AFHRL-TP-87-20

AIR FORCE

AD-A188 503

HUMAN RESOURCES

BASIC ATTRIBUTES TEST (BAT) SYSTEM: A PRELIMINARY EVALUATION

Thomas R. Carretta

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

November 1987
Interim Technical Paper for Period June 1963 - June 1986

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0180

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TP-87-20	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Manpower and Personnel Division 6b. OFFICE SYMBOL (If applicable) AFHRL/MOEA	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (If applicable) HQ AFHRL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 62703F; PROJECT NO. 7719; TASK NO 18; WORK UNIT ACCESSION NO. 45

11. TITLE (Include Security Classification)
Basic Attributes Test (BAT) System: A Preliminary Evaluation

12. PERSONAL AUTHOR(S)
Carretta, T.R.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM Jun 83 TO Jun 86	November 1987	34

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
05	09	
05	08	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Air Force Officer Qualifying Test (AFOQT), information processing, ~~Basic Attributes Test (BAT)~~ pilot selection, ~~cognitive abilities~~ Undergraduate Pilot Training (UPT)

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

Recent efforts to reduce attrition rates in Air Force Undergraduate Pilot Training (UPT) have resulted in the development of an experimental computer-administered test battery, the Basic Attributes Test (BAT) system. Included in the battery are several tests which measure information processing efficiency and speed that were identified in previous research as being related to pilot performance, particularly with regard to fast jet fighter aircraft. This paper evaluated three subtests used to assess cognitive abilities: Digit Memory (information input efficiency), Decision-Making Speed (low-level cognitive and high-level sensory perceptual-motor involvement), and Item Recognition (short-term memory storage, search and comparison operations). Each of the subtests was evaluated in terms of its ability to predict various flight performance measures and final training outcome. Of particular interest was the potential of the cognitive subtests to increase the validity of current selection procedures that rely mainly on paper-and-pencil measures. In addition, an integrated model containing both the current selection measure and experimental subtests did demonstrate significant relationships with several other performance measures including recommendations for fighter assignments after training. Keywords:

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office	22b. TELEPHONE (Include Area Code) (512) 536-3877 — 22c. OFFICE SYMBOL AFHRL/TSR

DD Form 1473, JUN 86 — *Previous editions are obsolete.*

AFHRL Technical Paper 87-20

November 1987

BASIC ATTRIBUTES TEST (BAT) SYSTEM: A PRELIMINARY EVALUATION

Thomas R. Carretta

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed and submitted for publication by

Lloyd D. Burtch
Chief, Cognitive Skills Assessment Branch

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

High rates of attrition among students in Undergraduate Pilot Training (UPT) are a major concern for the United States Air Force. Recent research and development efforts at the Air Force Human Resources Laboratory have attempted to reduce attrition rates by improving the method by which pilot candidates are selected. Currently, UPT students are chosen primarily on the basis of their scores on the Pilot composite of the Air Force Officer Qualifying Test (AFOQT). The present effort sought to determine the extent to which scores on three cognitive/perceptual subtests from an experimental test battery, known as the Basic Attributes Tests (BAT), added to the validity provided by the AFOQT Pilot composite score.

Scores from the three cognitive/perceptual tests--Digit Memory (information input efficiency), Decision-Making Speed (choice reaction time), and Item Recognition (short-term memory storage, search and comparison operations)--did not add significantly to the prediction of graduation or failure. However, the experimental subtests did demonstrate significant relationships with several other performance measures including recommendations for fighter or non-fighter assignments following UPT.

PREFACE

This work was completed under Work Unit 77191845 in support of a Request for Personnel Research (RPR 78-11, Selection for Pilot Training) submitted by Air Training Command training program managers.

This paper is intended to serve as an interim report regarding three of the cognitive/perceptual tests of the Basic Attributes Test (BAT) battery.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

BASIC ATTRIBUTES TEST (BAT) SYSTEM:
A PRELIMINARY EVALUATION

I. INTRODUCTION

Since World War I, the United States military has taken an active interest in developing tests to predict success in pilot training. Throughout World War II, tests of psychomotor ability, called apparatus tests, were commonly used in the selection and classification of aircrew personnel. Typically, these tests involved some form of rotary pursuit or compensatory tracking task using a mechanical or electrical device. These apparatus tests generally exhibited validities ranging from .20 to .40. A number of paper-and-pencil tests were also used with aircrew personnel, but given less consideration than the apparatus tests. Such tests included measures of general intelligence, mechanical comprehension, perception, vocabulary, and reading comprehension (North & Griffin, 1977).

Despite the demonstrated validities of psychomotor tests and their proven utility in reducing attrition in pilot training, the Air Force discontinued their use in 1955, because of problems with unreliable equipment and an administrative shift toward decentralized testing procedures. From then until now, pilot candidates have been chosen primarily on the basis of the Air Force Officer Qualifying Test (AFOQT), a paper-and-pencil test; physiological fitness; and previous flying experience (Bordelon & Kantor, 1986).

The Pilot composite score of the AFOQT is based on subtests such as verbal analogies, mechanical comprehension, scale reading, instrument comprehension, table reading, and aviation information. This composite score has demonstrated a reliable correlation with pilot training outcome in a number of studies (e.g., Acosta, 1985; Bordelon & Kantor, 1986; Hunter & Thompson, 1978; McGrevy & Valentine, 1974; Miller, 1966). However, beginning in the 1960s, concern with attrition rates in pilot training, along with the development of computer technology, produced a renewed interest in the utility of psychomotor testing (Long & Varney, 1975). Based upon studies that demonstrated the reliability and validity of psychomotor testing (e.g., Hunter & Thompson, 1978; McGrevy & Valentine, 1974), the Air Force initiated a project in 1981 to develop a computer-administered test battery for pilot selection and classification. The resulting product is the Basic Attributes Test (BAT) System, or BAT (Kantor & Bordelon, 1985).

The BAT consists of a number of tests designed to measure psychomotor aptitude, and perceptual and cognitive processes, as well as personality and attitudinal characteristics. The BAT tests were chosen on the basis of their being measures of psychological dimensions associated with pilot performance in previous research (e.g., Hunter, 1975; Hunter, Maurelli, & Thompson, 1977; McLaurin, 1973; Passey & McLaurin, 1966). Some of these tests were derived from earlier test batteries; others were adapted from tasks used in mainstream cognitive psychological research as measures of information processing proficiency, an ability identified as critical to pilot functioning in high-speed jet fighters (Imhoff & Levine, 1981).

This paper will focus on three of the cognitive perceptual tests: Digit Memory, Decision-Making Speed, and Item Recognition. Digit Memory was chosen to examine individual differences in short-term memory and sensory storage. Decision-Making Speed was adapted from a task used during World War II called Discrimination Reaction Time (Passey & McLaurin, 1966). Previous research indicates that this task includes three components: a perceptual response, a visualization response, and reaction time (Adams, 1957; Fleishman & Hempel, 1956). Finally, the third test, Item Recognition, was developed by Sternberg (1966) in order to study retrieval from short-term memory.

The general hypotheses guiding this effort parallel those used in previous research (e.g., Bordelon & Kantor, 1986; Kantor & Bordelon, 1985) that validated the psychomotor tests which form part of the BAT. That is, individual differences in performance on the tests should predict Undergraduate Pilot Training (UPT) performance and also should add significantly to the validity of the paper-and-pencil selection test, AFOQT, currently used for predicting training success. In particular, it is hypothesized that subjects with quicker reaction times and more efficient memories will be more likely to succeed in training. Furthermore, these differences should be better reflected in flight performance scores (check flight grades), which are more numerous and have a broader range than the dichotomous final training outcome measure (pass/fail). Moreover, the fact that the pass/fail rate is unevenly distributed (80% pass versus 20% fail) also makes it a less sensitive criterion.

