

FINDING THE ‘RIGHT STUFF’: DEVELOPMENT OF AN ARMY AVIATOR SELECTION INSTRUMENT

Lawrence C. Katz, Ph.D.
U.S. Army Research Institute for the Behavioral and Social Sciences
Fort Rucker, Alabama 36362

ABSTRACT

The Army Research Institute developed a computer-based, web-administered aviator selection battery to correct identified deficiencies of the Army’s current selection instrument. A rigorous six-stage process included the identification of viable predictor and criterion measures and the preliminary validation of the prototype instrument. This effort laid the groundwork for aviator classification into assigned aircraft/mission types and for UAS operator selection.

3) facilitate the transition of the final product into operation. We set out to accomplish these objectives by developing a prototype Selection Instrument for Flight Training (SIFT) via six broad tasks: Task 1: Review existing accession process and relevant literature; Task 2: Conduct training and job analysis; Task 3: Develop criterion and predictor measures; Task 4: Construct and pilot-test the prototype instrument; Task 5: Validate and assess utility and efficiency, and; Task 6: Produce a comprehensive technical report recommending a selection strategy for Army aviation.

1. INTRODUCTION

1.1 Background

The current test for selection into the Army flight school is the Alternate Flight Aptitude Selection Test (AFAST). Since it was first developed in 1988, the AFAST has been scrutinized for deficiencies, including minimal predictive validity and operational support, past expropriation of testing materials, general limitations of paper and pencil testing, a less-than-optimal selection strategy, and the likelihood that it does not measure the correct set of attributes given the considerable change in aviator requirements and in the applicant population. In June of 2004, this prompted the Army to seek to replace the AFAST with an alternate, computer-administered test for Army flight training with emphasis upon aptitudes required for Future Force aviator performance within the Future Combat Systems environment. The Army further recognizes the need to include or adapt existing tests within DOD, to have the capability of rapidly assessing the performance or validity of the new instrument, to revise the instrument when necessary, to have the capability of adapting its application to selection for related occupational categories such as Unmanned Aerial Vehicle Operators, and to develop an instrument for classification of aviators into appropriate aircraft/mission types.

1.2 Objectives

Our three main objectives were to: 1) develop an operation ready computer-based, web administered aviation selection battery that corrects identified deficiencies of the AFAST; 2) lay the groundwork for aviator classification and UAV operator selection; and,

2. REVIEW OF SELECTION STRATEGIES

2.1 Review Procedure

A focused review of aviator selection research, supplemented by relevant research from the general personnel selection domain, was conducted. The review identified more than 150 potentially relevant articles. Rather than rely entirely on a narrative summary, a spreadsheet was developed to summarize information about various test batteries and to facilitate comparison of the test batteries when deriving a recommended selection strategy.

2.2 Review Findings

Research focused specifically on pilot selection, as well as general selection research, clearly suggests that cognitive ability, or general intelligence (*g*), will be an important predictor of pilot performance. Researchers debate about the usefulness of identifying more specific abilities within the cognitive ability domain but most personnel selection research typically includes measures of different types of cognitive ability, including some combination of: general reasoning; spatial ability; mechanical reasoning; quantitative ability; verbal ability; multiple-task performance (also known as timesharing or divided attention), and information processing (e.g., perceptual speed and accuracy). Other research suggests that including measures of certain other abilities and characteristics is likely to enhance the validity of the overall selection process, including: aviation or helicopter knowledge; flight experience, and; normal-range personality characteristics. Traits that seem relevant for the pilot job include: conscientiousness; achievement orientation; stress tolerance (emotional stability); adaptability/cognitive flexibility; interpersonal

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 01 NOV 2006		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Finding The Right Stuff: Development Of An Army Aviator Selection Instrument				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences Fort Rucker, Alabama 36362				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADM002075., The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 17	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

and crew interaction skills; risk-taking; locus of control, and; interest in aviation.

3. JOB ANALYSIS

3.1 Job Analysis Procedure

The job inventory method is one of the most widely used approaches to conducting job analyses and was chosen as most appropriate for the present research. To begin this process, information concerning requirements of the Army aviator job was collected from available sources (e.g., job descriptions, training materials, subject matter expert [SME] interviews). The next step was to use this information to construct lists of all tasks believed to be relevant for the Army aviator position. These preliminary task lists were reviewed by small groups of job incumbents to ensure that they were comprehensive and relevant. Based on these meetings with SME groups, a final listing of tasks and activities was developed. This final task list, or Job Analysis Questionnaire (JAQ), included 101 task statements in 11 duty categories and 92 Knowledge, Skills, Abilities, and Other characteristics (KSAO) statements. It was then used to collect systematic, job descriptive information from a representative sample of 234 experienced Army aviators. The incumbents were asked to rate each task, activity, or KSAO regarding how critical or important it would be to successful job performance. They rated each statement on a 5-point scale from “not part of job” (0) or “unimportant” (1) to “critical” (5).

