Research Report 2002

Revalidation of the Selection Instrument for Flight Training

Victor Ingurgio
U.S. Army Research Institute

C. Veronica Crawford
Consortium of Universities of the Washington Metropolitan Area and Auburn University

July 2017

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

MICHELLE SAMS, Ph.D.
Director

Research accomplished for the Department of the Army.

Technical Review by

Dr. William Bickley, U.S. Army Research Institute
Dr. Louis Miller, U.S. Army Research Institute

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2017	Final	28 March 2013 to 25 September 2015

4. TITLE AND SUBTITLE Revalidation of the Selection Instrument for Flight Training	5a. CONTRACT OR GRANT NUMBER W5J9CQ-11-C-0040
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Victor J. Ingurgio (U.S. Army Research Institute) and C. Veronica Crawford (Consortium of Universities of the Washington Metropolitan Area and Auburn University)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 225
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Building 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Building 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Research Report 2002

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ARI Research POC: Dr. Victor J. Ingurgio, ARI—Fort Benning

14. ABSTRACT *(Maximum 200 words)*: The U.S. Army recently replaced the Alternate Flight Aptitude Selection Test (AFAST) with the Selection Instrument for Flight Training (SIFT). The SIFT is an aviator aptitude test created by the Army. Prospective students are required to complete and pass the SIFT as part of admission to flight training. Created in 2006 by the U.S. Army Research Institute, the SIFT addressed limitations of the AFAST that included negligible predictive and face validity, inadequate operational support, compromised security, and outdated testing format (i.e., paper and pencil testing). The present research analyzed the utility of the SIFT in predicting students' success during flight training. SIFT exams were collected for all prospective students over the course of 2 years. Classroom and flight grades were collected for all students who successfully passed the SIFT and went on to enroll in the Initial Entry Rotary Wing (IERW) Common Core flight training (Phase I). Results suggest that the SIFT provided weak to moderate prediction of IERW Phase I grades and flight performance, respectively. Future research should verify if the SIFT can predict students' success in their Go-To-War aircrafts (Phase II).

15. SUBJECT TERMS
Selection Instrument for Flight Training, SIFT, Predictive Validity, Initial Entry Rotary Wing, IERW Common Core

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited Unclassified	56	Scott Graham 706-545-2362

Research Report 2002

Revalidation of the Selection Instrument for Flight Training

Victor Ingurgio
U.S. Army Research Institute

C. Veronica Crawford
Consortium of Universities of the Washington Metropolitan Area and Auburn University

Fort Benning Research Unit
Scott E. Graham, Chief

United States Army Research Institute
for the Behavioral and Social Sciences

July 2017

Approved for public release; distribution is unlimited.

ACKNOWLEDGMENT

This research required the collaboration of various individuals and organizations. The authors are thankful for the cadre at the U.S Naval Medicine Operational Training Center located at Pensacola, Florida. Their cooperation and willingness to assist us in accessing the Selection Instrument for Flight Training (SIFT) exam scores throughout this project helped to make this research possible. In particular, Cory Moclaire, Research Manager of the Navy Medicine Operational Training Center (NMOTC), worked closely with us in the interpretation of the SIFT exam data. The authors also recognize and are appreciative of the various Aviation Center of Excellence Chief Warrant Officers from Fort Rucker who provided the Initial Entry Rotary Wing classroom data so that we could perform our analyses.

REVALIDATION OF THE SELECTION INSTRUMENT FOR FLIGHT TRAINING

EXECUTIVE SUMMARY

Research Requirement:

The Selection Instrument for Flight Training (SIFT) is a computer-based cognitive battery designed to select U.S. Army rotary wing aviators for flight training school. Developed by the Army and implemented in 2013, the SIFT exam served to replace the paper-and-pencil Alternate Flight Aptitude Selection Test (AFAST) due to noted limitations in administration and security of test materials and psychometric properties (i.e., construct and predictive validity; Kubisiak & Katz, 2006). Of particular note, since its inception, the predictive validity of AFAST decreased from .31 to .17 (Houston, 2006). Given the aforementioned concerns, in addition to the relatively small sample size (N = 240) of the original SIFT validation study, a revalidation study was requested to provide evidence that the predictive validity of the SIFT is an improvement over the AFAST. This revalidation effort was designed to answer the following questions: How well does the SIFT total score predict Initial Entry Rotary Wing (IERW) classroom grades and flight grades? Which SIFT subscales are the best predictors of IERW classroom grades and flight grades? Were there differences in SIFT predictions based on rank—between Commissioned Officers (2LT) and Warrant Officers (WO1)? Were there differences in SIFT prediction based upon the various SIFT exam testing sites? The aim of this research was to evaluate the SIFT to understand how well it may predict success in IERW training.

Procedure:

The present research analyzed the utility of the SIFT exam in predicting students' success during flight training. Prospective students are required to complete and pass the SIFT exam prior to the start of rotary wing flight training. SIFT exam scores were collected for all prospective students over the course of approximately 2 years. Classroom and flight grades, as well as the Order of Merit List were collected for all students who successfully passed the SIFT exam and went on to enroll in the IERW Common Core flight training. Their IERW data was matched and then correlated with their scores on the SIFT and its subscales.

Findings:

Results suggest that the SIFT exam total score was a negligible to weak predictor of the IERW classroom grades and a moderate predictor of daily flight grades. Further, the SIFT exam total score was a negligible predictor of the final Order of Merit List ranking. Additionally, the SIFT exam was a better predictor of Academic Average for Second Lieutenants than for Warrant Officer 1s. Yet, there were no differences in the predictiveness of SIFT scores with regard to exam testing sites.

Utilization and Dissemination of Findings:

The data documented in this report were used to inform the Aviation Center of Excellence and the Organization and Personnel Force Development at Fort Rucker, who provide instruction to students in Initial Entry Rotary Wing training. Based on the results contained in this report, training managers should use caution when using the SIFT to predict students' success in certain Common Core (Phase l) classes within the IERW program. Although there are no plans for follow-on work, future research could help to determine if the SIFT exam can be used as a tool for making platform selection decisions and whether it may help predict students' success in their Go-To-War aircraft (Phase II).

REVALIDATION OF THE SELECTION INSTRUMENT FOR FLIGHT TRAINING

CONTENTS

LIST OF TABLES

Page

LIST OF TABLES (continued)

Page

LIST OF TABLES (continued)

LIST OF FIGURES

REVALIDATION OF THE SELECTION INSTRUMENT FOR FLIGHT TRAINING

Introduction

Valid and reliable assessment of aviator aptitude has implications for financial costs and costs of human lives. Seeking ways to improve cost effectiveness while maintaining and advancing the training and operational readiness is one challenge for the U.S. aviation training program. Aviators possess specific knowledge and specialized skills that serve as invaluable resources to the Army's mission and operations. As military operations continue to adapt to the ever changing demands of domestic and international forces, ensuring the competence and effectiveness of military personnel, including Army aviators, is of prime importance. Given the complexity of mastering the skills to become an aviator, the Army aviation program has a vested interest in strengthening its ability to select qualified candidates who will successfully complete training. For students who fail to complete aviation training, it is estimated that the Army would have spent nearly $500,000 per student (Paullin, Bruskiewicz, Houston, & Damos, 2006).