It is also hypothesized that scores from the apparatus tests, taken together with scores from the AFOQT, should demonstrate stronger relationships with performance outcomes than does the AFOQT alone. That is, the apparatus tests must add to the ability to predict performance outcomes or there is no reason to go to the cost and effort to replace the current test system. On the other hand, if the apparatus tests do add to the validity of the test procedure, this is also evidence that the apparatus tests are measuring unique factors unrelated to those associated with current paper-and-pencil testing.

In addition to its concern with training attrition, the Air Force is interested in classifying pilots for advanced training as early in their careers as possible. Normally, pilots are recommended for one of two advanced training tracks at the end of UPT, which currently involves about 175 hours of flying time. On the basis of an evaluation by an Advanced Training Recommendation Board (ATRB), pilots go on to training for a Fighter-Attack-Reconnaissance (FAR) assignment or a Tanker-Transport-Bomber (TTB) assignment. In general, the students who perform best in UPT are selected for fast-jet training (i.e., FAR). Thus, it is expected that FAR-recommended pilots will demonstrate better scores on cognitive/perceptual tests than will the TTB-recommended pilots. The demonstration of a significant relationship would provide the Air Force with a tracking procedure that could take place early in UPT, resulting in more efficient and cost-effective training.

II. METHOD

Subjects

The subjects in the present effort were 1,273 Air Force officer candidates targeted for UPT. They were tested on the BAT system prior to their entry into UPT. The exact number of subjects varied from test to test, as the various tests comprising the BAT battery were not developed all at the same time. Further, UPT outcome measures (pass/fail outcome, ATRB ratings, check flight scores) were available for only a portion of the subjects, as many of the subjects had not yet completed UPT. Only subjects that had scores on all three tests and the AFOQT were included in the regression analyses that predicted performance on the UPT outcome measures (UPT pass/fail outcome, N = 512; ATRB rating, N = 410; check flight scores [see below], N = 115). A listing of the number of subjects available for each is presented in Table 1.

Procedure

Prior to entry into flying training, each subject was tested on the AFOQT. This test provided five composite scores based on a number of subtests: Verbal, Quantitative, Academic (verbal and quantitative combined), Navigator-Technical, and Pilot. Only the Pilot composite was used in this analysis, as that is the test score used in the operational selection of candidates for UPT. A breakdown of the subtests that contribute to each composite score is provided in Table 2.

Table 1. **Number of Subjects Available**

Test	Test only	UPT outcome (pass/fail)	ATRB (TTB/FAR)	Check flights
Digit Memory	1,273	512	410	115
Decision-Making Speed	1,067	512	410	115
Item Recognition	1,071	512	410	115

Table 2. **Construction of AFOQT Composite Scores**

AFOQT tests	Verbal	Quantitative	Academic Aptitude	Navigator-Technical	Pilot
Verbal Analogies	X		X		X
Arithmetic Reasoning		X	X	X	
Reading Comprehension	X		X		
Data Interpretation		X	X	X	
Word Knowledge	X		X		
Math Knowledge		X	X	X	
Mechanical Comprehension				X	X
Electrical Maze				X	X
Scale Reading				X	X
Instrument Comprehension					X
Block Counting				X	X
Table Reading				X	X
Aviation Information					X
Rotated Blocks				X	
General Science				X	
Hidden Figures				X	

Subjects also were tested with the BAT apparatus. The BAT apparatus consists of a super-microcomputer built within a self-contained unit with a glare shield and side panels designed to ensure consistency of testing conditions across subjects and test sessions. The subject responds to the various tests using, in combination or individually, a two-axis joystick on the right side of the apparatus, a single-axis joystick on the left side, and a keypad in the center of the test unit. The keypad includes the numbers 0 to 9, an ENABLE key in the center, and a bottom row with YES and NO keys and two others labelled S/L (for same/left responses) and D/R (for different/right responses). Figure 1 shows a typical test station.

The test battery as used in the present effort consisted of 15 tests lasting about 4 hours. After a test administrator initialized the system, the test session was self-paced by the subject. The test session included programmed breaks between tests, to avoid problems with mental and physical fatigue. The specific tests examined in this study are discussed below.

Digit Memory

The subject was presented with a simultaneous sequence of four digits in random order and given instructions to cancel the display and then respond as quickly as possible by pressing the

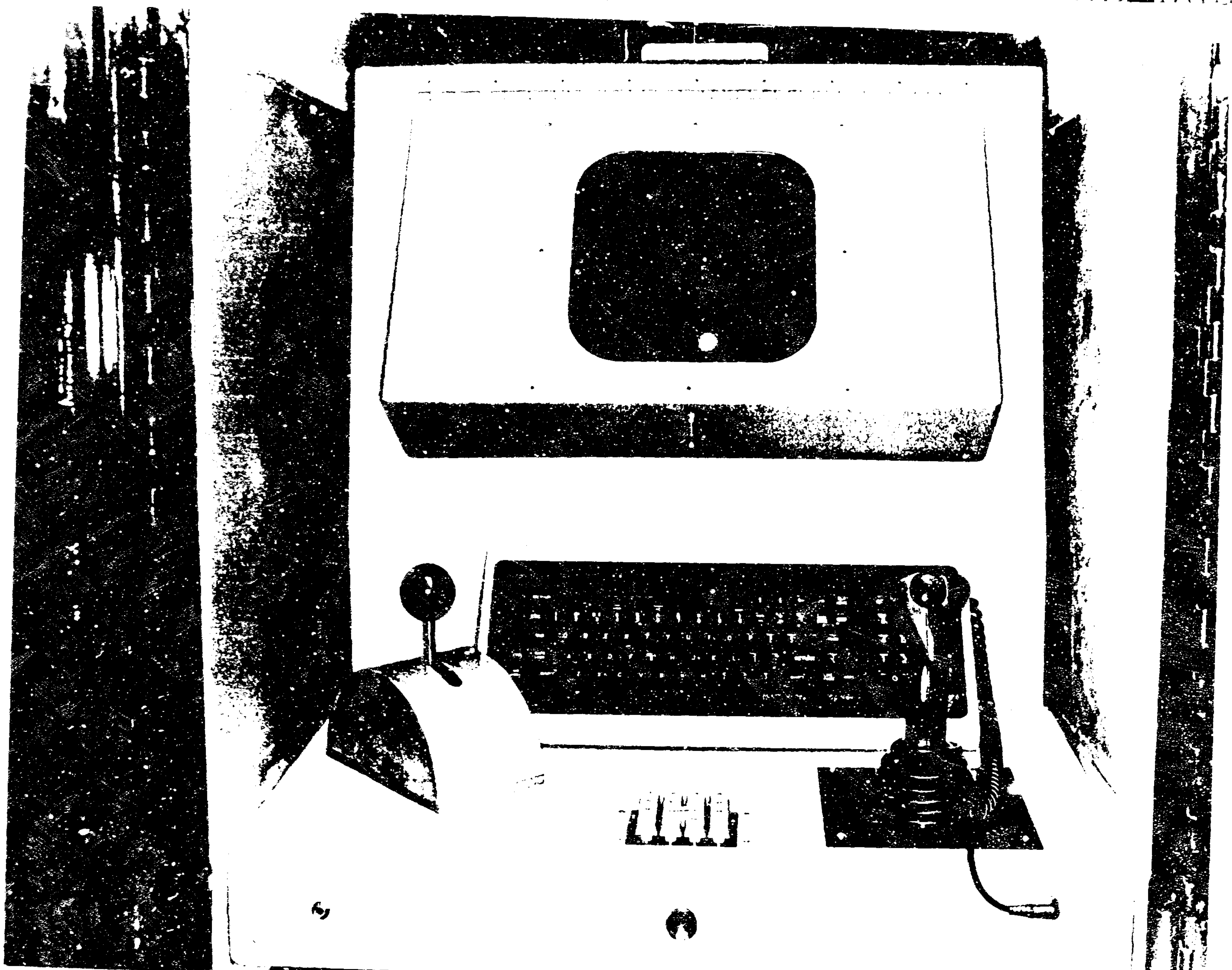

Figure 1. PORTA-BAT Test Station.

buttons on the data entry keypad in the same order as the presented digits. In addition to recording the accuracy of response (correct/incorrect) and overall response time, a measure of perceptual speed was taken as the amount of time it took the subject to identify the sequence of digits prior to actually entering a response. Key-in speed was the amount of time it took the subject to type the response sequence on the data entry keypad after the sequence of digits had been identified. There were 20 trials lasting approximately 5 minutes.

