figures has the simpler figure embedded within it and to indicate a choice by pressing the button corresponding to that figure. Speed and accuracy of response measures are taken. PSYCHOLOGICAL FACTORS: Field dependence/field independence.

MENTAL ROTATION (MRT) - The subjects are presented sequentially with a pair of letters and asked to make a speeded same-different judgement. The letter pair may be either in the same orientation, or rotated in space with respect to the other. A correct "different" judgement is associated with a mirror image and is not a function of relative rotation. In order to perform the task, the subject must form a mental image of the first letter (no longer displayed) and perform a point-by-point comparison with the second (which remains on the display). In addition, when the letters are rotated with respect to each other, the subject must mentally rotate the mental image of one letter into congruence with the other before undertaking the comparison. PSYCHOLOGICAL FACTORS: Mental-spatial transformation and classification.

ITEM RECOGNITION (ITM) - In the item recognition paradigm, a series of one to six digits is presented in a row on a CRT display, removed and followed, after a brief delay, by a single digit. The subject is instructed to remember the initial series of digits, then to decide if the single digit is one of those presented in the initial series. The subject is instructed to push one button (marked "yes"), if the single digit was in the series; another (marked "no") if not. The subject is instructed to make a response as quickly and accurately as possible. PSYCHOLOGICAL FACTORS: Short term memory store, search, and compare operations.

DECISION-MAKING SPEED (DMS) - In this choice reaction time task, one of a number of alternative signals is presented to the subject. The subject is required to respond to the signal with the matching response as quickly as possible. The key to this task is the amount of uncertainty that must be resolved in order to make the response decision. When more alternative signals may potentially be presented, greater uncertainty exists and the decision is made more slowly. This task consists of four subtasks each with three parts: in part one, two potential signals and two responses are defined (1 bit); in part two, four potential signals and responses (2 bits); and part three, eight potential signals and responses are defined (3 bits). In subtask one the subject knows both where and when a signal is to occur; in subtask two, where but not when; in subtask three, when but not where; and, finally, in subtask four the subject knows neither where nor when. PSYCHOLOGICAL FACTORS: Simple choice reaction time under various conditions of information load and conditions of spatial and temporal uncertainty. Low level cognitive, high level sensory-perceptual motor involvement (BAT-Version 4, 1985).

Finally, analysis of each of these subtests has indicated that the BAT has high test reliability. Appendix C is a table of the reliability estimates for each of the BAT tests used in this study.

Procedure

Subject Selection. Subject selection was accomplished by reviewing the flight records of all 67 TRW assigned pilots to determine which pilots met the requirements for participation in the study. Each squadron was then provided a list of potential subjects. The squadron was requested to indicate which of those pilots on the list would be available to participate in the research. The list of available subjects was then approved by the 67 TRW Deputy Commander of Operations (DCO). This procedure produced a total of 24 pilots who were to serve as subjects. When possible, subjects who dropped out of the study were replaced on a one-to-one basis by incoming pilots who met the prerequisites for participation in the study.

Pilots selected to participate in the study were initially notified by their squadron operations officer. Prior to beginning participation in the study each subject received instructions which described the purpose of the study and the procedures to be followed in accomplishing the sorties they were to fly. A copy of the instructions to subjects is at appendix D.

Ranking Task. Squadron supervisors served as judges. They were given the task of rank ordering only those pilots that were assigned to the squadron to which they themselves were assigned. Each judge was presented the ranking task on a one-to-one basis while in the privacy of his office. The researcher explained the overall purpose of the study to the judge and presented him an envelope which contained instructions on how to accomplish the ranking task and the stack of cards containing each subject's name. After reading the instructions the judge was given 30 minutes to rank-order his stack of cards. Upon completion of the ranking task the researcher collected the envelope and answered any questions the judge may have had about the research.

Following the ranking of subjects in each squadron, wing standarization evaluation personnel were tasked with rank-ordering all 22 subjects. Each of the four teams of judges were presented an envelope which contained instructions on how to accomplish the ranking task and the stack of cards containing each subject's name. After reading the instructions the judges were given 45 minutes to rank-order their stack of cards.

The rankings provided by the standardization evaluation teams and the rank-orderings accomplished by squadron supervisors were then combined. The sum totals from this combined set of rankings provided the rank-order of each subject.

Target Acquisition Score. Each subject was initially programmed to fly four reconnaissance sorties. However, due to a constraint on sortie availability, this plan was changed so that each subject flew three sorties. A repeated measures design was used to administer the day low level reconnaissance sorties to the subjects. To control for practice effect of repeated sorties and for possible WSO affect on task performance both sortie sequence and squadron WSO's were randomized for each subject. The sequence for flying the three sorties and for scheduling squadron WSO's to fly with a particular subject was randomly generated by computer. Squadron programmers were provided with this computer output which listed each subject's name, the order the sorties were to be flown in, and the names of four WSO's. (For example, Smith flies sortie 2 first and the WSO's to choose from for sortie 2 are Jones, James, Jackson, or Johnson.) When scheduling a subject to fly a sortie the programmer selected the first one of the four WSO's listed who was available to fly. If none of the listed WSO's were available, then the programmer was given the latitude to schedule any available WSO from within the squadron. Programmers were instructed, however, that a subject could not fly with the same WSO more than twice during the study.

This procedure was utilized due to the unobstrusive nature of the data collection. Squadron programmers scheduled subjects to fly these sortie profiles based on the availability of squadron aircraft and each subject's semi-annual flying requirements. The researcher had no control over squadron scheduling procedures. Therefore, the researcher was unable to control the affect of time of day or frequency of flights made by subjects. The procedure utilized did, however, provide for a randomization of sortie sequence and for the maximum amount of control possible over aircrew (pilot/WSO) pairing.

Squadron programmers scheduled pilots to fly a research sortie the day prior to the actual flight. When scheduled for a sortie the subject was given a copy of the ATO. He then constructed his map, proceeded to squadron intelligence for a threat and situation briefing, and planned his sortie in accordance with Tactical Air Command Regulation 55-4 (1983) requirements. Sorties were flown as preplanned sorties as described in TAC Operational Readiness Inspection (ORI) criteria for target tasking (Air Force Regulation 123-3, 1986). To simulate, as closely as possible, a combat situation the subject was instructed to fly the sortie as a combat profile as was required as part of his semi-annual flying requirements by Tactical Air Command Manual 51-50 (1985). Upon completing flight planning the crew proceeded to the aircraft a minimum of 45 minutes prior to scheduled takeoff time. During the sortie the subject was instructed to record the takeoff time, time over each target, collect Essential Elements of Information (EEI) on each target, and to ensure that the WSO had set the Aux Data Annotation System (ADAS) so that time over target would be recorded on the target film.

Following each sortie the subject debriefed intelligence on the status of his sortie, success or failure. The subject then proceeded to the Photographic Processing and Interpretation Facility (PPIF) and reviewed his film with a Photographic Interpreter (PI). As part of the film review the subject was required to identify each target. Target coverage was then confirmed by the PI. Finally, the frames of film with the best coverage of each of the targets was placed with the map.

The researcher scored the targets for each sortie. Targets were scored by placing the appropriate overlay for the sensor used over the frame of film. The target was then located and a score assigned based on the location of the target in relation to its position on the overlay. Possible scores for each target ranged from 0 to 50 points. Appendix E contains sample target overlay score sheets.

To determine the score for a sortie the points earned for each of the four targets was summed and then averaged. If a target was dropped due to weather or fuel then that target was considered an omission and was not included in computing that sortie's score. As with individual target scores, the score for each sortie ranged from 0 to 50 points.

The sortie scores were then used to produce two measures of task performance. The average of the three sortie scores produced the Target Acquisition Score (TAS). The variation in performance between sorties produced the TAS standard deviation. The TAS provided a measure of level of performance and had a range of 0 to 50 points. The standard deviation provided a measure of consistency of performance.

Basic Attributes Test. The BAT test was administered to all subjects on an as available basis throughout the period of flight data collection. Each subject was scheduled by his squadron to take the BAT test. As with the flying of sorties, the researcher had little influence on the sequence of subjects, time of day (however, all BAT tests were administered between the hours of 0800 to 1700), or activities accomplished prior to taking the BAT test (eg. extra duties performed or number of sorties flown prior to taking the test).

Prior to taking the BAT test the subject was informed of the purpose of the research, his involvement in the research, assured that his performance scores would be held in confidence and would not be released to any 67 TRW personnel, and that his scores would not be used to affect his career in any way. The subject was then placed at the portable test unit, PORTA- BAT, and administration of the test began.