3.2 Job Analysis Findings

The JAQ survey provided information about the tasks and attributes that are important for success as an Army aviator. The abilities that received the highest importance ratings across all platforms were: *Situational awareness*; *Helicopter operation*; *Psychomotor ability*; *Information processing*, and; *Decision making*. In addition, this analysis explored the manner in which the tasks, and the KSAOs required to perform those tasks, may vary depending upon airframe. The results of this report were used to help identify predictor measures for the SIFT project.

4. PREDICTOR MEASURES

Based on the above results, the following predictor measures were recommended for inclusion in a prototype battery for validity testing: 1) *Cognitive ability*; 2) *Perceptual Speed & Accuracy*; 3) *Personality or Temperament*; 4) *Motivation/Attitude*, and; 5) *Task Prioritization*.

4.1 Cognitive Ability

The job analysis and the published selection research indicated that cognitive ability, or general intelligence (*g*), is consistently an important predictor of aviator performance. The recommendation that stemmed from the review of existing predictor measures was that the Army should use either the cognitive tests from the U.S. Navy’s Aviator Selection Test Battery (ASTB) or the USAF’s Air Force Officer Qualification Test (AFOQT) as its cognitive ability predictor measure. The Army chose the ASTB, in large part because it is already web-enabled, and was granted permission from the Navy to administer it via the Navy’s web-based system (called the Automated Pilot Examination System, or “APEX”) during the preliminary validation study. The current version of the ASTB includes subtests measuring *Reading Comprehension*, *Mathematical Ability*, *Mechanical Comprehension*, *Spatial Apperception*, and *Aviation and Nautical Information*. The pilot test described herein used the operational version that is currently in use by the Navy to select aviators. Thus, the number of items and time limit for each subtest was determined by the Navy.

In return for allowing the Army to administer the ASTB via APEX, the Navy requested that this pilot test include all five of the ASTB cognitive subtests, although some of the items in the *Aviation and Nautical Information* subtest may demonstrate low face validity for Army aviation examinees. The Navy wished to examine a new data source, and the Army felt it would be worthwhile to examine performance on their Information subtest. Both Navy and Army researchers agreed that the Army would develop its own aviator selection score composite, and that the Army’s composite score might not include all of the ASTB subtests.

4.2 Perceptual Speed and Accuracy

The recommended selection strategy included measures of perceptual speed and accuracy. Three types of items were developed for this: Hidden Figures, Simple Drawings, and Panel Displays.

The Hidden Figures test is actually a measure of the extent to which one can distinguish simple shapes or objects that are “hidden” from obvious view by interfering lines and shapes in a more complex object, often referred to as Field Independence or Figure/Ground skill. This construct has been defined as the ability to hold the stimulus shape or object in mind so as to distinguish it from other well-defined perceptual material. This construct appears to be relevant for all pilots, including helicopter pilots, as they look for objects on the ground from various perspectives aloft.

Each of the 30 items in the Hidden Figures test requires the test taker to determine which of five simple figures (presented at the top of the screen) is hidden within a complex pattern. Only one of the five simple figures is included in each complex pattern, and the figure is always right side up and the same size as in the drawings at the top of the screen. This test is scored as the number correct minus a fifth of the number incorrect (i.e., there is a correction for guessing). There is a 6-minute time limit for this test.

The Simple Drawings test is the more typical perceptual speed and accuracy measure. Each item is a set of five simple drawings, mostly non-real-world or unidentifiable objects, where one and only one is not identical to the other four. The test taker must indicate which of the five drawings is unlike the other four. There are 100 items of this type, with a two-minute time limit. Again, there is a correction for guessing in the scoring for this test.

The Panel Displays test was developed for this project to measure perceptual speed and accuracy in a way that would seem face valid to applicants. It is identical in concept to the Simple Drawings, but the objects are actual helicopter gauge displays. In each item, a set of five gauges is presented, with one and only one gauge in a slightly different configuration from the other four, which are identical. A total of 40 items were developed for this test and a time limit of two minutes was set. There is a correction for guessing for this item type as well.