Decision-Making Speed

This test measured simple choice reaction time under varying degrees of information load and spatial and temporal uncertainty, as well as low-level cognitive and high-level sensory-perceptual motor involvement. The subject was presented with one of several alternative digits and required to respond by keying the matching digit as quickly as possible. The critical manipulation in this test was the amount of uncertainty that had to be resolved in order to make the response decision. When more alternative signals were potentially available for presentation, greater uncertainty existed and the decision should have been made more slowly.

The Decision-Making Speed test was comprised of four subtasks, each with three parts. In subtask one, the subject knew both where and when a signal was to occur; in subtask two, the subject knew where but not when; in subtask three, when but not where; and finally, in subtask four, the subject knew neither where nor when. Within each subtask, there were three parts. In part one, two potential signals and responses were defined. There were four potential signals and responses in part two, and eight potential signals and responses in part three. Therefore, degree of uncertainty of the signal was manipulated in three ways--location of occurrence, time of occurrence, and range of signal/response values. There were 12 trials within each part of each subtask, resulting in 144 trials (3x4x12) lasting altogether about 20 minutes.

Item Recognition

In this test, the subject was presented with a string of one to six digits on the screen. The string was removed and then followed, after a brief delay, by a single digit. The subject was instructed to remember the initial string of digits, then decide whether the single digit was one of those that had been presented in the initial string. The subject was instructed to press a keypad button marked YES if the single digit was in the initial string, or another marked NO if it was not. As with the Digit Memory and Decision-Making Speed subtests, the subject was urged to work as quickly and accurately as possible. There were two blocks of 24 trials each, and the entire test lasted about 20 minutes.

UPT Performance Criteria

UPT final training outcome was scored as a dichotomous variable, with pass = 1 and fail = 0. The ATRB ratings for advanced training leading to an assignment either as a TTB pilot or a FAR pilot were also scored in this manner, with TTB = 0 and FAR = 1. Final training outcome and ATRB recommendation were determined, in part, by a subject's performance on six check flights during UPT. A check flight involved an in-flight performance evaluation by an Instructor Pilot other than one with whom the student normally flew. Three of the check flights took place in a Cessna-built T-37, a low-performance jet trainer; and three took place in a Northrop T-38, a high-performance, supersonic jet trainer. The T-37 check flights included: mid-phase contact, a subject's first check flight; contact, in which the subject's ability to fly maneuvers and aerobatics by visual cues outside the plane was evaluated; and instrument, in which the subject had to fly maneuvers by reference to the display on the cockpit instrument panel. The T-38 check flights, in addition to contact and instrument, included evaluation of the subject's ability to fly in formation with other aircraft. Each subject received a check flight grade (1-unsatisfactory, 2-fair, 3-good, and 4-excellent) and an overall percentage score for all flights that were completed during training.

III. RESULTS AND DISCUSSION

AFOQT Pilot Composite

A regression equation that used only the AFOQT Pilot composite was found to be significantly related to both UPT pass/fail outcome ($r = .106$, $p \leq .05$) and ATRB rating TTB/FAR ($r = .136$, $p \leq .01$), but was statistically unrelated to check flight performance. A summary of these regression analyses is provided in Table 3.

Table 3. **AFOQT-Pilot Composite: Summary of UPT Outcome Regression Analyses**

Outcome measure	N	Mean	SD	Correlation with outcome AFOQT-Pilot
UPT pass/fail	512	0.801	0.400	.106*
ATRB TTB/FAR	410	0.549	0.498	.136**
T-37 midphase grade	115	2.56	1.19	.159
T-37 contact grade	114	2.96	0.94	.012
T-37 instrument grade	112	2.94	1.05	.160
T-38 contact grade	102	2.62	1.14	.009
T-38 instrument grade	100	2.89	1.11	.040
T-38 formation grade	98	2.87	1.05	.059
T-37 midphase percentage	115	85.48	8.36	.059
T-37 contact percentage	114	91.22	5.42	.120
T-37 instrument percentage	112	91.66	7.57	.070
T-38 contact percentage	102	91.53	5.76	.063
T-38 instrument percentage	100	92.27	6.13	.010
T-38 formation percentage	98	92.80	6.83	.071

*$p \leq .05$.
**$p \leq .01$.

Digit Memory

Descriptive Measures

Response measures were recorded for 1,273 subjects. Each trial provided an indication of the accuracy of the response (correct/incorrect), perceptual speed (RT_1), and key-in speed (RT_2). Responses on each of these measures were fairly consistent across the 20 trials. Percent correct ranged between 81% and 95% over the 20 trials. This was encouraging, as the primary variable of interest in tests of this type is response time only when correct responses are made. Average perceptual speed (RT_1) and key-in speed (RT_2) also were consistent across trials. The distributions for both response time measures were positively skewed. This was the result of a few extremely long response times. Table A-1 provides a summary of these measures.

Response times exceeding 7,500 milliseconds were treated as outliers. They were recoded to equal 7,500 milliseconds in order to reduce the effects of careless responding and develop a more reliable measure to use in subsequent analyses. These constituted less than 1% of all responses but significantly distorted the means and standard deviations.

Factor Structure

The most conceptually important measure provided by this test was perceptual speed (RT_1). A factor analysis was performed on the 20 trials for this measure in order to evaluate its internal consistency. There were only 1,067 subjects for this analysis due to some missing data on the last two trials. As can be seen in Table A-2 in the Appendix, inter-item correlations ranged between .211 and .625, with the strength of the correlations generally increasing after the first few trials. The low correlations on the early trials were attributed to the relatively large amount of variability for the response times on these trials. After the first few "practice" trials, the subjects' responses became more stable, thus increasing the strength of the correlations.

The goal of factor analysis is to identify one or more underlying dimensions (i.e., factors) that a group of variables is measuring. The perceptual speed scores were expected to yield one general underlying dimension. Two factors accounting for 52.9% of the total variance emerged from the factor analysis. The method used retained only those factors that had an eigenvalue greater than or equal to 1.0. After Varimax rotation, the principal factor accounted for 93.6% of the explained variance, indicating that the perceptual speed measure was internally consistent. A summary of the factor analysis is presented in Table A-3.

As the response measures appeared to be internally consistent, data reduction techniques were used to produce a few reliable measures for the regression analyses. First, based on techniques typically used on tests such as these, only data for correct responses were retained for further analyses. Second, Trials 1 through 5 were treated as practice trials and eliminated from further analyses, because responses on these early trials were relatively unstable and unreliable. Finally, scores for Trials 6 through 20 were reduced to a single score. Summary statistics were generated for percent correct, perceptual speed (RT_1), and key-in speed (RT_2) to be used in the regression analyses.

Inferential Measures

UPT Final Outcome/ATRB Rating. Once a set of reliable measures was identified, the next step was to examine their predictive validity with regard to UPT performance criteria (UPT final outcome, ATRB rating, check flight grades, and check flight percentage scores). Before proceeding, it should be noted that zero-order correlations between variables in the regression model and the outcome measures were tested only if the overall model showed significance.