The battery of tests were computer administered. The sequence and flow of the tests were controlled by the computer and were held constant for each subject. The BAT test began by informing the subject how to interact with the computer in order to accomplish the test. Next, it identified the purpose of the research being conducted and the purpose and composition of the BAT test itself. The program then produced a privacy act statement prior to recording the individual's social security number, and a statement of the rights of the individual in regard to the voluntary nature of participation in human research. Finally, the BAT tests were administed in the following order (a) biographical data, (b) two hand psychomotor test, (c) complex coordination, (d) embedded figures, (e) mental rotation, (f) item recognition, and (g) decision making speed.

Prior to beginning each subtest the subject was informed of the purpose of the test, how to accomplish the test, and, with some tests, given a practice session. The practice session, if present, was to familiarize the subject with the mechanics involved in accomplishing the test. No feedback of performance level was provided. Appendix F is a copy of the BAT test instructions presented to the subject.

Statistical Analysis Alpha was established at .05 for all analyses. Target acquisition scores were evaluated for consistency of performance between both sorties and subjects using a repeated-measures, treatment-by-subjects design (Brunig & Kintz, 1977). Rank-ordering was evaluated for internal consistency among judges within a group using Kendall's coefficient of concordance (Downie & Heath, 1974). Agreement of rank-ordering among groups of judges was evaluated using Spearman's rank-order correlation.

The validity of rank-ordering was determined by analyzing the relationship of rank-order to target acquisition score using Spearman's rank-order correlation. The relationship of each of the scales derived from the BAT subtests to target acquisiton score and TAS standard deviation were determined using multiple regression techniques ($SPSS^X$,1983).

CHAPTER III

RESULTS AND DISCUSSION

Testing the Reliability of Target Acquisition Scores Target film scores were used to produce two distinct measures of performance. Average film score for the three sorties, TAS, served as a measure of level of performance. TAS standard deviation served as a measure of consistency of task performance. For the 22 subjects TAS had a range of 15.83 to 42.5, a $M = 30.63$, and $SD = 6.95$. TAS standard deviation ranged from 1.44 to 15.28, with a $M = 6.25$, and $SD = 3.38$. TAS scores for the sample reflected a normal distribution of scores, skewness = -.190 and kurtosis = -.197.

A repeated-measures, treatments-by-subjects design was used to test the hypothesis which stated that performance would be consistent among sorties. A second purpose was to determine if there were significant differences in the consistency of subject's performance on each of the sorties. Table 1 displays the results of the analysis of variance which indicated that the sorties were consistent among themselves, $F\ (2, 42) = 1.91, p > .05$. Conversely, the analysis of variance indicated that subjects varied in their performance on each of the sorties, $F\ (21, 42) = 3.08, p < .01$.

The first hypothesis was retained. The fact that each of the sorties was performed similarly indicated that film scores were a consistent measure. The consistency of film scores provided support for their possible use as a measure of task performance. These results, then, supported the findings of Shartle and Hemphill (1950) and Peirce et al. (1979) who concluded that final mission outcome

TABLE 1

Analysis of Variance of the Consistency Between Sorties

Source	SS	df	ms	F	p
Total	5110.11	62	-	-	-
Subjects	2991.46	21	142.45	3.08	<.01
Treatments	176.28	2	88.14	1.91	>.05
Error	1942.36	42	46.25	-	-

served as a relevant criterion of performance. The results further indicated that no one sortie alone caused the observed variation in individual subject task performance.

Since the results of the analysis of variance indicated that pilots varied among themselves in their performance between sorties, the post hoc hypothesis that level of performance was related to consistency of performance was tested. A pearson correlation coefficient was computed between TAS and TAS standard deviation. The results suggested that high performers were more consistent in their performance than low performers, however the relationship was not statistically significant, r (20) = -.36, $p > .05$. Consequently, the hypothesis that there was a significant relationship between level and consistency of performance was rejected. These results implied that target acquisition score and standard deviation may have been measuring two different aspects of performance and that both should be retained as performance indices.

Relationship Between Rank-Order and Pilot Performance To establish the reliability of the rank-order system the consistency of each of the three groups of judges was tested. Kendall's coefficient of concordance was used to test the amount of agreement among judges in each separate group (Downie and Heath, 1974). Results of the analysis indicated that there was a high amount of agreement among judges in their rank-ordering of pilots based on the pilot's ability to fly day low level reconnaissance. The outcome for each group of supervisors was: (a) 12 TRS, W (7, 12) = .85, $p < .001$; (b) 91 TRS, W (7, 12) = .82, $p < .001$; and (c) SEFE, W (4, 24) = .69, $p < .001$.

The high amount of agreement among judges indicated that supervisor rankings were a reliable measure. That is, the rank-orders produced were repeatable or reproducible. Most notable was the high amount of agreement between 12 TRS and 91 TRS supervisors. This indicated that those most likely to be faced with the task of determining the performance capabilities of assigned pilots were most likely to agree on those capabilities.

To add support to the estimates of reliability obtained, a test for the relationship among groups of judges was performed. Of the original 24 subjects, 18 had been ranked by two different groups of judges. Of these 18, one group of nine had been rank-ordered by 12 TRS supervisors while the second group was rank-ordered by 91 TRS supervisors. Standardization evaluation personnel (SEFE) had rank-ordered all 18. The Spearman rank-order correlation was used to determine if there was a significant relationship between rank-orders assigned to each of these two groups. The results indicated a significant relationship between groups of judges in rank-ordering both groups of pilots; (a) 12 TRS to SEFE rankings, r_s (7) = .83, $p < .01$, and (b) 91 TRS to SEFE rankings, r_s (7) = .93, $p < .001$. Thus, the assumption of reliability of the ranking technique was provided further support. The rank-order system used in this study, then, supported the findings of McCormick and Ilgen (1985) who reported the statistical reliablity of rankings.

The reliability of the rankings implied that the rank-orders produced may have been a valid measure of task performance. However, such an assumption was not directly supported nor denied by determining the amount of agreement among supervisors alone. Support for the validity of supervisor rankings required the relationship between actual performance and rankings be tested.

To determine the relationship between a subject's rank-order and performance it was desirable to compare all subjects to one another. Since there was a significant relationship between all three groups of judges the rankings were combined so that in the final outcome each subject was rank-ordered in relation to all others. To accomplish this the rankings of the 18 judges were pooled together and summed. The total of the ranks was then used to assign a rank-order to each of the 22 subjects who had completed the study. In essence, this procedure acted to rank-order subjects according to the rankings produced by standardization evaluation. Squadron supervisor rankings acted as tie breakers. The combined rank-ordering acted to maintain the original rank-order produced by each individual group to the maximum extent possible. The relationship between the combined rank-order and each group's original rank-order was; (a) 12 TRS to combined rank-order, r_s (20) = .93, (b) 91 TRS to combined rank-order, r_s (20) = .91, and (c) SEFE to combined rank-order, r_s (20) = .97, all of which were significant at the $p < .001$ level of significance.

A Spearman rank-order correlation was used to determine if rank-order was related to a subject's TAS (level of performance). The results of this analysis indicated that a subject's rank-order was not significantly related to his TAS, $r_s = .24$, $p > .05$.

The hypothesis stating that rank-order was related to task performance was rejected. The results of the analysis of the rank-order system indicated a high amount of agreement in the rank-ordering of subjects both within and between the three groups of judges. However, the rankings produced were not significantly related to the measure of task performance. Therefore, the observation made by Shannon and Waag (1973) that supervisors ratings were related to their performance was not supported by this study.

Contamination of the Rank-Order System These results presented the problem of identifying the variables which may have explained differences between a pilot's performance and his assigned rank-order. Of possible variables, experience had been the most commonly used by the Air Force as a determinant of performance. In addition, Lovell and Steward (1970) had observed that performance improved as a function of experience. Therefore, four experience variables were identified which uld possibly explain performance and rank-order. First, was the amount of time a ject had been qualified as a pilot. This variable was defined as the number of hs which had expired since the subject's pilot aeronautical rating date. Second, was the subject's total flying time. This variable was defined as the total amount of flying time the subject had logged as a crew member in any Air Force aircraft. Analysis of these two variable indicated that number of months qualified as a pilot and total flying time were highly intercorrelated, $r(22) = .95$, $p < .001$. Consequently, only total time was used in further analyses since the number of months that an individual had been qualified as a pilot did not appear to make any unique contributions in explaining performance. Third, was the subject's RF-4C flying time. This variable was defined as the total amount of flying time the subject had logged only in the RF-4C aircraft. Fourth, was the amount of time the subject had been assigned to his present squadron. This variable was defined as the number of months which had passed since the subject had arrived at his assigned squadron and was named time on station. Time on station appeared important as it provided an indice from which to analyze the extent that rank-order was determined by familiarity with the subject. That is, time on station provided a means to determine if raters rank-ordered subjects merely by reference to the amount of time which they had known the subject.