4.3 Personality or Temperament

The Test of Adaptive Personality (TAP) and the Assessment of Individual Motivation (AIM) were judged to be the most reasonable measures to administer in the preliminary validation study, supplemented by new scales that would cover constructs not measured (well) by any of the existing inventories but identified as important to pilot performance in the first phase of this project. There is some redundancy in content coverage across the TAP and the AIM, but the two inventories use very different item formats.

The TAP consists of biodata items accompanied by Likert-type response scales (e.g., “Very often” to “Never”). It measures such personality constructs as *Peer Leadership*, *Work Motivation*, *Respect for Authority*, *Team orientation*, *Diplomacy*, and *Cognitive Flexibility*.

The AIM consists of personality statements and uses a forced-choice item format. Items are presented in sets of four (called tetrads). Within each tetrad, test-takers must pick the one statement that is most like them and,

among the remaining choices, the one statement that is least like them. It measures *Dominance*, *Work Orientation*, *Dependability*, *Adjustment*, and *Agreeableness*.

In addition, new scales were written to measure *Adaptability*, *Attention to Detail*, *Decisiveness*, *Multi-Tasking*, *Internal Locus of Control*, *Reasonable Risk Taking*, *Risk Tolerance*, and *Stress Tolerance*.

4.4 Motivation or Attitude

The recommended selection strategy included a measure of motivation and attitudes toward becoming an aviator. Pilot selection researchers have often measured motivation to become an aviator using a knowledge test format, that is, multiple-choice test questions that assess knowledge of aviation topics. For example, the ASTB, AFOQT, and AFAST all include an Information subtest that uses a multiple-choice knowledge test format. The logic is that persons who are more motivated to become an aviator will make an effort to learn about aviation and will thus possess more aviation knowledge than persons with lower motivation levels. Another way to measure motivation and attitudes toward becoming an Army aviator is to use a direct self-report approach.

It was decided to try both an indirect, knowledge-based approach and a direct measurement approach. First, a knowledge test called the Army Aviation Information Test (AAInfo) was developed. It was a 50-item, 5-option multiple choice test assessing such knowledge content areas as: *Major helicopter controls and parts*, *Basic operation of a helicopter*, and *Meteorological conditions impacting helicopter flight*.

Second, self-report items were developed, aimed at directly measuring motivation to become an Army aviator, called the Army Aviation Identification scale. The logic was that persons motivated to become an Army aviator would express great interest in becoming an aviator and believe that their values, attitudes, and skills are a good fit for the aviator job. This scale was included in the Army Aviation Biodata Inventory

4.5 Task Prioritization

The ability to prioritize tasks that are occurring quickly and simultaneously was measured using the “Popcorn Test”. In this test, blocks of varying size move horizontally across the computer screen from left to right at varying rates of speed. There are five lines upon which blocks may be moving at any given time, with up to three blocks moving on each line at any given time. The test-taker uses a mouse to control an on-screen cursor and scores points by “erasing” each block before it reaches the right edge of the stimulus box. Larger blocks

are worth more points than smaller blocks and faster-moving blocks are worth more points than slower-moving blocks. Block size and speed are multiplied to obtain a final score for each 90-second trial. Prior to the scored portion of the test, test-takers receive detailed instruction and one 30-second practice trial.

5. CRITERION MEASURES

To test the validity of the above subtests to predict performance in flight school, the following academic and flight training criterion measures were collected from the students tested: Academic grades for each training stage; Instructor put-up grades; Evaluator flight grades; Hours to complete each training stage; Number of set-backs at each training stage, and; Average daily flight grades from each training stage.

In addition, a Behavioral Summary Scale (BSS) was developed for this effort, to measure 10 performance-related characteristics on a 7-point scale from “Needs Improvement” to “Outstanding”. Included in the BSS were Instructor and Flight Evaluator ratings of students’: *Stress Tolerance, Adaptability, Work Ethic, Openness to Feedback, Problem Solving, Situational Awareness, Communication, Initiative, Integrity, and Overall Potential*.