The first set of regression analyses used UPT final outcome (pass/fail) as the performance criterion. A regression equation that used average perceptual speed (RT_1), standard deviation of perceptual speed, and percent correct for Trials 6 through 20 was unable to significantly predict UPT final outcome (multiple R = .069, n.s.). Similar results were obtained when average key-in speed, standard deviation of key-in speed, and percent correct were used as predictors of UPT final outcome (multiple R = .085, n.s.). Tables 4 and 5 provide summaries of these regression analyses.

Table 4. Digit Memory (Perceptual Speed): Summary of UPT Outcome Regression Analyses

Outcome measure	N	Mean	SD	Correlation with outcome PS-Mean	PS-SD	% Correct	Mult. R
UPT pass/fail	512	0.801	0.400	-.029	-.016	.060	.069
ATRB TTB/FAR	410	0.549	0.498	-.131	-.109	.102*	.166**
T-37 midphase grade	115	2.56	1.19	-.138	-.051	-.043	.145
T-37 contact grade	114	2.96	0.94	-.157	-.076	.036	.167
T-37 instrument grade	112	2.94	1.05	-.067	-.077	-.095	.124
T-38 contact grade	102	2.62	1.14	-.101	-.007	.059	.140
T-38 instrument grade	100	2.89	1.11	-.067	.006	-.162	.177
T-38 formation grade	98	2.87	1.05	-.237	-.299*	-.129	.330*
T-37 midphase percentage	115	85.48	8.36	-.224	-.083	-.050	.232
T-37 contact percentage	114	91.22	5.42	-.171	-.112	.064	.190
T-37 instrument percentage	112	91.66	7.57	-.051	-.033	-.122	.128
T-38 contact percentage	102	91.53	5.76	-.033	-.005	.020	.045
T-38 instrument percentage	100	92.27	6.13	-.062	-.008	-.030	.075
T-38 formation percentage	98	92.80	6.83	-.191	-.166	-.073	.209

*$p \leq .05$.
**$p \leq .01$.

Table 5. Digit Memory (Key-in Speed): Summary of UPT Outcome Regression Analyses

Outcome measure	N	Mean	SD	Correlation with outcome KS-Mean	KS-SD	% Correct	Mult. R
UPT pass/fail	512	0.801	0.400	-.014	-.054	.060	.085
ATRB TTB/FAR	410	0.549	0.498	-.042	-.089	.102	.132
T-37 midphase grade	115	2.56	1.19	.008	-.034	-.043	.060
T-37 contact grade	114	2.96	0.94	.144	-.079	.036	.231
T-37 instrument grade	112	2.94	0.85	.055	-.106	-.095	.151
T-38 contact grade	102	2.62	1.14	-.011	.020	.059	.065
T-38 instrument grade	100	2.89	1.11	-.167	-.102	-.162	.247
T-38 formation grade	98	2.87	1.05	-.008	-.143	-.129	.203
T-37 midphase percentage	115	85.48	8.36	.025	-.071	-.050	.109
T-37 contact percentage	114	91.22	5.42	-.027	-.218	.064	.247
T-37 instrument percentage	112	91.66	7.57	-.117	-.092	-.122	.182
T-38 contact percentage	102	91.53	5.76	-.031	-.042	.020	.047
T-38 instrument percentage	100	92.27	6.13	-.186	-.189	-.030	.222
T-38 formation percentage	98	92.80	6.83	.079	-.032	-.073	.129

The three perceptual speed measures (average perceptual speed, standard deviation of perceptual speed, and percent correct) were related significantly to ATRB rating (multiple R = .166, $p \leq .01$). Subjects who made quick, consistent, and accurate responses were more likely to receive a FAR rating. Although the direction of the correlations for the key-in speed measures were in the expected direction, they were not related significantly to ATRB rating (multiple R = .132, $p \leq .069$).

Check Flight Scores. Check flight grades (1, 2, 3, or 4) and check flight percentage scores were available for only 115 of the 512 subjects that had UPT final outcome scores.

Separate regression analyses were performed using average perceptual speed (RT_1), standard deviation for perceptual speed, and percent correct to predict each of the check flight grades and percentage grades. Results of the regression analyses indicated that the perceptual speed measures were predictive of performance only on the T-38 formation check flight grade (multiple R = .330) at the .05 level of significance. The T-38 formation flight is the final training flight during UPT. Performance on this flight was better for subjects who made quick and consistent decisions. Although the perceptual speed measures were not related significantly to performance on the other check flights, the zero-order correlations between the predictor variables and outcome measures were in the expected direction.

Similar but non-significant results were obtained when key-in speed was used instead of perceptual speed. A brief summary of these analyses is provided in Tables 4 and 5.

Decision-Making Speed

Descriptive Measures

Response measures (correct/incorrect and reaction time) were recorded for 1,071 subjects on each of the 144 trials. The data from each 12-trial set for each subject were summarized as a single score. This data reduction technique was used to make the data more manageable and to create a relatively small set of stable predictor variables (12 means instead of 144 scores). The resulting means and standard deviations are presented in Table A-4 in the Appendix.

As can be seen in Table A-4, the response times for subtask one (subject knew both where and when the signal would occur) were more variable than those in later subtasks. During these early trials, the subjects were unfamiliar with the test procedure and were less consistent in their response times. As a result, the trials from subtask one were treated as "practice trials" and eliminated from further analyses.

Examination of the cell means revealed that the location manipulation (subject did or did not know where the signal was to occur) did not significantly affect reaction time. As a result, the data were further collapsed into six cells: two subtasks (where the subject did or did not know when the signal would occur) with three parts in each (2 versus 4 versus 8 potential signals and responses).

Factor Structure

Decision-making speed under varying levels of uncertainty was the most conceptually important measure provided by this test. However, the consistency of decision-making speed and accuracy of responses under varying levels of uncertainty also are important determinants of decision-making ability. In order to evaluate the interrelationships among these variables, a factor analysis was performed using average decision-making speed, standard deviation of average decision-making speed, and percent correct for each of the six number of signals/responses (2 or 4 or 8) by time of occurrence (subject did or did not know when the signals would occur) combinations. Scores were available for 1,071 subjects.

The six average decision-making speeds correlated strongly with one another ($.419 \leq r \leq .684$) and with their respective standard deviations ($.567 \leq r \leq .711$), but were related only weakly to percent correct ($.058 \leq r \leq .216$). The six standard deviations were interrelated moderately, as were the six percent-correct measures. The standard deviations and percent-correct measures were not related statistically to each other. The inter-item correlations are provided in Table A-5 in the Appendix.

The factor analysis resulted in the identification of five initial factors that accounted for 62.0% of the total variance of the 18 measures. The number of factors was not surprising as the 18 measures included three distinct types of scores (average response times, standard deviations, and percent correct) obtained under varying conditions. After Varimax rotation, the principal factor accounted for 56.9% of the explained variance. This factor can be interpreted as a "general response latency" factor, as the average decision making speeds and standard deviations in all three signals/responses conditions where the subject knew when the signal would occur loaded heavily on this factor. Factors 2, 4, and 5 were defined primarily by the average decision-making speed and the standard deviation of decision-making speed for the separate signals/responses conditions when the time of occurrence of the signals was unknown. Finally, factor 3 was defined by the six percent-correct measures and can be thought of as an "accuracy index." Table A-6 provides a summary of the factor analysis.

These results suggested that the degree of uncertainty of signal/response was most important when the time of occurrence was unknown. A model of decision-making ability should consider changes in ability under varying levels of uncertainty in addition to a general accuracy of response variable.

The data were collapsed across the uncertainty of signal/response manipulation in order to produce a small set of reliable predictors to be used in the regression analyses. These included average decision-making speed and its standard deviation for the "when" and "not when" conditions, and overall percent correct. These measures were chosen to represent three important features of decision-making ability; namely, speed, consistency, and accuracy of responses under differing levels of uncertainty.