Additionally, it has been reported that Army aviation accidents have risen between the years of 2000-2007 (Aviation Week, 2008). Further, the Army reportedly experienced 120 aviation accidents in 2014 (Department of the Army, 2015), totaling an estimated $809.9 million in costs. Given the technical complexity, associated expenses, and significance of potential errors, assessing the aptitude of prospective aviators and selecting students for training is critical. In supporting Army aviation's mission of training, educating, and developing Army aviation professionals in the support of ground missions and procedures (U.S. Army Aviation Center of Excellence, 2014), this study sought to assess and re-validate its primary aviator aptitude test, the Selection Instrument for Flight Training (SIFT).

The SIFT exam is a computer-based cognitive battery that served to replace the paper-and-pencil Alternate Flight Aptitude Selection Test (AFAST). Created in 2006 by the United States Army Research Institute for the Behavioral and Social Sciences (ARI), the SIFT exam addressed limitations of the AFAST that included negligible predictive and face validity, inadequate operational support, compromised security, and outdated testing format (Kubisiak & Katz, 2006). Since its inception in 1988, the predictive validity of AFAST had decreased from .31 to .17 (Houston, 2006). Additionally, the AFAST did not adequately discriminate between students likely to pass training from students likely to fail training (Houston, 2006). Given the aforementioned concerns, coupled with the changes in aviation requirements and the demographics of prospective students, the SIFT exam was created to address these limitations and is currently being utilized to select students for aviation training. The SIFT exam has now been in official use since 1 October 2012. The present study seeks to assess the current utility of the test's ability to predict successful completion of rotary wing flight training.

Selection Instrument for Flight Training—Initial Development

To improve aviation selection, the creation of the SIFT exam was based on two phases. The first phase involved a review of the literature focused specifically on studies investigating aviator selection in particular and personnel selection in general (Katz, 2006). Experts within the field of aviator selection were also interviewed to include the latest research in test development (see Houston & Bruskiewicz, 2006 for a more detailed description of the development of the SIFT exam). The second phase involved a thorough job analysis in which subject matter experts provided recommendations regarding the knowledge, skills, abilities, and other related characteristics deemed essential to the role of aviator. Based on these two phases of research, the following predictors consistently appeared to be important considerations in aviator selection test development: cognitive ability, perceptual speed and accuracy, personality/temperament, motivation/attitude, and task prioritization. In the final SIFT test selection, cognitive ability, perceptual ability, and motivation/attitude were shown to have the most reliable psychometric properties.

Criterion Variables

In testing the initial validation of the SIFT exam and its subscales, the following criterion measures were used: academic grades from each flight training stage, instructor put-up grades, evaluator flight grades, hours taken to complete each training stage, number of set-backs at each training stage, average daily flight grades from each training stage, and the Behavioral Summary Scale (BSS), which was created specifically for the initial validation to test personality and temperament. Criterion measures were collected throughout the students' training sequence. Descriptive statistics and intercorrelations were computed for all criterion measures, and predictor composites were created and tested against criterion composites.

Initial Validation

Initial SIFT exam experimental testing took place prior to the first validation study to ensure the proper procedures for proctoring the SIFT exam, which included refining the timing standards and order of subscales and refining the wording of test questions (Katz, 2006). The initial validation took place in 2006 and was conducted with 240 Army aviation students across 12 testing sessions from October to December of 2005, averaging 20 students per testing session. Four hours were allotted for the SIFT exam testing. As with the initial experimental testing procedures, testing conditions were proctored during the first validation study to establish uniformity across testing sessions and simulate what would normally be expected in testing situations.

The administration of the SIFT subscales were counterbalanced to limit the influence of order effects. In determining which subscales to include in the final SIFT exam test selection, researchers considered the degree of incremental validity provided, the administration time, and logistical issues that could make test administration difficult. As such, the cognitive scales of the Navy's Aviator Selection Test Battery (ASTB), the Army Aviation Information Test (AAInfo), and the Perceptual Speed and Accuracy (PSA) subscales of Hidden Figures (HF) and Simple Drawings (SD) consistently met all three criteria and are currently part of the SIFT exam presently used to select students for Army aviation training. The SIFT is available only in a web-based format. The system is operated on a secure server that is monitored and controlled. Examples of test items similar to SIFT subscale items are presented in Appendix A; actual items are not provided for test security purposes.

Selection Instrument for Flight Training—Subscales

Perceptual Speed and Accuracy Test (PSA). To assess for perceptual ability, the subscales of Hidden Figures and Simple Drawings were selected for final SIFT exam inclusion. Hidden Figures (HF) measures the ability to discern simple shapes hidden within a more complex object. Examinees are required to select one of five figures hidden within a complex object. Hidden Figures was deemed to be a relevant task for all aviators, as they are required to perceive objects on the ground from various perspectives from above during flight. The original HF subtest consisted of 50 items and was revised to include a final test bank of 30 items. Simple Drawings (SD) measures perceptual speed and accuracy by having examinees indicate which object is different from four other objects. Simple Drawings (both the original and revised versions) is comprised of 100 items.

Army Aviation Information Test (AAInfo). The Army Aviation Information subscale was designed to assess one's knowledge of terminology and concepts relevant to Army aviation. It is largely believed that the more motivated a person is, the more likely a person will seek information about aviation, thus becoming more knowledgeable about aviation (Bruskiewicz, Houston, Paullin, O'Shea & Damos, 2007). The AFAST's helicopter knowledge subscale was revised to create the AAInfo (Katz, 2006). Items were based on readily available information that is accessible from multiple public sources, so as to not penalize examinees who may have limited means of accessing Army aviation material. The knowledge content of the AAInfo includes the following domains: major helicopter controls/parts, flight rules, types of helicopters, basic helicopter operations, and the impact of weather conditions on helicopter flying (Katz, 2006). Content validity was assessed by subject matter experts and revised prior to the administration of the SIFT experimental testing. The original AAInfo consisted of 50 items and was revised to include a total of 40 items.

Spatial Apperception Test (SAT). The Spatial Apperception Test is designed to measure one's ability to recognize simple changes in the position or attitude of an airplane by viewing the ground and horizon from the cockpit. Each problem in this test consists of six pictures (see Appendix A for an example): an aerial view at the upper left and five pictured choices below labeled A through E. Each pictured choice shows a plane in flight. The picture at the upper left shows the view that the pilot would have looking straight ahead from the cockpit from one of the five pictured planes. The examinee needs to determine which of the five sketches most nearly represents the position or attitude of the plane and the direction of flight from which the view would have been seen from the cockpit. This subtest consists of 25 items and is included as a subscale of the Cognitive Abilities Test.

Cognitive Abilities Test (AACog). The Army chose to modify the U.S. Navy's Aviator Selection Test Battery (ASTB) as its measure for cognitive ability as it was already formatted for online use, and the Navy provided permission to administer the test (Katz, 2006; Bruskiewicz et al., 2007). The ASTB is comprised of the following subscales: a Reading Comprehension Test (RCT)—20 items that assess the examinee's ability to extract information from passages of text, a Mathematical Skills Test (MST)—items that assess the examinee's computational skill and mathematical aptitude, and a Mechanical Comprehension Test (MCT)—items that assess the examinee's ability to perceive physical relationships and solve practical problems in applied mechanical science. Because they are adaptive tests, the number and difficulty of questions presented to different examinees varies on the MST and MCT.