6. PROTOTYPE BATTERY VALIDATION

Two-hundred forty incoming Army aviation students participated in a preliminary validation study, in which they were administered the predictor measures described above and criterion measures were subsequently collected. A total of 12 testing sessions were held between October and December of 2005, with about 20 participants in each session. Testing conditions were similar to those expected to occur operationally, and included a proctored setting with all items presented on computers via the Internet. To ensure uniformity in the testing procedures, a protocol was developed that specified the setup procedures, login processes, and introductory scripts to be used during each session. To reduce the potential impact of order effects, we counterbalanced the administration of tests. Following data quality checks for each test (assessing for missing data, response sets, extreme outliers, etc.), item-level and test-level psychometric analyses were conducted.

Criterion measures were collected regularly throughout the participants’ Primary Flight Training, Instrument Flight Training (2 stages), and Basic Warfighting Skills Flight Training. Means, standard deviations, and intercorrelations were computed for all criterion measures. Principal components analysis (PCA)

with a Direct Oblimin rotation was used to create composites for the BSS ratings. These BSS composites were then combined with the remaining criterion measures in a separate PCA that was used to create final criterion composites.

To assess the validity of the predictors at the four training stages, while minimizing the degrees of freedom required to conduct the validation analyses, predictor composites were created and tested against the criterion composites. Bivariate validity correlations were calculated for each predictor composite across all criteria and criterion composites.

7. RESULTS

The validity analyses were numerous and complex, and resulted in a number of different predictors looking like the “best bets,” depending on the specific criterion measure used. But there was a consistent pattern that emerged from these analyses, on which recommendations were based. In considering which tests to include in the operational test battery, three criteria were assessed: 1) the amount of incremental validity provided over the ASTB cognitive tests; 2) administration time, and; 3) Potential logistical or administrative issues that could make implementation difficult.

The resulting recommendation was to include the Cognitive Scales of the ASTB, the *Army Aviation Information Test*, and a composite score of *Perceptual Speed and Accuracy* subtests (Hidden Figures and Simple Drawings) in the final battery. These tests were consistently among the best predictors of performance across the various criterion measures and across the training stages.

8. CONCLUSIONS

This paper presented each of the tasks comprising the rational process whereby the Army Research Institute recently completed development of a new selection instrument for Army aviation. Three composite measures were identified as demonstrating good predictive validity for performance in Army aviation training. Other candidate tests that demonstrated some degree of incremental validity may be considered as experimental measures for which additional data should be collected. These include the AIM personality inventory and the *Peer Leadership* and *Stress Tolerance* subscales from the Army Aviation Biodata Inventory. In addition, other Biodata subscales, such as *Attention to Detail* and *Reasonable Risk-taking*, might be relevant for predicting job performance (independent of training

performance), and follow-up data are being collected for these analyses.

The next steps in this R&D effort will be to develop a classification instrument for Army aviation and a selection instrument for Unmanned Aviation Systems (UAS). Regarding the classification instrument, our objectives will be to develop a computer-based battery to determine the differential suitability of aviation students to the various Army aircraft, and to develop an automated algorithm to assign students to training tracks while they are still in initial training.

Concurrently, we are developing a computer-based, web-administered instrument to assess the relevant attributes of applicants for UAS operator training. We are using the same methodology described in this paper to produce a scientifically sound estimate of the likelihood that individuals will successfully complete training to successfully perform as UAV operators. We intend to maximize utilization of existing tests as may be found in use or under development within the Department of Defense, and to build on the work completed in the development of SIFT battery.



Finding the “Right Stuff”: Development of an Army Aviator Selection Instrument

**25th Army Science
Conference**

28 November 2006

Lawrence C. Katz, Ph.D.
U. S. Army Research Institute
Rotary Wing Aviation Research Unit
Fort Rucker, AL





Finding the “Right Stuff”

OBJECTIVES:

1. Develop an operation ready computer-based, web administered **aviator selection battery** that corrects six identified deficiencies of the current test
 - Review and revise the **selection strategy** to fit the current selection situation
 - Revise **job and task analyses** to fit current missions and training goals
 - Map resulting **KSAs** onto domains of observable behaviors
 - Develop a **computer-based, web administered** selection test to replace AFAST
2. Lay groundwork for aviator **classification** and **UAS operator** selection
3. Facilitate the **transition** of the final product into operation
 - Coordinate with and support Aviation Personnel Proponency Office in securing authorization and funding for **fielding and operation** of the test battery



Finding the “Right Stuff”

Task 1: REVIEW OF SELECTION & TEST STRATEGIES

- *Several pilot selection batteries have demonstrated reasonable validity for predicting training criteria – so validity not highly differentiating factor.*
- *Much less effort and success in predicting pilot performance beyond training.*
- US Navy is currently administering selection battery online (APEX).