Inferential Measures

UPT Final Outcome/ATRB Rating. The next step was to evaluate the predictive utility of these measures against UPT final outcome (pass/fail), ATRB rating, and the six check flight grades and percentage scores.

As with Digit Memory, the Decision-Making Speed measures were not related significantly to UPT final outcome (multiple $R = .107$, n.s.) but were related to ATRB rating (multiple $R = .229$, $p \leq .001$). A summary of the Decision-Making Speed regression analyses is presented in Table 6.

Check Flight Scores. As previously noted, check flight scores were available for only 115 of the 512 subjects with a UPT final outcome score. The multiple regression analyses indicated that the five Decision-Making Speed performance variables were helpful in predicting performance on the later check flight percentage scores (multiple R between .228 and .460). The five Decision-Making Speed summary variables were related most closely to check flight percentage scores for the T-37 instrument flight (multiple $R = .460$, $p \leq .001$) and T-38 contact flight (multiple $R = .312$, $p \leq .10$). One explanation for this finding was that the later flights placed greater demands on the pilot's ability to make quick, consistent, and accurate decisions than did the earlier flights. Performance on these flights improved as average decision-making speed and variability decreased. The check flight regression analyses are also summarized in Table 6.

Table 6. Decision-Making Speed:
Summary of UPT Outcome Regression Analyses

				Correlation with outcome					
Outcome measure	N	Mean	SD	RT-when	SX-when	RT-not when	SX-not when	% Correct	Mult. R
UPT pass/fail	512	0.801	0.400	.015	.049	.037	.035	.075	.107
ATRB TTB/FAR	410	0.549	0.498	-.108	.001	-.208*	-.108	.035	.229**
T-37 midphase grade	115	2.56	1.19	-.034	.031	-.103	-.058	-.158	.183
T-37 contact grade	114	2.96	0.94	.052	-.043	-.020	-.098	-.047	.198
T-37 instrument grade	112	2.94	1.05	.025	.103	-.079	-.213	-.063	.245
T-38 contact grade	102	2.62	1.14	-.068	.065	-.139	-.056	-.054	.166
T-38 instrument grade	100	2.89	1.11	-.081	.122	-.239	-.199	-.077	.302
T-38 formation grade	98	2.87	1.05	-.115	.102	-.190	-.058	-.184	.268
T-37 midphase percentage	115	85.48	8.36	-.005	.102	-.126	-.112	-.218	.278
T-37 contact percentage	114	91.22	5.42	.122	.022	-.021	-.115	.034	.261
T-37 instrument percentage	112	91.66	7.57	-.033	.137	-.214	-.432*	-.096	.460**
T-38 contact percentage	102	91.53	5.76	-.041	.092	-.206	-.267	-.049	.312
T-38 instrument percentage	100	92.27	6.13	-.126	-.003	-.153	-.129	.109	.238
T-38 formation percentage	98	92.80	6.83	-.069	.043	-.195	-.109	-.034	.228

*$p \leq .05$.
**$p \leq .01$.

Item Recognition

Descriptive Measures

Reaction time and accuracy of response (correct/incorrect) were recorded for 1,082 subjects on each of the 48 trials. The data from all trials that presented digit strings of the same length were summarized as a single score. As with the other tests, Digit Memory and Decision-Making Speed, this data reduction technique was used to make the data more manageable and to create a relatively small set of stable predictor variables (6 means instead of 48 scores). Table A-7 provides a summary of the response time means and standard deviations and the accuracy of response for each of the six lengths of the digit strings.

As indicated in Table A-7, the six string lengths (1-6) were not presented an equal number of times during the 48 trials. Each subject, however, did receive the same series of strings during the test.

Subjects' responses were extremely accurate across the 48 trials, with an average of 95.2% correct. This was encouraging, as it is a common practice with tasks of this type to calculate response time means and standard deviations based only on trials with correct responses. As expected, subjects generally took longer to respond as the length of the digit string increased. This suggested that the subjects needed to make more comparisons between the initial string (in memory) and the single digit as the length of the string increased.

Factor Structure

The most conceptually important measure provided by this test was average response time for correct responses for each of the six string lengths. However, it was felt that the task of memory search and comparison was qualitatively different for strings of different lengths (e.g., amount of rehearsal needed to maintain short-term memory, search and comparison strategy). As a result, for each of the six string lengths, the consistency of the standard deviations of response time and the percent correct were also of interest.

A factor analysis was performed that used 18 variables; namely, the average response time, standard deviation of response time, and percent correct--for each of the six string lengths. This was done in order to determine the interrelationships among these variables. There were 1,082 subjects for this analysis.

The inter-item correlation matrix, provided in Table A-8, yielded several interesting results. The average response times for the six string lengths were moderately to strongly related to each other ($.437 \leq r \leq .825$). Average response times for a given string length also were related strongly to the standard deviation of response time for that string length ($.641 \leq r \leq .715$). The standard deviations were moderately interrelated ($.206 \leq r \leq .386$), whereas the percent-correct scores were only marginally interrelated. Average response time and standard deviation measures were not statistically related to percent correct ($-.084 \leq r \leq .106$).

The 18 Item Recognition scores were expected to yield more than one factor, as the percent correct measure was conceptually different from the average response times and standard deviations. Before rotation, four factors were defined that accounted for 56.2% of the total item variance. After rotation, the principal factor accounted for 71.3% of the total explained variance and can be interpreted as a general "response latency" factor. Average response time and standard deviation of response time for string lengths 2, 3, and 4 loaded heavily on this factor. Factor 2 was defined primarily by the average response times and standard deviations for string lengths of 5 and 6, while factor 3 was similarly defined for string length 1. Finally, factor 4 can be interpreted as an "accuracy index," as it consisted of the six percent-correct measures. A summary of the factor analysis is provided in Table A-9 in the Appendix.

The factor solution suggested that a model that considered the average response time and its standard deviation for different string lengths, along with an overall accuracy measure, was appropriate. However, for practical purposes, the number of test variables needed to be reduced drastically. As a result, a model was developed that used a regression line for each subject's response times for the six string lengths. This method was chosen because the response times showed a linear relationship across the six string lengths and variability of response time was consistent for the different string lengths (homoscedastic). This method yielded a slope, intercept, and standard error for each subject. These three measures provided an indication of the subject's short-term memory storage and search ability for strings of differing lengths. A fourth variable, overall percent correct, was added to the model to reflect the results of the factor analysis. These four variables (slope, intercept, standard error, and percent correct) were used to predict UPT performance.

Subjects who had regression lines with low intercepts, small standard errors, and high slopes were expected to perform better on all of the UPT performance criteria. These subjects probably used a more efficient memory-searching strategy than did those whose baseline time (intercept) was high, who were inconsistent in their response times, and who took the same amount of time regardless of initial string length (little or no slope).

Inferential Measures

UPT Final Outcome/ATRB Rating. As with the Digit Memory and Decision-Making Speed measures, this test was not predictive of UPT final outcome (multiple R = .071, n.s.), but was related significantly to ATRB rating (multiple R = .261, $p \leq .0001$). Table 7 provides a summary of the Item Recognition regression analyses.