A multiple regression equation was used to determine if level of performance was related to total flying time, RF-4C flying time, and/or time on station. Table 2 displays the results which indicated that total flying time alone was significantly related to TAS, multiple $R = .63$, $p < .001$. Inclusion of RF-4C flying time into the model increased the predictive abilities of the model only sightly, multiple $R = .64$, $p < .001$. A test of the reduced to complete model indicated that the complete model did not add significantly to the prediction of TAS scores, $F\ (1/19) = .32$, $p > .05$. Time on station did not enter into the equation. As a result RF-4C flying time and time on station were deleted from the equation.

Next, a multiple regression equation was generated to determine if rank-order was predictable from total flying time, RF-4C flying time, and/or time on station. Table 3 displays the results which indicated that RF-4C flying time alone contributed most to the prediction of rank-order, multiple $R = .64$, $p < .001$. The addition of total flying time into the equation increased the predictive abilities of the model only sightly, multiple $R = .65$, $p < .001$. A test of the reduced to complete model indicated that the complete model did not add significantly to the prediction of

TABLE 2

TAS: Pilot Experience Regression Analysis

Source	SS	df	MS	F	p
Regression	410.59	1	410.59	13.42	< .001
Residual	611.83	20	30.59		
Total	1022.42	21			

Multiple R = .63

R Square = .40

Standard Error = 5.43

rank-order, $F(1/19) = .33$, $p > .05$. Time on station did not enter into the equation. Consequently, total flying time and time on station were deleted from the equation.

These results provided one possible explanation for the lack of relationship between rank-order and task performance. Results of the analyses indicated that task performance was dependent on total flying time. Conversely, rank-order was dependent on RF-4C flying time. As a result, the conclusion was drawn that rank-order was not related to performance since rankings were, at least in part, explained by different aspects of experience. Since rank-ordering failed to show promise as a criterion measure of task performance it was dropped from further analyses.

While not the specific purpose of this study, the results of this analysis both

TABLE 3

Rank-Order: Pilot Experience Regression Analysis

Source	SS	df	MS	F	p
Regression	362.64	1	363.64	13.87	< .001
Residual	522.86	20	26.14		
Total	885.50	21			

Multiple R = .64

R Square = .41

Standard Error = 5.16

supported and clarified the findings of Lovell and Steward (1970) and lend support to current TAC policy. As Lovell and Steward observed, this study indicated that performance did improve as a function of experience. However, this study clarified their findings by showing that as a predictor of task performance total flying time explained performance to a greater degree than experience in any one specific aircraft. Further, these results acted to validate current TAC policy which assigns more weight to total flying time than to assigned aircraft time in establishing crew member experience level and as a basis for selection for upgrade training (Tactical Air Command Manual 51-50, 1985).

Testing the Performance to BAT Test Relationship To test the hypothesis that at least one of the BAT tests would be related to task performance each of the tests were compared separately to measures of performance. Results from each test were transformed into a set of variables which reflected the subject's performance on the test. In all, a set of 16 variables were formed from the six BAT tests. Unfortunately the sample size was extremely small. Complete test data were available for only 19 of the 22 subjects who participated in the study. Consequently, a forward stepping multiple regression technique was employed to identify those variables, within each test, which were most highly related to reconnaissance pilot task performance.

Two Hand Coordination and Complex Coordination Tests The two hand coordination and the complex coordination tests reduced to form five distinct scores. The two hand coordination test reduced to left-right axis (horizontal) tracking error, and up-down axis (vertical) tracking error. The complex coordination test reduced to left-right axis (horizontal) tracking error, up-down axis (vertical) tracking error, and left-right axis (horizontal) tracking error of the short bar of light (rudder). Previous studies have indicated that only two of these five scores were needed. These were horizontal (X-1 axis) tracking error from the two hand coordination test and the vertical (Y-2 axis) tracking error from the complex coordination test (Kantor & Bordelon, 1984).

Horizontal tracking error of the two hand coordination test had a M = 5211.08 and a SD = 1658.80. Vertical tracking error of the complex coordination test had a M = 10713.33 and a SD = 3506.34. A multiple regression technique was used to determine the relationship of these two psychomotor scores to TAS and standard deviation. Table 4, at Appendix G, reveals that the psychomotor tests were not significantly related to target acquisition score, multiple $R = .06$, $p \leq .9731$. In addition, neither psychomotor test score was significantly related to score standard deviation, multiple $R = .17$, $p \leq .7875$.

This study indicated that psychomotor skills were not related to level of performance as a reconnaissance pilot. The failure of these tests to discriminate among reconnaissance pilots was most likely caused by the continuing selection process which occurs as new pilots are placed in advanced weapon systems. As Kantor and Bordelon (1985) indicated psychomotor tests were significantly related to prediction of successful completion of UPT. It is likely that completion of UPT and then FAR selection have acted to attenuate the range of psychomotor scores such that there is not a significant difference in psychomotor skills among those individuals flying fighter type aircraft.

Decision Making Speed Test The decision making speed test reduced to 12 scores based on the subject's knowing or not knowing when and where symbols would be presented. The conditions were (a) where and when the symbols would appear, (b) when but not where the symbols would appear, (c) where but not when the symbols would appear, and (d) a situation in which the subject did not know where or when the symbols would appear. For each condition there were three trials which differed in number of potential signals.

The accuracy of responses across all 12 trials of the subtest were consistently high with a range of 94 to 100 percent correct. This high accuracy of response was expected since prior studies have shown that the critical factor in this test was the response time of the individual. The results indicated that decision making speed decreased as the number of symbols increased and when the subject did not know when the symbols would be presented. Using these observations, Carretta (in press) has shown that the 12 score means could be further collapsed to form a stable set of five variables. These variables were the average and standard deviation of the when and when not conditions and the total percent correct.

A multiple regression technique was used to determine the relationship of the decision making speed test to TAS and standard deviation. Table 5, at Appendix G, displays the results which indicated that decision making speed was not significantly related to target acquisition score, multiple $R = .39$, $p \leq .2608$. However, decision making speed was related to TAS standard deviation, multiple $R = .67$, $p \leq .0267$. Total percent correct responses, average response time when it was not known when the symbol would be presented, and standard deviation of response time for the when not condition accounted for most of the relationship of the test to TAS standard deviation. Average response time and standard deviation for the when condition improved the relationship, multiple $R = .73$, however, the increase was not significant, $F(2/13) = 1.08$, $p > .05$.

Decision making speed variables were most closely related to consistency of task performance. An explanation for these findings is that the reconnaissance mission operated in a high speed, low altitude arena requiring rapid decisions in a dynamic environment. Those pilots who were able to make consistently rapid and accurate responses during these dynamic flight situations were able to perform the reconnaissance mission more consistently. Likewise, performance on the decision making speed test increased with the ability to make quick, accurate responses during periods of temporal uncertainty. These results support the findings of Carretta (in [illegible]) which [illegible]cated that decision making speed was most closely related to those UPT check flights requiring quick, consistent, and accurate responses.

Item Recognition Test The item recognition test reduced to seven scores. The[illegible] the overall percent correct and the average response time for each of the six different string lengths. As expected the accuracy of responses was very high with responses ranging between 94 to 100 percent correct. Carretta (in press) observed that response time increased linearly as string length increased. Consequently, the data were further reduced by creating a regression line for each subject. This produced a set of four variables (a) slope, (b) intercept, (c) standard error, and (d) overall percent correct responses (Carretta, in press).

A multiple regression technique was used to determine the relationship of item recognition variables to TAS and standard deviation. Table 6, at Appendix G, displays the results which revealed that item recognition was significantly related to TAS, multiple $\underline{R} = .84$, $\underline{p} \leq .0001$. Overall percent correct responses and slope alone were related to TAS. Intercept and standard error did not enter into the regression equation. Although in the expected direction, none of the item recognition variables were significantly related to score standard deviation, multiple $\underline{R} = .39$, $\underline{p} \leq .0974$.

These results indicated that subjects who had higher slopes and higher percent of correct responses performed better as reconnaissance pilots. As suggested by Carretta (in press) these subjects probably used a more efficient memory searching strategy than those subjects who took the same or less amount of time to respond to different string lengths. One explanation for this relationship is that for the reconnaissance pilot to successfully perform his mission he had to develop a mental image of a target from a map and then search and compare possible objects on the ground in an attempt to identify the desired target. Consequently, it is possible to conclude that the factors measured by the item recognition test (short term memory storage, search, and comparison operations) were also required to successfully perform as a reconnaissance pilot.