Domains used in Aviator Selection:

- *Cognitive Ability:* Most batteries measure specific cognitive abilities, but general intelligence, *g* factor, typically accounts for much of the variance.
- *Aviation/Helicopter Knowledge*
- *Aviation Interest*
- *Relevant Experience*
- *Psychomotor Skills* (e.g., tracking; complex coordination)
- *Personality* (Normal range)



Finding the “Right Stuff”

Task 2: JOB ANALYSIS

KSAOs that received highest importance ratings:

Situational Awareness

Helicopter Operation

Psychomotor Ability

Information Processing

Decision Making





Finding the “Right Stuff”

Task 3: PREDICTOR MEASURES FOR VALIDITY TEST

Recommended Predictor Measures:

Cognitive Ability: All cognitive subtests from Navy ASTB

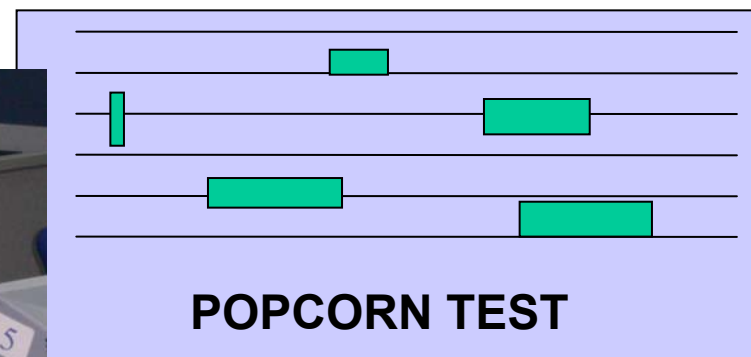
Perceptual Speed & Accuracy: Newly-developed (PSA)

Hidden Figures/Simple Drawings/Cockpit Displays

Personality/Temperament: Assessment of Individual Motivation (AIM; White & Young), Test of Adaptive Personality (TAP; Kilcullen)

Motivation / Attitude: Newly-developed (AAI), Army Aviation Identification scale (Kilcullen)

Task Prioritization: Popcorn Test





Finding the “Right Stuff”

Task 4: CRITERION MEASURES FOR VALIDITY TEST

1) Academic Grades

2) Flight Grades



3) Behavioral Summary Scale (BSS) Ratings

(Instructor & Evaluator 7-point Ratings):

- | | |
|--------------------------|-------------------------|
| -Stress Tolerance | -Problem Solving |
| -Initiative (Inst. Only) | -Integrity (Inst. Only) |
| -Adaptability | -Situational Awareness |
| -Work Ethic | -Communication |
| -Openness to Feedback | -Overall Potential |



Finding the “Right Stuff”

Behavioral Summary Scale (Example)