Table 7. Item Recognition:
Summary of UPT Outcome Regression Analyses

				Correlation with outcome				
Outcome measure	N	Mean	SD	Slope	Intercept	St. Error	% Correct	Mult. R
UPT pass/fail	512	0.801	0.400	-.015	-.035	-.067	-.007	.071
ATRB TTB/FAR	410	0.549	0.498	-.052*	-.183*	-.131	.055	.261**
T-37 midphase grade	115	2.56	1.19	.067	-.035	.017	.044	.093
T-37 contact grade	114	2.96	0.94	.043	-.069	-.053	.133	.137
T-37 instrument grade	112	2.94	1.05	.003	-.023	-.090	.057	.113
T-38 contact grade	102	2.62	1.14	-.050	.049	-.061	-.054	.167
T-38 instrument grade	100	2.89	1.11	-.140	-.035	-.083	.037	.231
T-38 formation grade	98	2.87	1.05	-.123	.000	-.057	-.158	.230
T-37 midphase percentage	115	85.48	8.36	.029	-.083	-.014	.015	.114
T-37 contact percentage	114	91.22	5.42	.038	-.076	-.084	.225	.232
T-38 instrument percentage	112	91.66	7.57	.041	-.125	-.148	.027	.158
T-38 contact percentage	102	91.53	5.76	-.033	.009	-.141	-.053	.243
T-38 instrument percentage	100	92.27	6.13	-.152	-.045	-.080	.002	.243
T-38 formation percentage	98	92.80	6.83	.075	-.115	-.060	-.052	.167

*$p \leq .05$.
**$p \leq .01$.

Check Flight Scores. Although the correlations were in the expected direction, the Item Recognition model was not related significantly to performance on the check flights. The predictor variables were related most closely to check flight percentage scores on the T-37 and T-38 contact flights (multiple R = .232 and .243) and the T-38 instrument flight (multiple R = .243). Table 7 provides a brief summary of these regression analyses.

An Integrated Model

Neither the AFOQT Pilot composite score nor any of the three BAT tests demonstrated a close, consistent relationship with all of the UPT performance criteria. One possible explanation was that these four cognitive measures were designed to assess performance only on simple tasks. Performance on the UPT outcome criteria, however, probably is determined more realistically by some combination of skills. Check flight grades and percentage scores, for example, were determined by the subjects' ability to perform a variety of complex maneuvers and operations during a particular flight. The specific skills that were related most closely to performance probably varied during the course of training.

It appeared that the AFOQT Pilot composite score and the three BAT tests were measuring, at least in part, different abilities, as each measure demonstrated a unique pattern of relationships to the UPT performance criteria. The Pilot composite score was related to both UPT final outcome and ATRB rating, but was unrelated to check flight performance. In contrast, none of the three cognitive tests was related to UPT final outcome. However, each of the BAT tests was related significantly to ATRB rating. Scores on the Digit Memory test were related to performance on only the T-38 formation flight. Decision-Making Speed was related most closely to performance on the later check flights. Scores on the Item Recognition test were not related significantly to performance on the check flights.

If the AFOQT Pilot composite score and the three BAT tests measured conceptually different skills, prediction of performance might be improved by use of an integrated model containing measures from more than one source. This method was used to predict UPT final outcome, ATRB rating, and check flight performance.

The "full model" regression equation used to predict UPT final outcome included the AFOQT Pilot composite score and all 12 predictors from the three computer-administered tests. This model (multiple R = .182, n.s.) did not differ significantly in predictive power from a "reduced model" that used only AFOQT Pilot composite score (r = .106) ($F[12,498]$ = 0.94, n.s.). That is, the Digit Memory, Decision-Making Speed, and Item Recognition measures did not improve the prediction of UPT final outcome beyond that provided by AFOQT Pilot composite score alone. The "integrated model" regression analyses are summarized in Table 8.

The "full model" was related significantly to ATRB rating (multiple R = .320, $p \leq .001$) and did improve prediction of performance significantly beyond that provided by AFOQT Pilot composite score alone (r = .136) ($F[12,498]$ = 3.88, $p \leq .01$).

The "full model" regression equation yielded moderate multiple correlations with both check flight grades (.311 to .431) and percentage scores (.355 to .503). This model was related significantly to performance only for the T-37 instrument percentage score (multiple R = .503, $p \leq .01$). The "full model" improved prediction of performance on the T-37 contact (multiple R = .431, $p \leq .10$) and T-38 contact (multiple R = .451, $p \leq .10$) percentage scores, but neither reached statistical significance at the .05 level. Although these results were encouraging, definite conclusions were difficult to reach, as the ratio of observations to predictors was low (less than 10 to 1) and some of the predictors were correlated strongly to each other. Results from the "full model" were compared to those from the individual tests for those instances

Table 8. Integrated Model: Summary of UPT Outcome Regression Analyses

					Multiple R			
Outcome measure	N	Mean	SD	AFOQT pilot	Digit memory	Decision-making speed	Item recognition	Integrated model
UPT pass/fail	512	0.801	0.400	.106*	.069	.107	.071	.182
ATRB TTB/FAR	410	0.549	0.498	.136**	.166**	.229**	.261**	.320**
T-37 midphase grade	115	2.56	1.19	.159	.145	.183	.093	.311
T-37 contact grade	114	2.96	0.94	.012	.167	.198	.137	.321
T-37 instrument grade	112	2.94	1.05	.160	.124	.245	.113	.356
T-38 contact grade	102	2.62	1.14	.009	.140	.166	.167	.354
T-38 instrument grade	100	2.89	1.11	.040	.177	.302	.231	.431
T-38 formation grade	98	2.87	1.05	.059	.330*	.268	.230	.408
T-37 midphase percentage	115	85.48	8.36	.059	.232	.278	.114	.365
T-37 contact percentage	114	91.22	5.42	.120	.190	.261	.232	.431
T-37 instrument percentage	112	91.66	7.57	.070	.128	.460*	.158	.503**
T-38 contact percentage	102	91.53	5.76	.063	.045	.312	.243	.451
T-38 instrument percentage	100	92.27	6.13	.010	.075	.238	.243	.377
T-38 formation percentage	98	92.80	6.87	.071	.209	.228	.167	.355

*$p \leq .05$.
**$p \leq .01$.

where they had shown a significant relationship to performance. Comparisons between the "full model" and individual test models suggested that the "full model" did not increase predictive power with regard to the check flights.

IV. CONCLUSIONS

The AFOQT Pilot composite score showed a low positive but statistically significant relationship to UPT final outcome and ATRB rating, but was unrelated to check flight performance.

The three sets of measures obtained from the BAT tests were sufficiently reliable to be used in selection systems. None of the three tests was related to UPT final outcome, but all three were predictive of ATRB rating. Digit Memory and Decision-Making Speed models were related significantly to performance on some of the later check flights.

The failure of the integrated model to consistently improve the prediction of UPT performance may have occurred for several reasons. For instance, performance on some of the tests simply may not have been related to the criterion measures. The skills measured by these simple cognitive tests may not reflect the complex combination of skills that is required in order to perform well during UPT. Further, the three tests may have been too conceptually similar to one another to provide unique contributions to the prediction of flight training performance. Strong interrelationships among predictors from the different tests (mostly means and standard deviations) may have limited the usefulness of an integrated model to improve prediction of UPT performance beyond that provided by the individual tests. An integrated model that uses predictor variables from tests that assess more distinctly different skills (e.g., information processing, spatial relations, and psychomotor ability) or more complex skills (e.g., time-sharing tasks) may be more successful in predicting flight training performance.

The ATRB results suggest that these three cognitive tests may be most useful in situations where it is desirable to classify pilot candidates into specialized training tracks at an early stage (e.g., Specialized Undergraduate Pilot Training) or when only TTB-rated or FAR-rated candidates are needed (e.g., Euro-NATO Joint Jet Pilot Training, Air National Guard).

Future research efforts will cross-validate the current findings when more data become available, and will examine an integrated model based on a combination of tests that are both more complex and more conceptually distinct from one another.

REFERENCES

Acosta, H.M. (1985). Development of the Basic Attributes Testing System (BATS): Computer-based aircrew selection and classification research (Report 85-0015). Maxwell AFB, AL: Air Command and Staff College, Air University.