Flight Training

Army aviation requires extensive training to ensure readiness for organizational missions. All initial rotary wing Army flight training is delivered at Fort Rucker, Alabama over the course of 32 weeks. Training begins with acceptance into the Initial Entry Rotary Wing (IERW) Aviator Common Core and is divided into two phases (Table 1). The first phase, Common Core, has four stages. The first stage of Common Core —Phase I training, pre-flight instruction, is completed over the course of 2 weeks. Students are oriented to basic knowledge of aircraft systems, meteorological conditions, and aerodynamics. The second stage of Common Core —Phase I training is completed in 8 weeks, where students are trained in basic maneuvers and emergency procedures. Additionally, students begin to take their first flights at this stage. All students are trained in TH-67 and OH-58D aircrafts during this initial training. Stage three of Common Core —Phase I training consists of 8 weeks where students learn about flight instruments and navigation in more detail, and they also perform simulation training. Stage four of Common Core —Phase I training is completed over the course of 4 weeks, where students are trained in basic warfighting skills (e.g., target identification and artillery and fire support). In this stage, students are trained in the OH-58D aircraft. However, this aircraft is scheduled to be replaced by the UH-72A sometime in 2019. Appendix B details the Phase I courses by training stage. At the end of stage four, successful students become instrument qualified and receive flight ratings, to include their Order of Merit List ranking, which is used upon graduation to assign students to a specialized aircraft. At the culmination of Phase I training, students would have received more than 170 hours of flight instruction (U.S. Army Aviation Center of Excellence, 2015).

Table 1

Flight Training Sequence (Adapted from Maine Army National Guard, 2015)

Phase I				Phase II
Initial Entry Rotary Wing Common Core Training TH-67, OH-58D aircraft				Go-To-War Aircraft UH-60A, CH-47D, OH-58D, AH-64A/D
Stage 1: Pre-Flight	Stage 2: Primary	Stage 3: Instruments	Stage 4: Basic Warfighter Skills (BWS)	
2 weeks	8 weeks: 46.4 aircraft hours	8 weeks: 15.7 aircraft hours and 37.5 simulated hours	4 weeks: 21.5 aircraft hours	Between 14.4–24.8 weeks: 40–89.4 aircraft hours and 10.5–64.5 simulated hours

Once students graduate from IERW Common Core Training (Phase I), they enroll into Phase II, Go-To-War Aircraft, an advanced aircraft course. Helicopter ratings achieved during the IERW Common Core phases of the flight training are used to qualify and place students in their designated aircrafts. Advanced training is completed in approximately 14–24 weeks, dependent upon the type of aircraft (i.e., UH-60A, CH-47D, OH-58D, and AH-64A/D) students are assigned. Once all flight training is completed (both Phases I and II), newly designated aviators are assigned the Military Operation Specialty (MOS) that aligns with their selected aircraft (Paullin et al., 2006).

During the IERW Common Core and Go-To-War Aircraft phases of flight training, students receive academic grades during each stage of training and flight hours are tracked. Students who may need additional hours to master flight skills are provided “extra hours” to meet the training standards. As such, flight hours are used as an indication of how efficient a student is in mastering a specific training module. Some students who do not meet the training standards may be reassigned to the next training class.

Revalidation Analyses

The goal of the present study is to perform a revalidation of the SIFT exam. Ten years have passed since the initial validation study by Houston and Bruskiewicz (2006), allowing sufficient time to re-assess whether the SIFT exam continues to have reliable and valid utility in predicting aviation students' knowledge and proficiency through the Initial Entry Rotary Wing (IERW) training course. This study sought to answer the following questions:

- How well does the SIFT exam total score predict IERW classroom grades and flight grades?
- Which SIFT subscales are the best predictors of IERW classroom grades and flight grades?
- Were there differences in SIFT exam prediction between Commissioned Officers (2LT) and Warrant Officers (WO1)?
- Were there differences in SIFT exam prediction among various testing sites?

Within the initial validation plan, several predictors and criteria were combined into composite scores in anticipation of having a relatively small sample size (O'Shea, Oppler, & Houston, 2006). Given the available data collected for this study, we retained all the original predictors. The current revalidation study was able to collect a larger sample, which allowed for a more in-depth analysis of the SIFT exam's predictive ability of students' classroom performance and flying proficiency. Therefore, all classroom grades and flight training grades were analyzed as outcome measures.

Method

Participants

A total of 9,083 SIFT exams were collected from testing sites across the United States and internationally. From this population, we collected IERW classroom and flight grades for 472 students from February 2013 – March 2015 from Fort Rucker, Alabama. The SIFT cut-score for selection into IERW is set at 40. Twenty-eight training classes were completed during this time frame, with an average class size of 10 students. Nine participants were eliminated from IERW training due to the following reasons: cheating, failed the oral defense, fit deferred (unfit, usually due to an injury), and self-eliminated. Consequently, these elimination data were separately evaluated. Therefore, a total of 463 participants were included in this revalidation.

Procedure

This study was approved by the U.S. Army Human Research Protections Office. ARI and the Navy Medicine Operational Training Center (NMOTC) partnered to access, collect, and analyze data. In the first phase of data collection, SIFT exam scores were collected by NMOTC from all domestic and international testing sites and provided to ARI to support the revalidation study. The data were password protected and stored with encryption to protect participants' identifying information. In the second phase of data collection, IERW students' classroom grades and flight proficiency grades were collected by the instructor cadre at Fort Rucker and delivered to ARI for the 28 IERW training periods. Like the SIFT exam data, these data were also protected and encrypted.

ARI began receiving data from NMOTC in February 2013. These data included identifiers that were used to match the SIFT exam scores to the classroom and flight grade data from the aviation students who were enrolled in the IERW training program through March

2015. Predictor measures included all SIFT subscales from the SIFT exam. Criterion measures included IERW academic classroom grades and flight proficiency grades, as well as the Order of Merit List. These classroom grades ranged from 0–100, with 100 indicating perfect mastery of content. Flight proficiency grades were assessed by the number of hours needed to attain mastery in the flight navigation training phase. An inverse relationship exists between hours of flight proficiency and mastery of skills in flight navigation, so that the more hours needed to attain mastery indicates less efficiency in training performance.

Measures

Predictors. In an effort to be consistent with the prior validation study, the present study analyzed the following variables and composites as predictors:

- The Army Aviation Cognitive Test (AACog) composite, as measured by the following SIFT subscales:
  - Mechanical Comprehension Test (MCT)
  - Math Skills Test (MST)
  - Reading Comprehension Test (RCT)
  - Spatial Apperception Test (SAT)

- The Perceptual Speed and Accuracy (PSA) composite, as measured by:
  - Simple Drawings subscale (SD)
  - Hidden Figures subscale (HF)

- Motivation, as indirectly measured by the Army Aviation Information subscale (AAInfo). Although the Army Aviation Information subscale is a knowledge test, Martinussen's (1996) meta-analysis of 134 studies that focused on military aircrew selection points out that most psychologists interpret such tests as measures of motivation.

Table 2 describes these predictor variables.