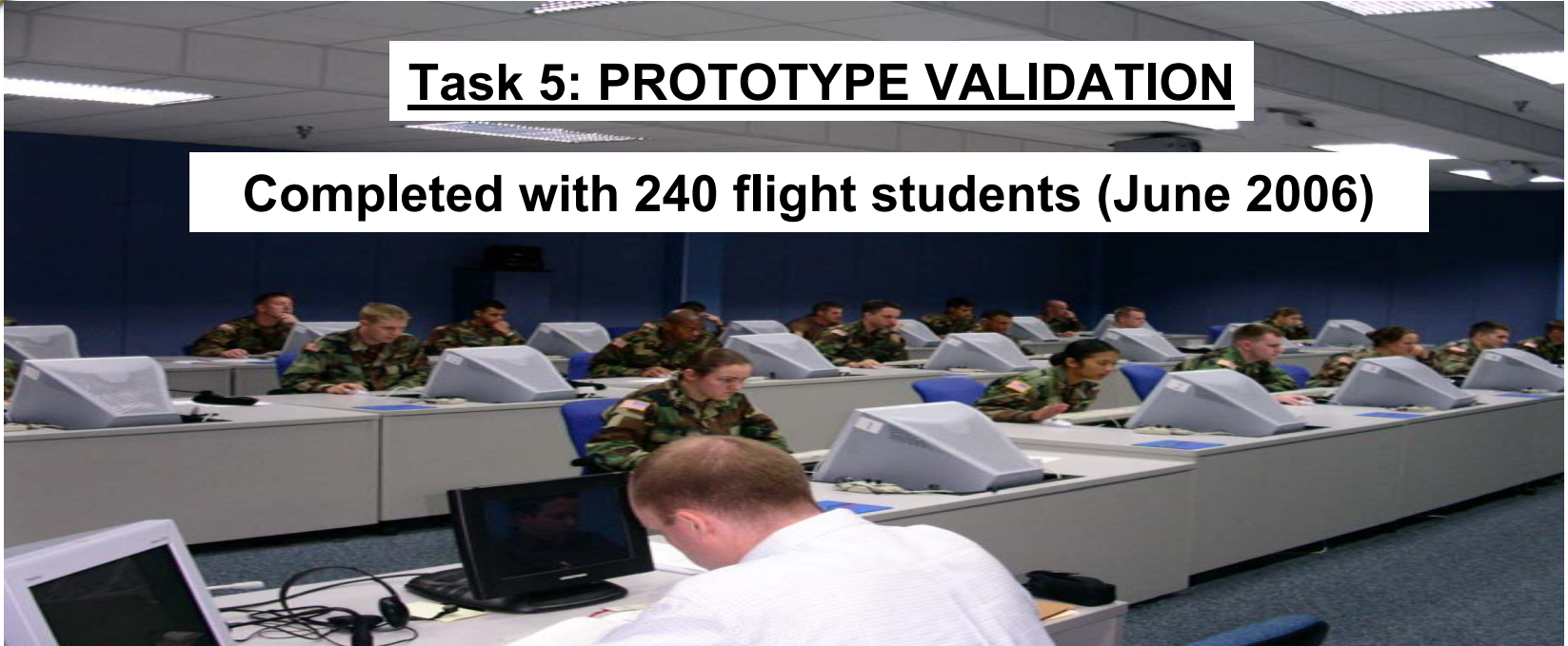
1. Stress Tolerance: Maintains composure in challenging and threatening situations; able to function well when presented with a large number of tasks to perform.								
①—————②			③—————④—————⑤			⑥—————⑦		
<i>Needs Improvement</i> • Becomes flustered when faced with many tasks. • Does not remain calm when challenged. • Shows frustration when presented with stressful situations.			<i>Satisfactory</i> • Usually functions well when faced with many tasks, but may occasionally get flustered. • Usually remains calm. • Occasionally shows frustration.			<i>Outstanding</i> • Performs extremely well during overload situations. • Always remains calm. • Manages frustration by developing constructive solutions.		
2. Adaptability: Adjusts easily to changing situations or conditions; quickly adapts and changes priorities when necessary.								
①—————②			③—————④—————⑤			⑥—————⑦		
<i>Needs Improvement</i> • Has difficulty adapting to mission changes. • “Locks in” on a single strategy.			<i>Satisfactory</i> • Usually able to adapt to mission changes. • Usually able to change strategy when necessary..			<i>Outstanding</i> • Very effectively adapts to mission changes. • Very effectively changes strategy when necessary.		



Finding the “Right Stuff”

Task 5: PROTOTYPE VALIDATION

Completed with 240 flight students (June 2006)



- **Criteria for Recommending Predictors:**
 - **Incremental validity over ASTB Cognitive Scales**
 - **Administration time (4-hour limit)**
 - **Logistical/administrative issues**



Finding the “Right Stuff”

Task 6: SIFT RECOMMENDATIONS

-based on incremental validities

Include:

- **ASTB Cognitive Scales (without “Aviation/Nautical Info”)**
 - **Math - Verbal Test (MVT)**
 - **Mechanical Comprehension Test (MCT)**
 - **Spatial Apperception Test (SAT)**
- **Army Aviation Information Test**
- **Perceptual Speed and Accuracy Test (2 subtests)**
 - **-Hidden Figures and Simple Drawings**



Finding the “Right Stuff”

- **Results:**
 - Army Aviation replacement for AFAST
 - Improved predictive validity (\$225k/student)
 - Computer-based (easily modified)
 - Internet deliverable (APEX)
 - Collaboration with Navy and Air Force
 - Numerous products developed
 - Foundation for Classification & UAS selection



Finding the “Right Stuff”

Future Developments:

- ARI will conduct validation beyond training
- ARI continues to collect criterion measures
- ARI R&D FY 07
 - Classification Instrument for Army Aviation
 - Selection Instrument for Unmanned Aviation Systems



Finding the “Right Stuff”

QUESTIONS?