Adams, J.A. (1957). The relationship between certain measures of ability and the acquisition of a psychomotor criterion response. Journal of General Psychology, **56**, 121-134.

Bordelon, V.P., & Kantor, J.E. (1986). Utilization of psychomotor screening for USAF pilot candidates: Independent and integrated selection methodologies (AFHRL-TR-86-4, AD-170 353). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Fleishman, E.A., & Hempel, W.E. (1956). Factorial analysis of complex psychomotor performance and related skills. Journal of Applied Psychology, **40**, 96-104.

Hunter D.R. (1975). Development of an enlisted psychomotor/perceptual test battery (AFHRL-TR-75-60, AD-A020 544). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Hunter, D.R., Maurelli, V.A., & Thompson, N.A. (1977). Validation of a psychomotor/perceptual test battery (AFHRL-TR-77-28, AD-A044 525). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Hunter, D.R., & Thompson, N.A. (1978). Pilot selection system development (AFHRL-TR-78-33, AD-A058 418). Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Imhoff, D.L., & Levine, J.M. (1981). Perceptual-motor and cognitive performance task battery for pilot selection (AFHRL-TR-80-27, AD-A094 317). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

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

Long, G.E., & Varney, N.C., (1975). Automated pilot aptitude measurement system (AFHRL-TR-75-58, AD-A018 151). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

McGrevy, D.F., & Valentine, L.D., Jr. (1974). Validation of two aircrew psychomotor tests (AFHRL-TR-74-4, AD-777 830). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Miller, R.E. (1966). Relationship of AFOQT scores to measures of success in undergraduate pilot and navigator training (PRL-TR-66-14, AD-656 303). Lackland AFB, TX: Personnel Research Laboratory.

McLaurin, W.A. (1973). Validation of a battery of performance tests for prediction of aerospace ground equipment course grades (AFHRL-TR-73-20, AD-774 586). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

North, R.A., & Griffin, G.R. (1977). Aviator selection 1919-1977 (Special Report-77-22). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Passey, G.E., & McLaurin, W.A. (1966). Perceptual-psychomotor tests in aircrew selection: Historical review and advanced concepts (PRL-TR-66-4, AD-636-606). Lackland AFB, TX: Personnel Research Laboratory.

Sternberg, S. (1966). High-speed scanning in human memory. Science, 153, 652-654.

APPENDIX A: TABLES

Table A-1. Digit Memory: Cell Means and Standard Deviations

Trial	N	% Correct	Perceptual Speed (RT_1) Mean (MS)	Perceptual Speed (RT_1) SD	Key-in Speed (RT_2) Mean (MS)	Key-in Speed (RT_2) SD
1	1273	85.5	1922.7	1755.8	3185.2	1769.2
2	1273	84.7	2178.8	1858.3	3059.6	1986.7
3	1273	85.5	2132.9	1975.2	2926.0	2001.0
4	1273	91.5	1859.3	1546.8	2722.0	1546.4
5	1273	89.8	1666.5	1344.6	2614.1	1396.8
6	1273	91.5	1501.3	1190.4	2787.4	1263.9
7	1273	85.9	1664.9	1141.7	2879.9	1248.8
8	1273	91.7	1502.6	1161.8	2306.4	1294.8
9	1273	88.8	1475.8	1257.9	2409.7	1258.5
10	1273	85.6	1618.0	1333.5	2843.7	1588.1
11	1273	94.7	1450.6	1019.3	2336.1	1209.0
12	1273	93.6	1394.9	1020.7	2212.1	1160.3
13	1273	87.1	1671.8	1192.5	2781.6	1330.2
14	1273	89.1	1620.8	1729.4	2346.7	1856.8
15	1273	81.1	1616.2	1030.8	2336.8	1171.1
16	1273	89.3	1566.6	1040.0	2575.3	1096.5
17	1273	92.5	1572.0	1312.4	2778.6	1500.7
18	1273	94.0	1318.5	848.9	2396.2	1023.9
19	1067	92.4	1685.6	1069.0	2757.1	1250.4
20	1067	94.6	1373.0	1213.9	2280.3	1352.4
Mean		89.4	1639.6		2626.2	
Median		89.6	1360.4		2432.2	

Table A-2. Digit Memory: Inter-Item Correlation Matrix for Perceptual Speed

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1.000																			
2	.340	1.000																		
3	.349	.363	1.000																	
4	.332	.396	.498	1.000																
5	.346	.346	.485	.487	1.000															
6	.306	.354	.434	.424	.549	1.000														
7	.342	.386	.440	.550	.555	.555	1.000													
8	.313	.367	.443	.446	.523	.592	.625	1.000												
9	.302	.340	.375	.391	.413	.462	.554	.551	1.000											
10	.290	.352	.393	.419	.523	.497	.531	.583	.507	1.000										
11	.311	.406	.356	.474	.427	.458	.559	.541	.524	.524	1.000									
12	.327	.384	.367	.456	.513	.533	.567	.595	.500	.540	.552	1.000								
13	.306	.341	.373	.459	.465	.525	.477	.488	.409	.487	.495	.541	1.000							
14	.211	.309	.257	.336	.341	.377	.401	.443	.366	.395	.494	.476	.455	1.000						
15	.295	.280	.361	.391	.469	.519	.521	.489	.438	.459	.488	.498	.521	.432	1.000					
16	.339	.322	.371	.418	.462	.505	.507	.505	.491	.510	.516	.512	.529	.414	.603	1.000				
17	.286	.305	.318	.382	.411	.473	.428	.577	.389	.468	.453	.518	.510	.411	.498	.555	1.000			
18	.286	.294	.329	.399	.410	.493	.492	.485	.458	.446	.456	.482	.508	.335	.490	.546	.533	1.000		
19	.322	.344	.337	.384	.478	.474	.482	.489	.406	.470	.485	.550	.540	.416	.489	.557	.579	.553	1.000	
20	.259	.289	.297	.310	.373	.399	.422	.463	.399	.366	.479	.392	.391	.368	.399	.446	.485	.474	.486	1.000

__Note.__ N = 1,067.

Table A-3. Digit Memory:
Rotated Factor Solution for Perceptual Speed (RT_1)

Trial	Communality	Factor 1	Factor 2
1	.2294	.2358	.4169
2	.2934	.2493	.4809
3	.4288	.1952	.6251
4	.4736	.2804	.6285
5	.5068	.3708	.6077
6	.5121	.4898	.5217
7	.6024	.4328	.6443
8	.5765	.5009	.5707
9	.4374	.4411	.4928
10	.4929	.4792	.5131
11	.5129	.5329	.4785
12	.5553	.5593	.4924
13	.5057	.5870	.4014
14	.3401	.4916	.3138
15	.4957	.5981	.3715
16	.5673	.6614	.3603
17	.5482	.6971	.2495
18	.5015	.6363	.3109
19	.5578	.6804	.3079
20	.3850	.5572	.2732

Factor	Eigenvalue	% of Variance	Cumulative %
1	8.92	93.6	93.6
2	0.61	6.4	100.0

Note. N = 1,067.

Table A-4. Decision-Making Speed: Cell Means and Standard Deviations

Subtask	Part	Response time		% Correct
		Mean	SD	
Subject Knows				
Where and When	2	609.5	334.0	96.2
	4	593.6	117.4	97.1
	8	919.3	160.2	94.3
Where only	2	639.8	122.3	98.0
	4	740.0	97.0	97.1
	8	1067.6	137.5	95.2
When only	2	507.7	107.3	94.4
	4	606.0	110.5	97.1
	8	919.7	176.2	95.3
Neither	2	663.4	138.1	96.1
	4	766.5	115.9	97.1
	8	1065.1	170.0	95.2

Note. N = 1,071.