Table 2

Summary of SIFT Predictor Variables

Predictor variable	Predictor subscale	Description	Number of items/time limit
Army Aviation Cognitive Test (AACog)	MCT	Mechanical Comprehension Test—Assesses the examinee's ability to perceive physical relationships and solve practical problems in applied mechanical science.	varies/ 15 minutes
	MST	Math Skills Test—Assesses the examinee's computational skill and mathematical aptitude.	varies/ 25 minutes
	RCT	Reading Comprehension Test—Assesses the examinee's ability to infer meaning from text.	20 items/ 25 minutes
	SAT	Spatial Apperception Test—Assesses the examinee's ability to perceive spatial relationships from differing visual orientations.	25 items/ 10 minutes
Perceptual Speed and Accuracy (PSA)	SD	Simple Drawings Standard Score—Assesses the examinee's ability to rapidly detect the unique object within a group of similar objects.	100 items/ 2 minutes
	HF	Hidden Figures Standard Score—Assesses the examinee's ability to rapidly identify symbols contained within a larger, complex pattern.	30 items/ 5 minutes
Army Aviation Information Test (AAInfo)	AAI	Army Aviation Information Test Standard Score—Assesses the examinee's knowledge of terminology and concepts relevant to Army aviation.	40 items/ 30 minutes
SIFT total score	SIFT_Z	Final raw SIFT: *z*-score	Between 4, -4
	SIFT_T	Final SIFT: *t*-score (the only score that examinees see)	Between 0, 80
Total items (can vary from 215 to 275, where the MCT and the MST can add up to 60 items total to the SIFT)/minimum time to complete (Examinees are given 4 hours to complete)			Maximum of 275 items/ 1 hour and 52 minutes

Criterion. Similar to the original validation study, the following criterion variables in Table 3 were collected: academic grades from each training phase and stage and hours to flight proficiency. In addition to these variables, the Order of Merit List rankings (OML) were used as a criterion variable. The OML is a ranking system in which all students at the end of Phase I receive a class rank that qualifies them to make a suggestion with regard to their aircraft of interest in Phase II, Go-To-War Aircraft training. Their suggestion is just a preference, as the needs of the Army come first.

Table 3

Sequence of IERW Training

Phase and stage		Criterion variables
Phase 1	Stage 1–Pre Flight	No grades are collected during this stage of training.
	Stage 2–Primary	Common Core Proficiency 1 Grade (CC Prof Grd 1) Time to Common Core Proficiency: Milestone 1 (CC Prof Time Grd 1) Academic Average Grade (Academic AVG Grd) Average Flying Grade (Daily Flying AVG Grd) Extra Hours Awarded (Extra Hrs)
	Stage 3–Instruments	Common Core Proficiency 2 Grade (CC Prof Grd 2) Time to Common Core Proficiency: Milestone 2 (CC Prof Time Grd 2)
	Stage 4–Basic Warfighter Skills (BWS)	Hours to Radio Communications Procedures Proficiency (RCP Hrs) Hours to Instrument Takeoff Proficiency (I. Takeoff Hrs) Hours to Radio Navigation Proficiency (R. Nav Hrs) Hours to Holding Procedures: Visual High Frequency Omnidirectional Range Proficiency (HP_VOR Hrs) Hours to Holding Procedures: Loss of Control Proficiency (HP_LOC Hrs) Hours to Holding Procedures: Global Positioning System Proficiency (HP_GPS Hrs) Hours to Holding Procedures Proficiency (HP Hrs) Hours to Precision Approach: Instrument Landing System Proficiency (PA_ILS Hrs) Hours to Precision Approach Global Positioning System/Localizer Performance with Vertical Guidance Proficiency (PA_GPS/LPV Hrs) Hours to Precision Approach: Precision Approach Radar Proficiency (PA_PAR Hrs) Hours to Precision Approach Proficiency (PA Hrs)

		Hours to Unusual Altitude Recovery Proficiency (UAR Hrs)
		Hours to Straight and Level Flight Proficiency (S&L Flight Hrs)
		Hours to Climb and Descent Proficiency (C&D Hrs)
		Hours to Standard Rate Turn Proficiency (SRT Hrs)
		Hours to Steep Turn Proficiency (ST Hrs)
		Hours to Climbing/Descending Turn Proficiency (C&D Turns Hrs)
		Hours to Acceleration and Deceleration Proficiency (Acc/Dec Hrs)
		Missed Approach Proficiency (MA Hrs)
Phase 2	Go-To-War Aircraft	Basic Instruments Grade (BI Prof Grd)
		Time to Basic Instruments Milestone (BI Prof Time Grd)
		Advanced Instruments Grade (AI Prof Grd)
		Time to Advanced Instruments Milestone (AI Time Prof Grd)
		Basic Warfighter Skills Grade (BWS Prof Grd)
		Time to Basic Warfighter Skill Milestone (BWS Time Prof Grd)

Results

All statistical analyses were performed using IBM's Statistical Package for the Social Sciences (SPSS Statistics v.21). The results are presented here in the same order as the initial validation report, except where noted. In the IERW student sample, males were slightly more represented compared to the initial validation study—93.3% of the revalidation study IERW sample being males, compared to 90.4% in the initial validation study. The revalidation participants were slightly older compared to the initial validation study with participants averaging just under 28 years of age, compared to about 23 years of age in the initial validation study. Descriptive statistics for all demographic variables comparing the SIFT examinee population (N = 9,083) to the present study's IERW sample (N = 463) and the eliminated IERW students (N = 9) are provided for comparison in Tables 4-6.

Table 4

Age and Gender Comparison

	Age				Gender			
					Male		Female	
Database	*N*	Range	Mean	*SD*	*n*	%	*n*	%
SIFT examinees	9,083	N/A	N/A	N/A	8,250	90.8	833	9.2
IERW students	463	20–36	27.8	3.5	432	93.3	31	6.7
IERW eliminated students	9	28–34	31.4	1.9	7	77.8	2	22.2

Note. Numbers are rounded to the nearest tenth.

For the IERW sample, 407 participants identified themselves as Caucasian (87.9%), 20 identified themselves as African American (4.3%), 13 identified themselves as Asian (2.8%), two identified themselves as Native American (0.4%), five identified themselves as Pacific Islanders (1.1%), and 16 participants identified themselves as Other (3.5%). Additionally, 432 participants identified themselves as Not Hispanic (93.3%) and 31 participants (6.7%) identified themselves as Hispanic.

Table 5

Racial/Ethnic Background Comparison

Database	*N*	Race	*n*	%	Ethnicity	*n*	%
SIFT examinees	9,083	Caucasian	7,593	83.6	Not Hispanic	8,112	89.3
		African American	550	6.1	Hispanic	971	10.7
		Asian	323	3.6			
		Native American	71	0.8			
		Pacific Islander	86	0.9			
		Other	460	5.1			
IERW students	463	Caucasian	407	87.9	Not Hispanic	432	93.3
		African American	20	4.3	Hispanic	31	6.7
		Asian	13	2.8			
		Native American	2	0.4			
		Pacific Islander	5	1.1			
		Other	16	3.5			
IERW eliminated students	9	Caucasian	8	88.9	Not Hispanic	8	88.9
		Asian	1	11.1	Hispanic	1	11.1

Note. Numbers are rounded to the nearest tenth.

Warrant officers (65.7%) were more represented in the IERW revalidation sample, compared to the initial validation study (54.6%) with First Lieutenants, Second Lieutenants, and Captains making up 34.3% of the IERW revalidation sample compared to 45.4% for the initial validation study (Captains were not represented in the initial validation study). Most eliminated students were WO1s.