<u>Mental Rotation Test</u> The mental rotation test reduced to seven scores. These were overall percent correct and the mean response time for each of six

variations in rotation. Symbols were rotated at 0, 60, 120, 180, 240, and 300 degrees. The accuracy of response was very high regardless of the rotation of the symbol. Responses ranged between 54 to 100 percent correct, with $\underline{M}$ = 91.61, and $\underline{SD}$ = 11.04. The average time for a correct response also remained reasonably stable regardless of the number of degrees of rotation of the symbols, range 708.78 to 2242.36 ms. Consequently, as in past studies the data were further reduced to form three variables; (a) average response time for a correct response, (b) standard deviation of response time, and (c) overall percent correct responses (T. R. Carretta, January 16, 1987).

A multiple regression technique was used to determine the relationship of the mental rotation variables to TAS and standard deviation. Table 7, at Appendix G, displays the results which revealed that mental rotation was significantly related to TAS, multiple $\underline{R}$ = .69, $\underline{p} \leq .0056$. Average response time for a correct response and the percent of correct responses explained the relationship of mental rotation to TAS. Standard deviation of average response time was also related to TAS, $\underline{r}$ = -.41. However, response time standard deviation was highly intercorrelated with correct average response time, $\underline{r}$ = .72, and so did not add significantly to the regression equation. The mental rotation scores were not related to TAS standard deviation, multiple $\underline{R}$ = .39, $\underline{p} \leq .2587$.

As with item recognition, these results indicated that those subjects who were able to make quick and accurate responses performed better as reconnaissance pilots. An explanation for these findings is that during a tactical reconnaissance mission the pilot was required to identify a series of targets on the ground. This involved comparing a target displayed on a map, or a photograph of the target, with an object which may have been the desired target on the ground. This process involved performing a mental-spatial transformation and classification process similar to that required by the mental rotation test. As a result those subjects who were able to quickly identify symbols regardless of their rotation to one another performed better on the reconnaissance mission and the mental rotation test. These subjects were probably more efficient at performing spatial transformations in time compressed situations.

Embedded Figures Test Embedded figures test scores were transformed into two variables (a) percent correct responses and (b) average response time for a correct response (T. R. Carretta, January 16, 1987). The percent of correct responses was extremely low, with a range of 37 to 86 percent. This indicates that the probablitliy of a subject correctly answering an item was near the chance level. Consequently, the item reliability of the test was questionable.

A multiple regression technique was used to determine the relationship of both embedded figures variables to TAS and standard deviation. Table 8, at Appendix G, displays the results which indicated that while in the expected direction the embedded figures test was not significantly related to TAS, multiple $R = .419$, $p \leq .0742$. In addition, the embedded figures test was not related to score standard deviation, multiple $R = .449$, $p \leq .0535$.

These results indicated that the failure of the embedded figures test to discriminate between high and low performing pilots may have been due to the performance of the the test itself. This supported past observations which revealed that the probability of correctly answering an item was below the level of chance. In addition, past results have indicated that the embedded figures test was not related to UPT outcome (pass/fail), UPT checkride scores, or ATRB board results (T. R. Carretta, January 16, 1987). It is unfortunate that this test has performed so poorly.

It seems logical that the psychological factors of field dependence/independence should play a critical role in reconnaissance pilot performance.

Since these data were collected the embedded figures test has been deleted from the BAT battery. In its place a new version of the test which presents three-dimensional embedded figures is being developed. It would seem beneficial to try out this test in future studies of reconnaissance pilot performance.

An Integrated Model Each of the six subtests were expected to measure some unique skill required to demonstrate successful performance as a pilot. If each of the subtests did measure conceptually different skills, then the relationship between the tests and performance might have been improved by using measures from more than one of the subtests in an integrated model. Unfortunately, the limited sample size obtained in this study did not permit the testing of the integrated model. Future studies intended to cross validate these findings would do well to focus on the development of such a model.

CHAPTER IV

CONCLUSIONS AND RECOMMENDATIONS

The final outcome of a mission as measured by its success or failure has been suggested in the past as a possible unidimensional criterion measure of task performance. In this study, a tactical reconnaissance pilot's performance during a mission was determined by his ability to photograph assigned targets. From these photographs, target film scores were derived to develop two unique measures of performance. To add empirical support to the face validity of film scores as a criterion measure, the hypothesis was proposed that film scores between sorties would be consistent. The test of the hypothesis provided support for this proposal by indicating that film scores were a consistent measure of task performance.

While performance between sorties as a whole was a stable measure, results of the analysis indicated that pilots varied in their ability to consistently perform the reconnaissance mission. Consequently, a test of the post hoc hypothesis which stated that high performers would be more consistent in their performance of the mission was performed. Contrary to expectations a test of this hypothesis indicated that level of task performance was not necessarily related to consistency of performance. Consequently, two separate measures of task performance were justified. These measures were level of performance and consistency of performance.

A benefit to future research would be the development of a performance assessment procedure which reduces the time and cost of measuring task performance. Rank-ordering seemed to be one such possibility. Consequently, the hypothesis was proposed which stated that the rank-ordering of squadron pilots by their supervisors would be both a reliable and valid measure of performance. As expected supervisors had a high amount of agreement among themselves on the rank-order assigned to groups of pilots. In addition, rankings among groups of supervisors were highly correlated with each other. This supported the hypothesis that the rank-order system was a reliable measure.

To determine the validity of the rank-order system, tests of the relation between task performance, as measured by TAS, and rank-order were performed. The results indicated a very low relationship between rank-order and task performance. A test of the relationship between experience variables, task performance, and rank-order provided one possible explanation. The results revealed that task performance was related to total flying time. Conversely, rankings had more in common with RF-4C flying experience. Thus the conclusion was drawn that performance and rankings were both dependent on experience. However, experience could be partitioned to explain performance and rank-order separately. As a result the hypothesis stating that rank-order was related to performance was not supported. These conclusions suggested that rating systems, especially rank-ordering systems, should be approached with caution when used as measures of task performance.

Placing new pilots in the type of mission and aircraft in which they can perform to their maximum potential is a highly desirable goal. Identification of tests which accurately predict a pilot's future level of performance in a particular mission or aircraft could contribute towards achieving this goal. The BAT seems to be a positive first step toward identifying such a group of tests. To test this possibility the hypothesis was proposed that at least one of the BAT subtests would be related to reconnaissance task performance. The results of this study supported the hypothesis and demonstrated that three of the BAT tests (Decision Making Speed, Item

Recognition, and Mental Rotation) were highly related to tactical reconnaissance task performance. In conclusion, the BAT test battery, especially the cognitive tests, demonstrated excellent promise in identifying those pilots who perform well in the tactical reconnaissance mission.

Recommendations The sample upon which this study was conducted was extremely small. Limited access to personnel, aircraft availability, and time prevented obtaining the desired sample size. Consequently, it should be recognized that a cross validation study should be accomplished prior to forming opinions or taking actions based on the results of this study.

The use of results from sorties flown in aircraft is both costly and time consuming. In addition, the environment in which the pilot interacts is ever changing and unpredictable. The dynamics of the environment input uncontrollable variables which produce error in performance measurement techniques. These problems present the challenge of continuing to develop a criterion measure which allows control of undesirable variables without exceeding available resources. One such possibility is the use of partial motion simulators designed to measure pilot control inputs and reaction time.

Future studies should include an assessment of the affect that the Weapon Systems Officer has on mission results. Such a quantitative measure of WSO performance could provide beneficial information. First, WSO performance measurement could provide a means for exploring the possible differences in level of performance among WSO's. Such results would be of value when studies on improving the navigator selection process are begun. Second, determining the affect the WSO has on mission results could provide a means of measuring the increased survivability and probability of mission success which an additional crew member may bring to the cockpit.

To further assess the value of the BAT test a cross validation study on a larger sample is required. Additional findings which support the BAT to task performance relationship could justify the use of select BAT subtests as predictors of a pilot's ability to perform the tactical reconnaissance mission. In addition, the revised embedded figures test, when complete, should be tried out again on pilots assigned to this weapon system.

Finally, BAT test results from pilots flying different weapon systems should be analyzed to determine if performance on different BAT tests are dependent on the type of aircraft the pilot flies. Results from such a study could possibly be used to either validate or to suggest improvements in the current ATRB rating system and MPC assignment policy.