Table A-5. Decision-Making Speed: Inter-Item Correlation Matrix

Variable	RTW2	RTW4	RTW8	RTN2	RTN4	RTN8	SxW2	SxW4	SxW8	SxN2	SxN4	SxN8	PtW2	PtW4	PtW8	PtN2	PtW4	PtW8
RTW2	1.000																	
RTW4	.606	1.000																
RTW8	.555	.638	1.000															
RTN2	.572	.613	.547	1.000														
RTN4	.492	.676	.552	.635	1.000													
RTN8	.419	.537	.607	.570	.684	1.000												
SxW2	.711	.251	.376	.247	.158	.154	1.000											
SxW4	.380	.615	.371	.291	.237	.215	.380	1.000										
SxW8	.380	.299	.708	.275	.200	.257	.442	.343	1.000									
SxN2	.379	.341	.400	.639	.238	.255	.386	.352	.409	1.000								
SxN4	.193	.287	.221	.199	.567	.291	.123	.186	.099	.143	1.000							
SxN8	.183	.225	.235	.179	.253	.604	.125	.120	.148	.119	.159	1.000						
PtW2	.191	.111	.086	.123	.098	.108	.021	.010	.044	.000	-.017	.029	1.000					
PtW4	.116	.123	.072	.088	.150	.043	.016	.005	-.023	-.016	.035	-.048	.141	1.000				
PtX8	.050	.051	.084	.100	.091	.068	-.023	-.070	.020	.018	-.050	-.007	.110	.163	1.000			
PtN2	.114	.141	.112	.224	.174	.177	-.035	.019	.010	.055	.015	.010	.228	.201	.229	1.000		
PtN4	.142	.167	.121	.189	.246	.152	.016	.036	.012	-.002	.054	.036	.227	.316	.241	.327	1.000	
PtN8	.047	.045	.056	.130	.086	.058	-.028	-.027	-0.27	.008	-.017	-.073	.142	.176	.315	.199	.180	1.000

Note. The variable labels refer to, respectively, average response time for the "when" and "not when" conditions (2, 4, or 8 stimuli/responses); standard deviations for the "when" and "not when" conditions (2, 4, or 8 stimuli/reponses); and percent correct for the "when" and "not when" conditions (2, 4, or 8 stimuli/responses). N = 1,071.

Table A-6. Decision-Making Speed: Rotated Factor Solution

Variable	Communality	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
RTW2	.7026	.7223	.3366	.1998	.0988	.1337
RTW4	.6715	.4271	.5841	.1587	.2320	.2624
RTW8	.6723	.5848	.2378	.1438	-.4225	.2730
RTN2	.8637	.2741	.3892	.2509	.1923	.7329
RTN4	.9030	.1526	.8329	.2314	.3104	.1903
RTN8	.9988	.1528	.3991	.1583	.8672	.1978
SxW2	.6219	.7861	.0515	-.0304	.0416	-.0186
SxW4	.3390	.4848	.2705	-.0621	-.0746	-.0618
SxW8	.4835	.6196	-.0533	-.0184	.0387	.0580
SxN2	.5864	.4212	.0663	-.0364	.0603	.1002
SxN4	.2846	.0886	.5105	-.0464	.0262	-.0292
SxN8	.3430	.1018	.1297	-.0517	-.0242	.0770
PtW2	.1401	.0749	.0155	.3636	-.0166	.0119
PtW4	.2108	.0322	.1158	.4324	.0260	.1620
PtW8	.2279	-.0190	-.0794	.4651	.2422	.1944
PtN2	.2781	-.0175	.0450	.5119	.0495	.6331
PtN4	.3407	.0286	.1322	.5665	.1171	.0155
PtN8	.1948	-.0348	-.0304	.4314	.5596	.0016

Factor	Eigenvalue	% of Variance	Cumulative %
1	5.04	56.9	56.9
2	1.51	17.1	74.0
3	1.11	12.5	86.5
4	0.61	6.9	93.4
5	0.59	6.6	100.0

Note. N = 1,071.

Table A-7. Item Recognition: Cell Means and Standard Deviations

		Response time		
String length	Number of trials	Mean	SD	% Correct
1	10	800.1	292.9	95.5
2	7	850.0	278.7	95.0
3	7	937.2	307.3	93.9
4	7	932.5	281.6	96.6
5	8	1027.7	300.4	95.3
6	9	1051.5	326.4	95.0

Note. N = 1,082.

Table A-8. Item Recognition: Inter-Item Correlation Matrix

Variable	RT1	RT2	RT3	RT4	RT5	RT6	SX1	SX2	SX3	SX4	SX5	SX6	Pt1	Pt2	Pt3	Pt4	Pt5	Pt6
RT1	1.000																	
RT2	.728	1.000																
RT3	.609	.782	1.000															
RT4	.640	.766	.789	1.000														
RT5	.544	.713	.772	.759	1.000													
RT6	.437	.624	.751	.749	.825	1.000												
Sx1	.715	.407	.390	.367	.321	.306	1.000											
Sx2	.332	.675	.414	.361	.321	.247	.273	1.000										
Sx3	.356	.441	.706	.406	.375	.348	.284	.346	1.000									
Sx4	.332	.352	.354	.667	.361	.285	.245	.243	.283	1.000								
Sx5	.342	.371	.334	.373	.641	.314	.258	.261	.203	.265	1.000							
Sx6	.382	.376	.402	.420	.456	.665	.386	.206	.227	.227	.253	1.000						
Pt1	-.037	.003	.020	.069	.064	.106	-.027	-.058	-.005	.016	-.052	.007	1.000					
Pt2	.052	.031	-.002	.010	.021	.053	.014	-.030	-.031	-.020	.014	.041	.143	1.000				
Pt3	.016	.003	-.018	-.029	-.030	.008	-.084	-.025	.012	-.052	-.049	-.014	.155	.091	1.000			
Pt4	-.006	-.021	-.048	-.006	-.001	.015	-.025	-.083	-.082	-.081	-.057	.010	.093	.143	.073	1.000		
Pt5	.097	.059	.031	.047	.026	.040	.033	.001	.016	.002	-.047	-.001	.150	.133	.186	.119	1.000	
Pt6	.035	.036	.041	.048	.058	.065	-.026	-.040	.024	-.007	-.009	-.002	.136	.124	.116	.107	.147	1.000

Note. These variable labels refer to average and standard deviation of reponse time and percent correct for the six string lengths.
N = 1,092.

Table A-9. Item Recognition: Rotated Factor Solution for Item Recognition

Variable	Communality	Factor 1	Factor 2	Factor 3	Factor 4
RT1	.8178	.5272	.2762	.6745	.0953
RT2	.8631	.8344	.3348	.2231	.0712
RT3	.8290	.7424	.5173	.0911	.0438
RT4	.7961	.6370	.6037	.1394	.0688
RT5	.8439	.5066	.7624	.0679	.0382
RT6	.9285	.3164	.8963	.0526	.1496
Sx1	.7235	.2427	.2150	.7847	-.0532
Sx2	.3842	.5965	.0535	.1241	-.1017
Sx3	.3386	.5430	.1916	.0793	-.0262
Sx4	.2338	.3798	.2697	.1117	-.0658
Sx5	.2489	.2909	.3636	.1329	-.1201
Sx6	.4105	.1365	.5655	.2681	.0161
Pt1	.1543	-.0213	.0704	-.0671	.3800
Pt2	.1151	-.0356	.0296	.0529	.3320
Pt3	.1363	.0214	-.0622	-.0468	.3603
Pt4	.0987	-.0955	.0197	.0196	.2978
Pt5	.1964	.0554	-.0480	.0603	.4329
Pt6	.1999	.0235	.0237	-.0273	.3436

Factor	Eigenvalue	% of Variance	Cumulative %
1	5.87	71.3	71.3
2	0.96	11.6	82.9
3	0.79	9.6	92.5
4	0.62	7.5	100.0

N = 1,082.