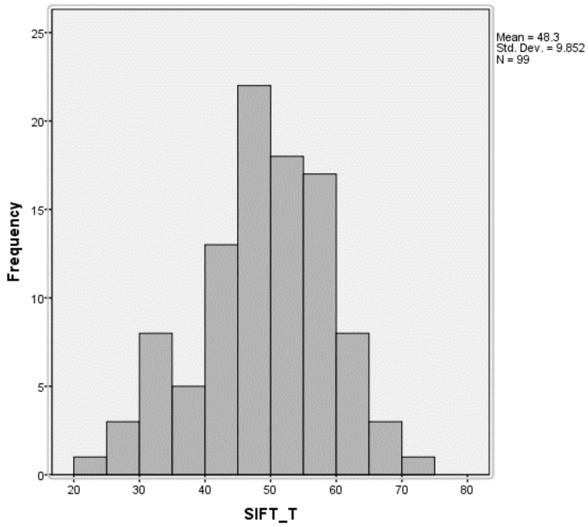
Table 6

Rank Comparison

Database*	*N*	Rank	*n*	%
IERW students	463	WO1	304	65.7
		1LT	11	2.4
		2LT	146	31.5
		CPT	2	0.4
IERW eliminated students	9	WO1	8	88.9
		1LT	1	11.1

*SIFT examinees: Rank information for the SIFT examinee population was not available.
Note. Numbers are rounded to the nearest tenth.

In comparing the SIFT examinee population scores to the "ideal" normative statistical values (in a perfect world, the mean = 50, and *SD* = 10), the SIFT examinee scores have been increasing. Retesting may be impacting the scores; if one fails their first attempt at the SIFT exam, they can re-take it after 6 months. As C. Moclaire (personal communication, September 29, 2014) puts it, "we've seen the pass rate come up a bit, which should be expected, as gouge gets out on the test (for example, study guides become available leading to higher scoring) and more people who failed the first time have the chance to re-test". These analyses have been scrubbed for individuals who retested after not passing the SIFT on their first attempt and their non-passing SIFT scores were not included in the analyses presented here. The SIFT examinee total scores for the first 99 exams collected were analyzed in September of 2012 (Figure 1), where the mean was 48.3 (*SD* = 9.9). Next, the average SIFT exam score in December 2013 was 50.4 (*SD* not available), and in September 2014, the first 7,089 SIFT exam scores (Figure 2) had a mean of 51.8 (*SD* = 11.2). Finally, as of May 2015, a total of 9,240 exams were taken with a mean score of 52.4 (*SD* = 11.2).



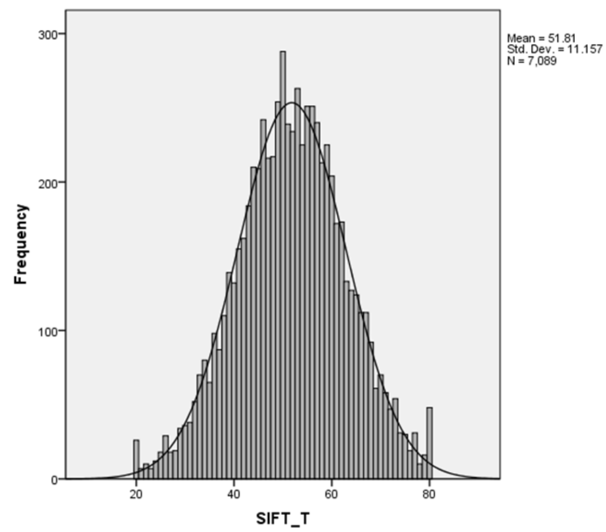
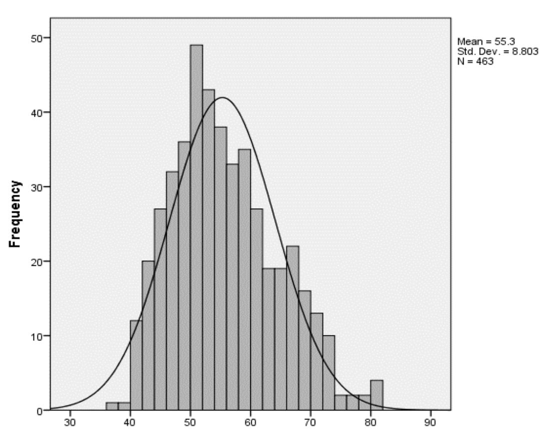


Figure 1. Histogram of first 99 SIFT scores

Figure 2. Histogram of the first 7,089 SIFT scores

In comparison, the present study's IERW student (Figure 3) mean score (N = 463) is 55.3 (SD = 8.8) and the SIFT total population of examinees, including the IERW students (Figure 4) mean score (N = 9,083) was 52.5 (SD = 11.1). O'Shea, Bruskiewicz and Houston (2006) found that AAInfo total scores were impacted by the examinees perception of prior knowledge; those who had indicated that they knew more about helicopters, spent more time learning about helicopters, and had private licenses to fly, had significantly higher scores on the AAInfo test. Across the board, we see that those admitted to the IERW program had higher subscale scores. However, the only significant difference was for the AAInfo test ($t(9{,}081) = 7.36$, $p < .001$). As a reminder, the highest score one can earn on the SIFT exam is an 80.



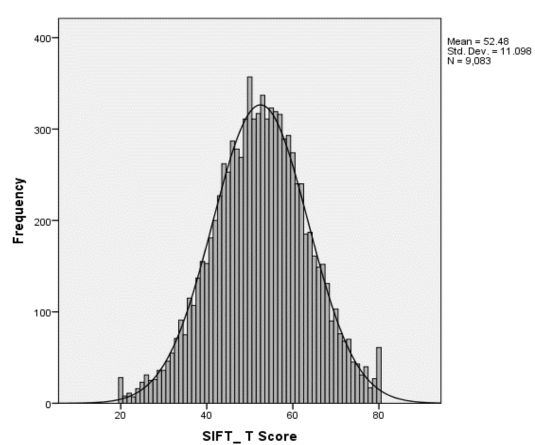


Figure 3. Histogram of IERW students for this study.

Figure 4. Histogram of SIFT examinees for this study.

Further, there were essentially no differences between the SIFT examinees, the IERW student sample, and the IERW eliminated students with regard to average SIFT exam completion times (Table 7). The SIFT exam completion time analysis was not analyzed in the initial validation study, but is included here to determine if exam completion time may have been indicative of success in IERW, which it was not.

Table 7

Overall SIFT Scoring and Time to Complete Comparisons

	SIFT score		Exam completion time		
Database	Mean	*SD*	Range	Mean	*SD*
SIFT examinees	52.5	11.1	46 minutes–2 hours and 27 minutes	1 hour and 41 minutes	17 minutes
IERW students	55.3	8.8	58 minutes–2 hours and 27 minutes	1 hour and 42 minutes	16 minutes
IERW eliminated students	51.0	10.3	1 hour and 30 minutes–2 hours and 1 minute	1 hour and 45 minutes	12 minutes

Note. Times are rounded to the nearest minute.

Table 8 shows the four possible locations that students accepted into the IERW course may have taken the SIFT exam, the respective average completion times, as well as the average SIFT total scores. Of the 463 students enrolled in an IERW course during the February 2013-March 2015 timeframe, more than half (54.2%) of the examinees were serving in active duty when they took the SIFT exam, with another 20.5% coming from the National Guard. On average, SIFT exam total scores were approximately 55 across the testing locations among all the IERW students. Further, duration of testing times was nearly the same across all SIFT testing locations, taking 1 hour and 45 minutes or less to complete. Again, this analysis was not performed in the initial validation study and was performed here to determine the impact of location and time to complete the SIFT exam on IERW success, for which there was no evidence.