References

Air Force Regulation (AFR) 123-3, Tactical Air Command (TAC) Sup. 1. (1986). Langley AFB, Va: Headquarters, Tactical Air Command.

Basic Attributes Test (BAT)-Version 4, Hardware, System, & Individual Test Descriptions. (1985). San Antonio, Tx: Air Force Human Resources Laboratory.

Basic Attributes Test (BAT)-Version 5 [Computer program]. (1985). San Antonio, Tx: Air Force Human Resources Laboratory.

Bruning, J. L., & Kintz, B. L. (1977). Computational handbook of statistics (2nd ed.). Illinois: Scott, Foresman and Company.

Carretta, T. R. (in press). The Basic Attributes Tests (BAT) System: a preliminary evaluation of three cognitive subtasks. (AFHRL Report No. TR-86-XX). San Antonio, Tx: Air Force Human Resources Laboratory.

de Leon, P. (1977). The peacetime evaluation of the pilot skill factor in air-to-air combat (Report No. R-2070-PR). Santa Monica, Ca: Rand Corporation.

Downie, N. M. & Heath, R. W. (1974). Basic statistical methods (4th ed.). New York: Harper and Row.

Erikson, R. A., & Burge, C. J. (1978). Launch opportunity for air-to-ground, visually delivered weapons (Report No. NWC TP 6005), China Lake, Ca: Naval Weapons Center. (From Mixon, T. R. (1982) An annotated bibliography of objective pilot performance measures: Part II. Monterey, Ca: Naval Post Graduate School)

Fleishman, E. A., & Ornstein, G. N. (1971). An analysis of pilot flying performance in terms of component abilities. In W. Ronan, & E. Prien (Eds.), Perspectives on the measurement of human performance. (pp.346-362). New York: Appleton-Century Crofts.

Gerathewohl, S. J., (1978). Psychophysiological effects of aging - developing a functional age index for pilots. II. Taxonomy of physiological factors. (Report No. FAA-AM-78-14). Washington, DC: Federal Aviation Administration. (From Rehman, J. T. (1982) Pilot performance measurement: an annotated bibliography. Atlantic City, N.J: Federal Aviation Administration Technical Center)

Imhoff, D. L. & Levine, J. M. (1981). Perceptual-motor and cognitive performance task battery for pilot selection. (Report No. AFHRL-TR-80-27). San Antonio, Tx: Air Force Human Resources Laboratory.

Kantor, J. E., & Bordelon, V. P. (1985). The USAF pilot selection and classification research program. Aviation, Space, and Environmental Medicine, 56, 258-261.

Kantor, J. E., Carretta, T. R., & Quebe, J. C. (1986). Aircrew selection systems. Manuscript submitted for publication.

Kerlinger, F. N. (1973). Foundations of behavioral research. (2nd ed.) New York: Holt, Rinehart and Winston, Inc.

Lovell, C. C., & Steward, W. A. (1970). Aircraft and pilot performance in dive bombing. (Report No. RM-6329-PR). Santa Monica, Ca: Rand Corporation.

McCormick E. J., & Ilgen, D. R. (1985). Industrial and organizational psychology (8th ed.). New Jersey: Prentice Hill.

Obermeyer, R. W., & Vreuls, D. (1974). Combat-ready crew performance measurement system: Phase IIIA crew performance measurement. (Report No.

AFHRL-TR-74-108(IV)). Northridge, Ca: Manned Systems Sciences, Inc.

Pierce, B. J., De Maio, J., & Eddowes, E. E. (1979). Airborne performance measurement methodology application and validation: F-4 pop-up training evaluation. (Report No. AFHRL-TR-79-7). San Antonio, Tx: Air Force Human Resources Laboratory.

Shannon, R. H., & Waag, W. L. (1973). The prediction of pilot performance in the F-4 aircraft. (Report No. NAMRL-1186). Pensacola, Fl: Naval Aerospace Medical Research Laboratory.

Shannon, R. H., (1980). The validity of task analysis information to human performance research in unusual environments. Proceedings of the Human Factors Society 24th Annual Meeting (pp. 325-329).

Shartle, C. L., & Hemphill, J. K. (1950). Aircrew effectiveness criteria. (Report No. HRRL 19). Ohio State University, Columbus Oh: Personnel Research Board Research Foundation.

Smith, J. F., & Flexman, R. E. (1972). An instructional manual for using performance record sheets designed for primary pilot training. (Report No. AFHRL/FT TR-72-11). Williams AFB, Az: Air Force Human Resources Laboratory.

SPSS, Inc. (1983). SPSS-X USER'S GUIDE. New York: McGraw-Hill.

Super, D. E., & Crites, J. O. (1962). Appraising Vocational Fitness. (rev. ed.). New York: Harper and Row Publishers.

Tactical Air Command Manual (TACM) 51-50. (1985). Langley AFB, Va: Headquarters, Tactical Air Command.

Tactical Air Command Regulation (TACR) 55-4, Volume II. (1983). Langley AFB, Va: Headquarters, Tactical Air Command.

Appendix A

Basic Attributes Test Battery

Test	Psychological Factors
Perceptual Speed	Information input efficiency
Dot Estimation	Compulsiveness vs. decisiveness
Time Sharing	Higher order tracking ability, learning rate, and time sharing ability as a function of differential load in a task involving one continuous and one discrete-events subtask
Encoding Speed	Verbal processing ability at increasing levels of information complexity
Mental Rotation	Mental-spatial transformation and classification
Item Recognition	Short term memory store, search, and compare operations
Immediate/Delayed Memory	Continuous short term memory storage and retrieval operations
Decision Making Speed	Simple choice reaction time under various conditions of information load and conditions of spatial and temporal uncertainty. Low level cognitive, and high level sensory-perceptual motor involvement
Risk-Taking	Effects of uncertainty on decision-making
Embedded Figures	Field dependence/independence

Self Crediting Word Knowledge	Self-assessment ability/self confidence
Activities Interest Inventory	Survival attitudes
Automated Aircrew Personality Profiler	Personality factors to be extracted
Two Hand Coordination and Complex Coordination (Stick Rudder)	Low to moderate order tracking and timesharing ability in pursuit and compensatory tracking tasks involving multi-axis continuous events (BAT - Version 4, 1985)

AIR RECONNAISSANCE REQUEST/TASK MESSAGE		CLASSIFICATION UNCLASSIFIED	FROM
RCVD DTG Z	INITIALS / DRAFTER / TEL NO	INFO:	TO
TRAP DTG Z	INITIALS / RELEASER / TEL NO		T.O. Z
			NLT Z
NOT FOR TRANSMISSION	AIR RECCE REQUEST/TASK NO:		MISSION NO
ORIGINATOR SERIAL NO	L		BAT 02 - PRIMARY

TYPE OF SENSOR	M-1	[X] PHOTO [ ] ELECTRONIC [ ] IR [ ] BEST SUITABLE
TECHNIQUE	M-2	[ ] VERTICAL [X] OBLIQUE [ ] PANORAMIC [ ] BEST POSSIBLE
MAP SERIES	N	IR 123 or VR 1166
TARGET LOCATION	O-1	MJ 315670
TARGET DISCRIPTION	O-2	HX 09 CAT 9B 2, 4 ROAD INTERSECTION
TARGET LOCATION	O-3	LJ 135645
TARGET DISCRIPTION	O-4	HX 18 CAT 9B 2, 4 ROAD INTERSECTION
TARGET LOCATION	O-5	KJ 506594
TARGET DISCRIPTION	O-6	HE 89 CAT 3A EL SITE
TARGET LOCATION	O-7	KK 660221
TARGET DISCRIPTION	O-8	HX 113 CAT 9B 2, 4 ROAD INTERSECTION
OBJECT AND	P-1	
RESULTS DESIRED	P-2	
PHOTO SCALE	Q	[X] BEST POSSIBLE
NO. PRINTS, PLOTS, RPTS.	R	
DELIVERY ADDRESS	S-1	
DATE TIME REQUIRED	S-2	
LTIOV	T	Z
TOT	U-1	Z
PRIORITY	U-2	[X] PRI-1 [ ] PRI-2 [ ] PRI-3
TOT SECURITY CLASS.	U-3	UNCLASSIFIED
REMARKS	U-4	
TASKING INSTRUCTIONS	V	
SORTIE NO		CLASSIFICATION UNCLASSIFIED

BERG FORM JAN 89 1

Appendix C

Basic Attributes Test Reliability Estimates

Subtest	N	Reliability
Psychomotor Tests	1237	
X-1 Axis		.97 *
Y-2 Axis		.98 *
Decision Making Speed	276	
Response Time		.87 *
Correct Responses		.82 *
Item Recognition	276	.96 **
Mental Rotation	276	.92 **
Decision Making Speed	276	
Response Time		.87**
Correct Responses		.82**

Note. * indicates that the reliability coefficient was calculated using Cronbach's coefficient alpha. ** indicates that the reliability was calculated using an odd/even split-half reliability. (Estimates of reliablity were obtained from T. R. Carretta, Air Force Human Resources Laboratory, San Antonio, Texas, March 13, 1987 and had been extracted from unpublished test reviews).