Table 8

SIFT Average Testing Time to Complete and Scores by Location for IERW Students

			Testing time		SIFT score		
Location	*n*	%	Mean	*SD*	Mean	*SD*	Range
Military Entrance Processing Station	61	13.2	1 hour and 45 minutes	14 minutes	55.8	9.2	40–78
Active duty	251	54.2	1 hour and 42 minutes	16 minutes	55.3	8.6	40–80
National Guard	95	20.5	1 hour and 43 minutes	17 minutes	55.2	8.8	38–80
University	56	12.1	1 hour and 42 minutes	17 minutes	55.1	9.4	37–78

Note. Times are rounded to the nearest minute.

Descriptive statistics of the SIFT predictor variables are provided in Table 9, and the values are represented as standardized *z*-scores; however, the values given for the composite variables (AACog and PSA) are the additive *z*-scores of their individual subscales.

Table 9

Descriptive Statistics for SIFT Predictor Variables

Predictor variables	*N*	Min	Max	Mean	*SD*
Simple Drawings	463	-5.92	3.46	-.05	1.07
Hidden Figures	463	-2.46	4.24	.26	1.21
Army Aviation Information	463	-3.36	2.11	.03	1.01
Mechanical Comprehension	463	-1.95	3.58	.01	0.92
Math Skills	463	-2.59	2.87	.06	0.82
Reading Comprehension	463	-2.51	2.93	.52	0.92
Spatial Apperception	463	-2.35	2.54	.66	0.92
Army Aviation Cognition Composite	463	-7.38	11.59	1.28	2.90
Perceptual Speed and Accuracy Composite	463	-5.23	5.41	.21	1.86
SIFT	463	-1.33	3.61	.53	0.89

Note. Minimum, maximum and mean values are represented as *z*-scores, and additive *z*-scores for the two composite variables.

Descriptive statistics for students' IERW classroom grades and flight proficiency in hours are provided for consideration in Tables 10 and 11, respectively. Sample size (N's) for these tables vary due to the fact that not everyone accepted into the IERW program actually finished the program.

Table 10

Descriptive Statistics for Criterion Variables for IERW Common Core Training—Classroom Grades and Time to Proficiency

	Classroom grades and times to proficiency				
Criterion variable	*N*	Min	Max	Mean	*SD*
Grade: Common Core Proficiency 1	438	63	100	77.39	5.06
Time to proficiency: Common Core (Milestone 1)	463	1	73	16.92	6.49
Grade: Common Core Proficiency 2	430	76	98	90.33	3.86
Time to proficiency: Common Core (Milestone 2)	463	1	28	0.93	4.15
Grade: Basic instruments	378	63	100	79.61	6.19
Time to proficiency: Basic Instruments Milestone	463	1	16	5.38	3.28
Grade: Advanced instruments	361	72	100	80.82	4.73
Time to proficiency: Advanced Instruments Milestone	463	1	43	1.59	7.04
Grade: Basic Warfighter Skills	361	72	99	81.23	5.4
Time to proficiency: Basic Warfighter Skill Milestone	463	1	52	9.53	9.5
Grade: Academic average	463	44	100	92.56	8.39
Grade: Daily flying average	463	63	93	82.77	3.39
Time to proficiency: extra hours	23	1	20	4.67	4.27

Note. Grade means and standard deviations have been rounded to the nearest hundredth.

Table 11

Descriptive Statistics for Criterion Variables for IERW Common Core Training—Flying Proficiency

		Flying proficiency (Notated in hours to proficiency)			
Criterion variables	*N*	Min	Max	Mean	*SD*
Radio communication procedures	359	1	29	5.29	5.84
Instrument takeoff	365	1	31	7.33	6.46
Radio navigation	382	1	29	5.93	6.39
Holding procedures—Vor	372	2	32	6.96	5.23
Holding procedures—LOC	367	1	36	6.47	6.18
Holding procedures—Global Positioning System	366	1	31	5.46	4.58
Holding procedures	148	1	37	4.38	7.30
Precision approach—ILS	355	1	28	3.54	4.30
Precision approach—GPS.LPV	356	1	29	3.88	5.67
Precision approach—PAR	330	1	31	2.37	3.53
Precision approach	64	1	15	2.31	2.81
Unusual altitude recovery	170	1	11	1.57	0.76
Straight and level flight	345	2	11	5.25	4.05
Climbs and descents	384	2	11	4.52	2.35
Standard rate turn	396	2	11	3.54	2.31
Steep turn	344	1	11	3.65	3.13
Climbing and descending turn	388	2	9	4.52	1.92
Acceleration and deceleration	392	2	9	3.88	1.66
Missed approach	367	1	27	4.43	4.25
Total hours to flying proficiency	401	3	248	73.28	39.80

Note. Numbers have been rounded to the nearest hundredth. Flying proficiency is notated by hours required to master proficiency in designated flying module.

Intercorrelational and Correlational Analyses

For the IERW sample, intercorrelational analyses among the SIFT subscales were conducted and compared to the analyses from the initial validation study (Katz, 2006). While most of the relationships between the subscales are statistically significant, the strength of the relationships between the subscales are weaker for this study when compared to the relationships from the initial validation study. For this study, the AAInfo test exhibited positive relationships with all the other subtests, with the exception of Math Skills. The AAInfo exhibited positive relationships with all subscales in the initial validation study. Math Skills (MST) and Mechanical Comprehension (MCT) continue to be the most strongly related to each other (r = .51, $p < .01$), although the strength of the correlation decreased by .07 from the initial validation study. Intercorrelational analyses from the initial validation study and the present study are provided for comparison in Appendix C. In all cases, the intercorrelations were lower in this revalidation study.

Correlational analyses between the SIFT subscales and classroom grades (Tables 12 and 13) and between the SIFT subscales and flight proficiency (Table 14 and 15) are provided. Analyses suggest that although not all SIFT subscales were significantly related to all classroom grades and flight proficiency, significant relationships existed among several of the SIFT subscales. Most notably, the SIFT exam total score, the Army Aviation Cognitive (AACog) Composite (i.e., combination of all cognitive subscales, to include mechanical comprehension, math skills, reading comprehension, and spatial apperception), and the Army Aviation Information (AAInfo) test were most strongly correlated with classroom grades, Order of Merit List rankings, and flight proficiency. Of the SIFT subscales, Spatial Apperception (SAT) correlated most highly with classroom performance and flight performance.

Table 12

Correlations Among SIFT Predictor Measures and Classroom Grades—Part 1

Predictor Measure	Common Core Proficiency 1 Grade	Time to Common Core Proficiency (Milestone 1)	Common Core Proficiency 2 Grade	Time to Common Core Proficiency (Milestone 2)	Basic Instruments Grade	Time to Basic Instruments Milestone
	N = 438	*N* = 463	*N* = 430	*N* = 463	*N* = 378	*N* = 463
Simple Drawings	.01	.05	.11*	-.10*	-.03	-.01
Hidden Figures	.06	.05	.12*	-.04	.04	-.01
Perceptual Speed and Accuracy Composite	.05	.07	.14**	-.09	.01	-.02
Army Aviation Information	.24**	-.06	.26**	.03	.29**	-.01
Mechanical Comprehension	.07	.01	.19**	-.08	.12*	-.08
Math Skills	.02	.05	.10*	-.10*	.07	-.13**
Reading Comprehension	.01	.07	.06	-.05	.05	-.00
Spatial Apperception	.17**	-.08	.21**	-.07	.22**	.05
Army Aviator Cognitive Composite	.10*	.03	.20*	-.11*	.17*	-.06
SIFT exam total score	.15**	.03	.26**	-.10*	.19**	-.05

Note. *$p < .05$, one tailed. **$p < .01$, one tailed. Numbers are rounded to the nearest hundredth.