Appendix D

Instructions to Pilot Subjects

You have been selected to participate in a research program designed to add an objective measure in the selection and placement of Air Force pilots into various weapon systems. Your participation in this study will consist of flying four day low level visual reconnaissance sorties and taking the Basic Attributes Test (BAT). The purpose of these tasks is to determine if BAT test scores are related to target acquisition and flying performance.

Instructions: You will be scheduled to fly four sorties, BAT 1 thru 4, in a randomized order. When planning and flying these sorties please ensure that you fly them in the following order:

1st sortie - BAT _____

2nd sortie - BAT _____

3rd sortie - BAT _____

4th sortie - BAT _____.

Fly these sorties as combat profile reconnaissance training sorties. That is, accomplish the entire 'recce cycle'. The desired type of target coverage is specifically specified on each frag. Be sure to download each camera used to obtain photographic coverage of the assigned targets.

WHEN FLYING THESE SORTIES PLEASE STICK TO THE RULES AND REMEMBER TARGET REATTACKS ARE NOT AUTHORIZED.

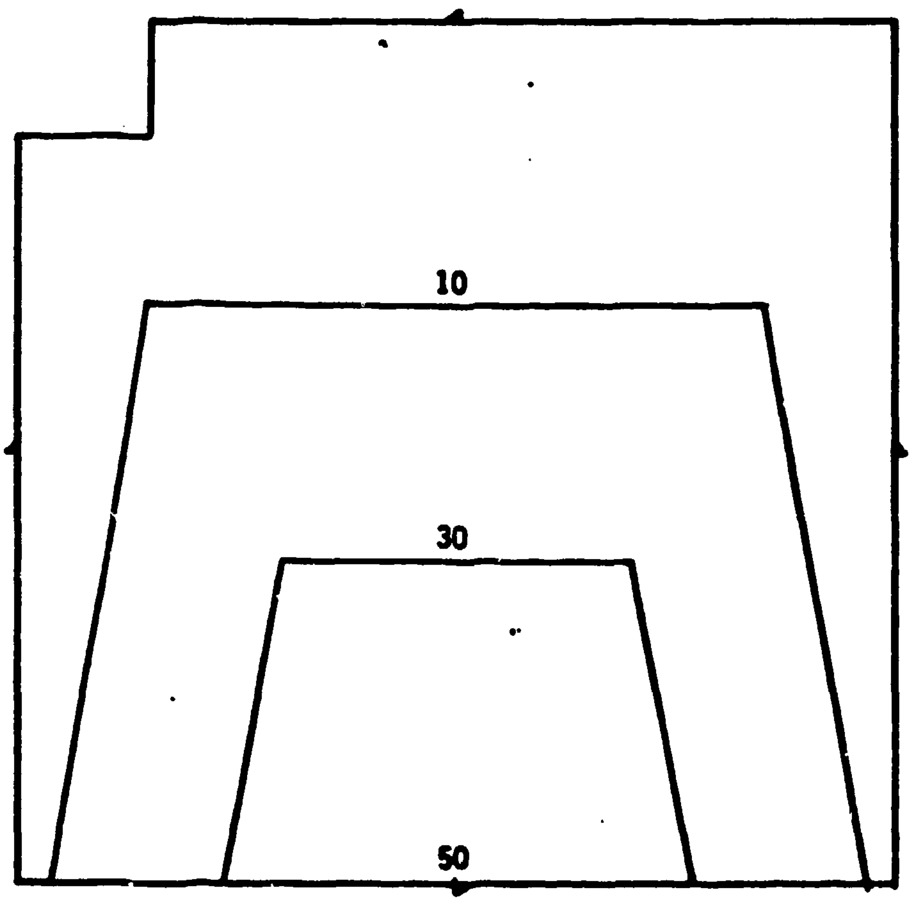
On the camera card write "BAT - HOLD". After mission debrief, pilot and WSO should proceed to the PPIF to review the mission film. When reviewing the film, identify the frame which best covers the target and mark the target with grease pencil. Remind the PI to hold the film until reviewed by the researcher.

BAT TEST: You will need to take the Basic Attributes Test (BAT) sometime during the course of this research project (prior to 31 Nov 86). The test will be explained to you prior to your taking the test. The test will last between 1 1/2 and 2 hours. When scheduled for the BAT go to bldg 1305, which is adjacent to the base small arms range. The BAT test is located to the right as you enter the building. PLEASE BE ASSURED THAT THE INFORMATION COLLECTED IN THIS STUDY WILL NOT BE USED TO AFFECT YOUR CAREER IN ANY WAY. THE INFORMATION COLLECTED CONCERNING YOU PERSONNALLY WILL NOT BE RELEASED TO ANY WING PERSONNEL NOR TO ANY AIR FORCE PERSONNEL NOT ASSOCIATED DIRECTLY WITH CONDUCTING THIS STUDY.

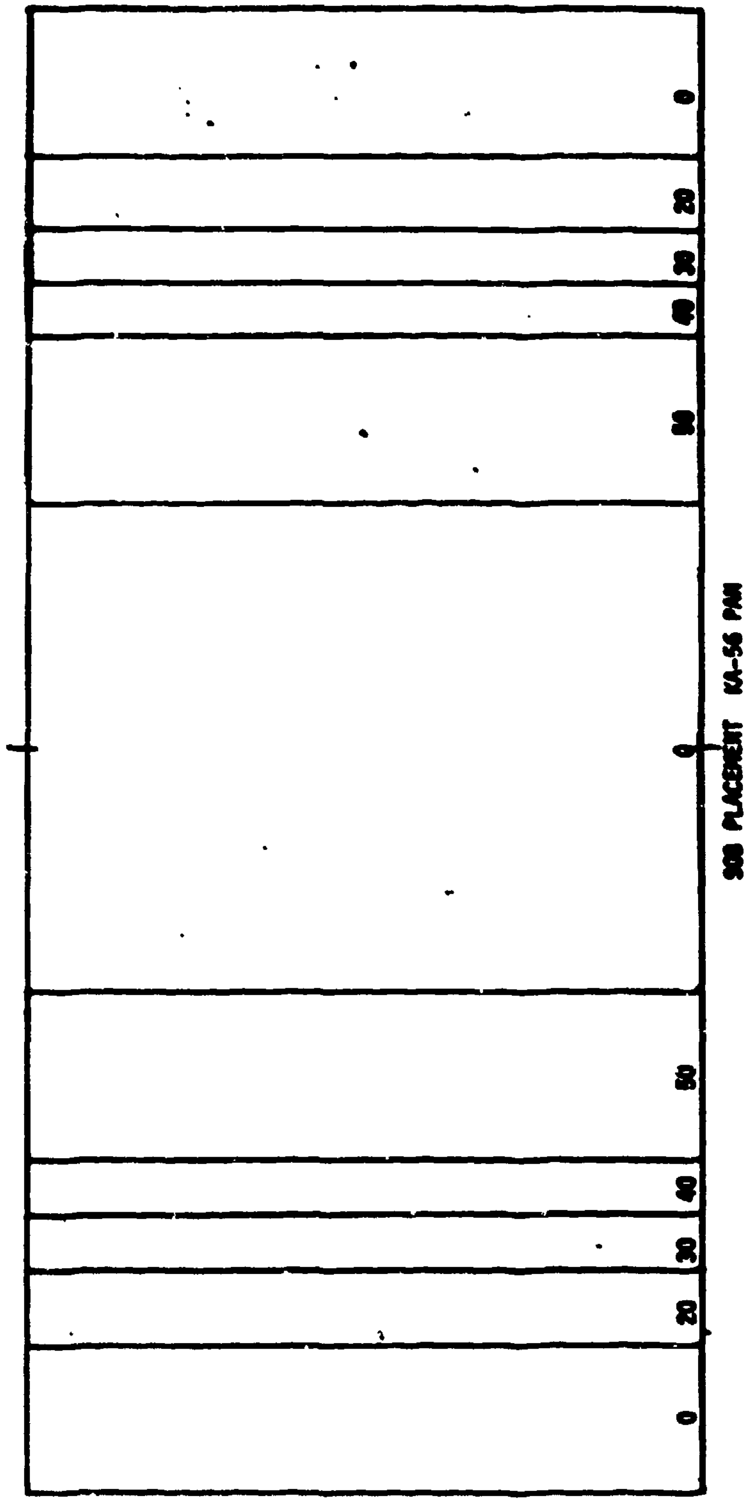
THANK YOU FOR YOUR COOPERATION

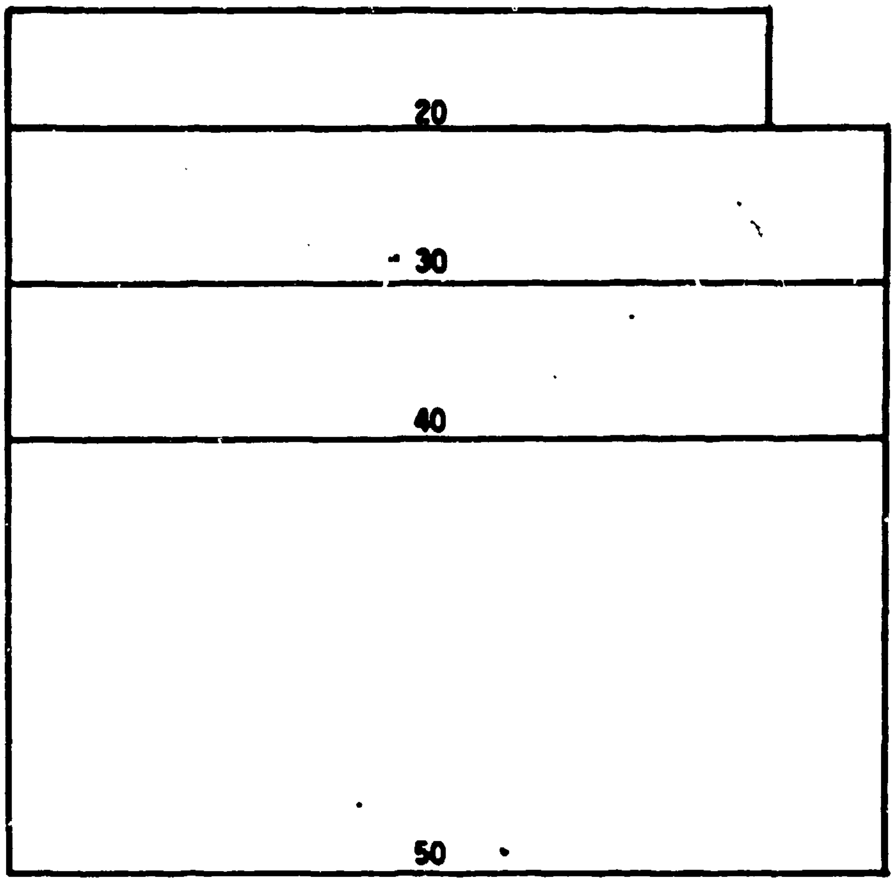
IF YOU HAVE ANY COMMENTS OR QUESTIONS

CONTACT ME: phone # 3070.



FORWARD OBLIQUE PLACEMENT KS-72/87


0
20
30
40
50
0
PLACEMENT KA-56 PAN
50
40
30
20
0



LSOB PLACEMENT KS-87

Appendix F

BAT Test-Version 5 Instructions

BASIC ATTRIBUTES SYSTEM

Welcome :

Today you will be taking a battery of experimental tests. These tests have been developed to improve the personnel selection and classification system. Your performance on this test will remain confidential and will be used for research purposes only and not for any decisions regarding your career.

It is very important, however, that you do your very best on these tests, just as if your future career depended upon the results. This is because your performance will be used to decide how to use these tests to select people for training or other special assignments.

Your performance will not affect you, but it may affect the future. Please do your best.

PRIVACY ACT STATEMENT

This information is being collected under authority of 10 USC, 8012, secretary of the Air Force, powers, duties, delegation by compensation and by executive order No. 9397, 22 Nov 1943, numbering system for federal accounts relating to individual persons.

Purpose:

The information collected will be used solely for research and development purposes. Use of the social security account number is necessary to make positive identification of the individual and records.

Routine uses:

Information provided by respondents will be treated as confidential and will be used for official research purposes only. **Individual identity will not be revealed** The research information obtained will be used to improve the utilization of personnel resources within the armed forces.

Voluntary disclosure

Cooperation and disclosure of this information is voluntary. Failure to provide information would hinder the ability of the armed forces to best utilize its personnel resources. Your cooperation in this effort is appreciated.

BIOGRAPHICAL DATA

Enter your social security number.

Enter your last name, first name, middle intial.

Enter your age.

Enter your handedness (right or left).

PSYCHOMOTOR TESTS

The first two tests in the battery evaluate basic psychomotor abilities. Each test will run for five minutes. Before each test, you will be given three minutes of practice. Do not touch any of the controls until you are told to do so.

TWO HAND COORDINATION TEST

An airplane will move in a circle on the screen. This will be your target. You control the gunsight (+) appearing on the screen with the right control stick moving the gunsight left and right, and left control stick moving the gunsight up and down.

Your task will be to move the gunsight using both hands so that it will stay as close as possible to the airplane.

COMPLEX COORDINATION (STICK AND RUDDER) TEST

Now for the second test. You will see a large cross made from dots centered on the screen. You will also see a small cross. and a small vertical bar. The small cross will wander all over the screen, the vertical bar will wander only left or right.

The right control stick will move the cross, the left control stick will move the vertical bar.

Your task is to keep the small cross centered on the dotted cross, and the vertical bar in the center of the lower dotted cross.

Press the enable bar for a demonstration of what you are trying to achieve on the screen.

To move the small cross and veritical bar, push the stick toward the position of the cross and bar. If the small cross is in the upper right portion of the screen, move the right control stick to its upper right quadrant to move the cross to the center. If the vertical bar is on the right side of the screen, move the left control stick right to move the vertical bar to the left.

Keep the small cross as close as you can to the center of the screen and at the same time keep the verical bar centered.

Grasp the right control stick with your index finger on the grey trigger, and PRESS THE ENABLE BAR to begin a three minute practice session.

PSYCHOLOGICAL and COGNITIVE TESTS

EMBEDDED FIGURES TEST

In this task we are interested in how well you can pick out simple geometric figures when they are embedded in more complex figures. A warning signal will occur prior to each trial, followed shortly be the display. The display will consist of a simple geometric figure and two complex figures labeled "Figure 1" and "Figure 2". You must decide which complex figure contains the simple figure.

Place the index finger of your left hand on the button labeled 1, and index finger of your right on the key labeled 2. When you have decided which complex figure contains the simple figure, press the corresponding button on the response pad.

MENTAL ROTATION TEST

In this task, we are concerned with your ability to handle letters in various orientations. You will see a warning marker, followed by the appearance of one letter, of which you should form a mental image to preserve its orientation. When it disappears, a second will be presented, and you must decide if it is the same as the first. If the image of the first can be slid over the second to correspond exactly, they are the same. If you decide that the images are "identical" except possibly for rotational orientation, press the "S" button for "same" on your response pad. If they are "different", like mirror images, press the "D" button on your response pad. REMEMBER to respond as quickly as possible without making mistakes.

The first 12 trails will be for practice and will not be scored. Remember to form the image, then to mentally rotate it if necessary, then decide whether or not you have an exact match. Work without errors.

ITEM RECOGNITION TEST

This task measures how quickly and accurately you can recognize items that you have just seen.

For each question ONE OR MORE digits will appear on the screen. Try to remember each of the digits.

A few seconds after they have disappeared, you will see "STAND BY" followed immediately by a single digit.

You must decide whether the digit is one of those you just memorized. If the digit is one you just memorized - press the YES button, if it is not press the NO button.

Please respond as quickly as possible without making errors. Speed is very important. YOU WILL BE MAKING YOUR RESPONSES ON THE KEY PAD.

DECISION MAKING SPEED TEST

In this task we are interested in how quickly you can recognize which of several potential signals has occurred and execute an appropriate response under a variety of conditions. During the 12 sets of trials to follow, three (3) variables are going to change:

1) TIME: During half of the trials you will know exactly when a trial is about to occur because you will see a "stand by" cue at the center of your display about 2 seconds prior to a "target" signal. During the remaining trials, once a set of trials begins, you will not be "cued", but will have to respond as quickly and accurately as you can WHENEVER a signal occurs.