Table 13

Correlations Among SIFT Predictor Measures and Classroom Grades—Part 2

	Advanced instruments grade	Time to advanced instrument milestone	Basic warfighter skills grade	Time to basic warfighter skill grade	Academic grade	Flying grade	Extra hours	Order of Merit List
Predictor Measure	*N* = 361	*N* = 463	*N* = 361	*N* = 463	*N* = 463	*N* = 463	*N* = 23	*N* = 391
Simple Drawings	-.02	-.11*	-.04	-.03	-.01	.07	.42*	.01
Hidden Figures	.06	-.03	.04	-.06	.03	.10*	-.17	-.04
Perceptual Speed and Accuracy Composite	.02	-.08	.01	-.06	.02	.11*	.10	-.03
Army Aviation Information	.24**	.02	.21**	.02	.13**	.37**	.12	-.24**
Mechanical Comprehension	.07	-.09	.03	-.21**	.16**	.23**	-.54**	-.14**
Math Skills	.06	-.09	.03	-.17**	.13**	.15*	.04	-.10**
Reading Comprehension	.02	-.07	.04	-.04	.12*	.10*	-.01	-.03
Spatial Apperception	.19**	-.07	.18**	.03	.12*	.32**	-.59**	-.10
Army Aviator Cognitive Composite	.13*	-.12*	.10*	-.14**	.19**	.30**	-.60**	-.20**
SIFT exam total score	.15**	-.11*	.12*	-.11*	.16**	.33**	-.27	-.16**

Note. *p < .05, one tailed. **p < .01, one tailed. Numbers are rounded to the nearest hundredth.

Table 14

Correlations Among SIFT Predictor Measures and Flight Proficiency—Part 1

Predictor measure	PA *N* = 64	UAR *N* = 170	S&L flight *N* = 345	C&D *N* = 384	SRT *N* = 396	ST *N* = 344	C&D Turn *N* = 388	Acc/Dec *N* = 392	MA *N* = 367	Total hrs to flight proficiency *N* = 401
Simple Drawings	.17	.07	-.01	.01	-.00	-.02	.06	.03	-.06	-.03
Hidden Figures	-.07	-.05	-.02	-.11*	-.02	-.06	-.04	-.02	-.13**	-.08*
Perceptual Speed & Accuracy Composite	.04	.00	-.02	-.07	-.01	-.05	.00	.01	-.12*	-.07
Army Aviation Information	-.16	-.11	-.09*	-.11*	-.06	-.09	-.20**	-.27**	-.04	-.12**
Mechanical Comprehension	-.03	-.11	-.10**	-.01	-.04	-.01	-.17**	-.19**	-.18**	-.21**
Math Skills	.06	-.05	-.03	-.01	-.02	-.08	-.11*	-.08	-.08	-.13**
Reading Comprehension	-.13	-.02	-.09*	-.09*	-.08	-.12*	-.05	-.06	-.11*	-.11*
Spatial Apperception	-.21	-.01	.00	-.12*	-.05	-.05	-.14**	-.15**	-.06	-.06
Army Aviator Cognitive Composite	-.14	-.05	-.08	-.09	-.07	-.10	-.18**	-.18**	-.16**	-.19**
SIFT Exam total score	-.12	-.06	-.09	-.11*	-.06	-.10*	-.16**	-.18**	-.17**	-.18**

Note. **p* <.05, one tailed. ***p* <.01, one tailed. Numbers are rounded to the nearest hundredth.

Table 15

Correlations Among SIFT Predictor Measures and Flight Proficiency—Part 2

	RCP	I. Takeoff	R. Nav	HP_Vor	HP_LOC	HP_GPS	HP Hrs	PA_ILS	PA_GPS. LPV	PA_PAR
Predictor Measure	*N* = 359	*N* = 365	*N* = 382	*N* = 372	*N* = 367	*N* = 366	*N* = 148	*N* = 355	*N* = 356	*N* = 330
Simple Drawings	-.01	.03	.03	-.08	-.00	-.02	.03	.02	.01	.03
Hidden Figures	-.03	-.04	.02	-.05	-.03	-.06	-.04	-.00	-.04	.09
Perceptual Speed & Accuracy Composite	-.03	-.01	.03	-.08	-.02	-.05	-.01	.01	-.02	.07
Army Aviation Information	-.08	-.07	-.08	-.14**	-.06	-.06	.10	-.07	-.16**	-.16**
Mechanical Comprehension	-.08	-.08	-.14**	-.17**	-.14**	-.11*	-.02	-.16**	-.14**	-.09
Math Skills	-.03	-.07	-.07	-.09*	-.08	-.07	.10	-.07	-.07	.06
Reading Comprehension	-.05	-.07	-.04	-.09*	-.08	-.03	-.02	-.04	-.05	-.08
Spatial Apperception	-.07	-.09*	-.08	-.08	.02	-.08	-.03	-.09	-.13**	-.01
Army Aviator Cognitive Composite	-.08	-.12*	-.12*	-.16**	-.10	-.11*	.01	-.13*	-.14**	-.04
SIFT Exam total score	-.09*	-.10*	-.09*	-.18**	-.09*	-.11*	.03	-.10*	-.14**	-.04

Note. *$p < .05$, one tailed. **$p < .01$, one tailed. Numbers are rounded to the nearest hundredth.

The Army Aviation Information test and the Spatial Apperception subscale were the best predictors of IERW classroom grades and flight performance. The Math Skills and Mechanical Comprehension subscales contributed to the prediction of IERW classroom grades and flight performance to a lesser degree. Finally, we performed correlational analyses with regard to rank to determine if the SIFT exam total score, SIFT subscales, or composites correlated with IERW classroom grades, flight performance, and warfighter skills. The SIFT exam total score correlated higher with Academic Average for 2LTs ($r = .33$) than for WO1s ($r = .07$). The Army Aviation Information subscale correlated with the Order of Merit List for 2LTs ($r = .43$). Also, the combination of the Army Aviation Information and the Spatial Apperception subscales correlated with Daily Flight Grades for both WO1s ($r = .42$) and 2LTs ($r = .45$). Further, the combination of the Reading Comprehension, Mechanical Comprehension, and the Simple Drawings subscales were correlated to IERW Class Grades for 2LTs ($r = .41$).

Regression Analyses

In assessing whether the SIFT exam total score, the Army Aviation Information subtest, and the composite scales of Army Aviation Cognitive and Perceptual Speed and Accuracy predict performance in IERW training, full (simple linear) regression analyses were conducted (Tables 16-19).