2) SPATIAL LOCATION: During half the trials the signal will always occur at the center of your display; during the complementary half of the trials, signals will occur at one of four corners of your display.

3) SIGNAL SET SIZE: During one third of the trials presented, the signal will be either a "1" or a "4". Another third of the trials will present signals "1" or "2" or "3" or "4" on the display. The final third will present any one of the digits from

"1" to "8". When you see a digit appear on the screen your task is to press the corresponding digit key on the response pad as quickly and accurately as you can.

Before each set of trials you will be told: 1) whether "time" is cued (you know when) or not, 2) whether the signals always occur at display center or not, and, finally 3) whether there are two possible signals ("1" or "4") or four possible signals ("1" or "2" or "3" or "4") or all eight digits are possible signals. Your task remains the same under ALL conditions; hit the CORRECT digit on the response pad as QUICKLY as possible when you see it on the display.

For example, before a set of trials you might see:

1) Potential target digits are limited to "1" or "4".

2) The target may occur only at the center of your display.

3) You will see "stand by" followed closely by a digit on every trial (cued condition).

The above, in fact, describes the first of twelve sets of trials you are about to complete. REMEMBER this is a speed task, so work as QUICKLY as possible to hit the CORRECT response (Basic Attributes Test - Version 5, 1986).

TABLE 4

Psychomotor Tests: Pilot Performance Regression Analysis

Score	Mean	St. Dev.	Correlation with TAS	Correlation with St. Dev.
X-1 Axis	5211.08	1658.80	.040	-.148
Y-2 Axis	10713.33	3506.34	-.035	.062

Analysis of Variance: TAS

Source	SS	df	MS	F	p
Regression	2.864	2	1.432	.0272	.9731
Residual	840.528	16	52.533		
Total	843.392	18			

Multiple R = .058

R Square = .003

Standard Error = 7.248

Table 4

Analysis of Variance: SD

Source	SS	df	MS	F	p
Regression	7.048	2	3.524	.2425	.7875
Residual	232.554	16	14.535		
Total	239.602	18			

Multiple R = .172

R Square = .029

Standard Error = 3.812

TABLE 5

Decision Making Speed: Pilot Performance Regression Analysis

Score	Mean	St. Dev.	Correlation with TAS	Correlation with St. Dev.
Mean DMS - when	688.17	90.38	-.266	-.046
Mean DMS - not when	715.75	64.70	-.044	-.001
St. Dev. DMS - when	233.20	64.95	-.062	-.049
St. Dev. DMS - not when	238.14	68.16	.093	-.327
Percent Correct	98.27	1.51	.149	-.345

Analysis of Variance: TAS

Source	SS	df	MS	F	p
Regression	130.415	2	65.208	1.463	.2608
Residual	712.978	16	44.561		
Total	843.393	18			

Multiple R = .393

R Square = .155

Standard Error = 6.675

Table 5

Analysis of Variance: SD

Source	SS	df	MS	F	p
Regression	107.464	3	35.821	4.066	.0267
Residual	132.139	15	8.809		
Total	239.603	18			

Multiple R = .670

R Square = .449

Standard Error = 2.968

TABLE 6

Item Recognition: Pilot Performance Regression Analysis

			Correlation with	
Score	Mean	St. Dev.	TAS	St. Dev.
Slope	-16.23	25.92	.629	-.194
Intercept	849.77	203.40	-.376	.152
Standard Error	296.99	188.51	-.564	.205
Percent Correct	97.42	2.07	.658	-.391

Analysis of Variance: TAS

Source	SS	df	MS	F	p
Regression	596.434	2	298.217	19.321	.0001
Residual	246.960	16	15.435		
Total	843.394	18			

Multiple R = .841

R Square = .707

Standard Error = 3.929

Appendix G

Table 6

Analysis of Variance: SD

Source	SS	df	MS	F	p
Regression	36.719	1	36.719	3.077	.0974
Residual	202.883	17	11.943		
Total	239.602	18			

Multiple R = .391

R Square = .153

Standard Error = 3.455

TABLE 7

Mental Rotation: Pilot Performance Regression Analysis

			Correlation with	
Score	Mean	St. Dev.	TAS	St. Dev.
Percent Correct	91.61	11.04	-.019	.340
Mean Response Time	1173.69	372.58	-.626	.048
Standard Deviation	466.18	293.77	-.411	.028

Analysis of Variance: TAS

Source	SS	df	MS	F	p
Regression	402.002	2	201.001	7.286	.0056
Residual	441.391	16	27.587		
Total	843.393	18			

Multiple R = .690

R Square = .477

Standard Error = 5.252

Appendix G

Table 7

Analysis of Variance: SD

Source	SS	df	MS	F	p
Regression	37.259	2	18.629	1.473	.2587
Residual	202.343	16	12.646		
Total	239.602	18			

Multiple R = .394

R Square = .155

Standard Error = 3.556

TABLE 8

Embedded Figures: Pilot Performance Regression Analysis

			Correlation with	
Score	Mean	St. Dev.	TAS	St. Dev.
Percent Correct	62.79	13.14	-.224	.450
Mean Response Time	14614.70	5365.57	-.419	.389

Analysis of Variance: TAS

Source	SS	df	MS	F	p
Regression	148.037	1	148.037	3.619	.0742
Residual	695.356	17	40.903		
Total	843.393	18			

Multiple R = .419

R Square = .176

Standard Error = 6.396

Appendix G

Table 8

Analysis of Variance: SD

Source	SS	df	MS	F	p
Regression	48.418	1	48.418	4.305	.0535
Residual	191.184	17	11.246		
Total	239.602	18			

Multiple R = .449

R Square = .202

Standard Error = 3.353

Appendix H

List of Abbreviations

ADAS	Aux Data Annotation System
AFHRL	Air Force Human Resources Laboratory
AFMPC	Air Force Military Personnel Center
AFOQT	Air Forcer Officer Qualification Test
ATC	Air Training Command
ATO	Air Reconnaissance Request/Task Message
ATRB	Advanced Training Recommendation Board
BAT	Basic Attributes Test
CRT	Cathode Ray Tube
DMS	Decision Making Speed Test
DCO	Deputy Commander of Operations
EEI	Essential Elements of Information
EMB	Embedded Figures Test
FAA	Federal Aviation Administration
FAIP	First Assignment Instructor Pilot
FAR	Fighter-Attack-Reconnaissance
FEBA	Forward Edge of Battle Area
FOB	Forward Oblique Photography
FSP	Flight Screening Program
GCC	Graduated Combat Capability
IP	Instructor Pilot
ITM	Item Recognition Test

LOWAT	Low Altitude Awareness Training
MAJCOM	Major Command
MR	Mission Ready
MRT	Mental Rotation Test
ORI	Operational Readiness Inspection
OT & E	Operational Test and Evaluation
PCSM	Pilot Candidate Selection Method
PI	Photographic Interpreter
PPIF	Photographic Processing and Interpretation Facility
PS 2	Complex Coordination Test (Stick and Rudder)
RAM	Reconnaissance Air Meet
R&D	Research and Development
SEFE	Standardization Evaluation Flight Examiner
SOB	Side Oblique Photography
TAC	Tactical Air Command
TAS	Target Acquisition Score
TRS	Tactical Reconnaissance Squadron
TRW	Tactical Reconnaissance Wing
TTB	Tanker-Transport-Bomber
2-HN	Two-Hand Coordination Test
UPT	Undergraduate Pilot Training
USAF	United States Air Force
USN	United States Navy
WSO	Weapon Systems Officer

VITA

CENSUS: Timothy G. Kinney was born on April 13,1951 in Denver, Colorado. His parents are Mr. and Mrs. William O. Kinney. He is married and has three children.

TRAINING: Timothy G. Kinney graduated from Golden High School, Golden, Colorado, June 1969. He receivedhis Bachelor of Arts degree from the University of Northern Colorado of Greeley, Colorado, in 1973. He received his Master of Arts degree from Houston Baptist University of Houston, Texas, in 1986. He has done graduate work in 1985, 1986, and 1987 at St. Mary's University.

EXPERIENCE: Since 1973 he has served as an officer in the United States Air Force. From 1973-1978 he flew tactical reconnaissance aircraft at Kadena Air Base, Okinawa, Japan; from 1978-1982 he taught Undergraduate Pilot Training at Laughlin Air Force Base, Del Rio, Texas; from 1982-1985 he instructed operational pilots in tactical reconnaissance aircraft and developed continuation and upgrade flying training programs at Bergstrom Air Force Base, Austin, Texas.

PERMANENT ADDRESS: 814 Braun Court
Golden
Colorado

TYPIST: Timothy G. Kinney