The SIFT exam total score was shown to predict both classroom performance and flight performance. Students who earned higher SIFT exam total scores earned higher academic averages [$F(849, 31,666) = 12.36$, $p < .001$] and had higher Order of Merit List rankings [$F(374, 13,927) = 10.44$, $p < .001$]. Additionally, students who earned higher SIFT exam total scores earned higher daily flight averages [$F(591, 4,727) = 57.60$, $p < .001$] and needed less time to meet flight proficiency standards [$F(21,560, 612,164) = 14.05$, $p < .001$].

Like the SIFT exam total score, both the Army Aviation Cognitive (AACog) composite and the Army Aviation Information (AAInfo) test predicted IERW classroom grades and flight performance. Students who earned higher scores on AACog earned higher academic averages [$F(1,205, 31,309) = 17.74$, $p < .001$], had higher Order of Merit List rankings [$F(274, 14,027) = 7.61$, $p < .01$], earned higher daily flight averages [$F(473, 4,844) = 45.03$, $p < .001$], and needed less time to meet flight proficiency standards [$F(22,795, 610,929) = 14.89$, $p < .001$]. Students who earned higher scores on AAInfo test earned higher academic averages [$F(567, 31,947) = 8.19$, $p < .001$], had higher Order of Merit List rankings [$F(21, 347) = 23.43$, $p < .001$], earned higher daily flight averages [$F(729, 4,588) = 73.28$, $p < .001$], and needed less time to meet flight proficiency standards [$F(9,603, 624,121) = 6.14$, $p < .01$]. The Perceptual Speed and Accuracy composite predicted only daily flight average. Students who earned higher scores on the Perceptual Speed and Accuracy composite earned higher daily flight averages [$F(60, 5,257) = 5.26$, $p < .05$].

Table 16

Full Regression Analysis: SIFT Exam Total Score

	Predictor SIFT exam total score				
	R	*R* squared	*df*	*F*	Significance
Criterion variables					
		Classroom performance			
Academic average	.16	.03	849, 31,666	12.36	< .001
Order of Merit List rankings	.16	.03	374, 139,267	10.44	< .001
		Flight performance			
Daily flying average	.33	.11	591, 4,727	57.60	< .001
Total hours to flight proficiency	.18	.03	21,560, 612,164	14.05	< .001

Note. df are rounded to the nearest whole number. Decimals are rounded to the nearest hundredth.

Table 17

Full Regression Analysis: PSA

	Predictor Perceptual Speed and Accuracy Composite (PSA)				
	R	*R* squared	*df*	*F*	Significance
Criterion variables					
			Classroom performance		
Academic average	.02	.00	8, 32,507	0.11	> .05
Order of Merit List rankings	.03	.00	9, 14,292	0.24	> .05
			Flight performance		
Daily flying average	.10	.01	60, 5,257	5.26	< .05
Total hours to flight proficiency	.07	.01	3,111, 630,613	1.97	> .05

Note. *df* are rounded to the nearest whole number. Decimals are rounded to the nearest hundredth.

Table 18

Full Regression Analysis: AACog

	Predictor Army Aviation Cognitive Composite (AACog)				
	R	*R* Squared	*df*	*F*	Significance
Criterion variables					
		Classroom performance			
Academic average	.19	.04	1,205, 31309	17.74	< .001
Order of Merit List rankings	.14	.02	274, 14,027	7.61	< .01
		Flight performance			
Daily flying average	.30	.09	473, 4,844	45.03	< .001
Total hours to flight proficiency	.19	.04	22,795, 610,929	14.89	< .001

Note. df are rounded to the nearest whole number. Decimals are rounded to the nearest hundredth.

Table 19

Full Regression Analysis: AAInfo

	Predictor Army Aviation Information Composite (AAInfo)				
	R	*R* squared	*df*	*F*	Significance
Criterion variables					
			Classroom performance		
Academic average	.13	.02	567, 31,947	8.19	< .001
Order of Merit List rankings	.24	.05	21, 347	23.43	< .001
			Flight performance		
Daily flying average	.37	.14	729, 4,588	73.28	< .001
Total hours to flight proficiency	.12	.02	9,603, 624,121	6.14	< .01

Note. df are rounded to the nearest whole number. Decimals are rounded to the nearest hundredth.

Stepwise Regression Analyses

Follow-up analyses were conducted to investigate which SIFT subscales best predict IERW classroom grades and flight performance. To answer this question, forward stepwise regression analyses with inclusion criteria were conducted. The Durbin-Watson test for autocorrelation indicated that the residual plots for these analyses were randomly dispersed, and therefore, no cross-validation analyses were performed. Stepwise analyses were deemed to be the most appropriate statistical test due to the large number of predictor variables. Each of the seven SIFT subscales were placed into the model as independent variables. Because part of this revalidation effort was to determine if SIFT scores can predict success in the IERW course, IERW classroom grades (i.e., Academic Average and Order of Merit List rankings) and flight performance (i.e., Daily Flying Average and Total Hours to Flight Proficiency) were identified in the model as dependent variables. All statistical analyses are provided in Table 20.

In assessing which SIFT subscales best predict Academic Average, the Mechanical Comprehension subscale was the only predictor to be entered into the model; no other predictor accounted for unique variance [$F(806, 3{,}179) = 11.72$, $p < .001$]. In looking at the practical significance of the Mechanical Comprehension subscale in predicting Academic Average, the correlation coefficient was .16, indicating that 3% of the variance of Academic Average can be accounted for by IERW students' grades on the Mechanical Comprehension subscale. The Army Aviation Information test accounted for all the unique variance in Order of Merit List rankings, no other subscales were entered into the model [*F*(812, 13,489) = 23.43, *p* < .001] with 6% of the variance of Order of Merit List rankings being accounted for by AAInfo. The combination of the Army Aviation Information test, the Spatial Apperception subscale, and the Math Skills subscale accounted for the most unique variance in Daily Flight Averages [*F*(1,048, 4,270) = 37.54, *p* < .001] with 20% of the variance being accounted for. The Mechanical Comprehension subscale was the only predictor to be entered into the model for Hours to Flight Proficiency [*F*(29,128, 604,595) = 19.22, *p* < .001] with 5% of the variance being accounted for.

Table 20

Stepwise Regression Analyses (Forward Selection with Inclusion Criteria): IERW Classroom Grades and Flight Performance

	R	*R* squared	*df*	*F*	Significance
		Classroom performance			
Academic average					
Mechanical comprehension subscale	.16	.03	806, 3,179	11.72	< .001
Order of Merit List rankings					
Army aviation information test	.24	.06	812, 13,489	23.43	< .001
		Flight performance			
Daily flight average					
Army aviation information test AND Spatial apperception subscale AND Math skills subscale	.44	.20	1,048, 4,270	37.54	< .001
Total hours to flying proficiency					
Mechanical comprehension subscale	.21	.05	29,128, 604,595	19.22	< .001

Note. df are rounded to the nearest whole number. Decimals are rounded to the nearest hundredth.

Discussion

This revalidation study was designed to replicate the initial validation study with respect to the predictor and criterion variables analyzed. Both were similar regarding the demographic breakouts of age, ethnicity, and gender. The revalidation study had a slightly higher percentage of WOs than the initial validation study. However, this revalidation study differed from the initial validation study in that the check ride grades, the number of setbacks, and Behavioral Summary Scales (BSS) were not included as criterion variables.

For this study, we were interested in using more objective measures of classroom grades and flight performance and opted not to collect and analyze the check ride grades, setbacks, and the BSS measures, as they were deemed to be more subjective measures. Further, this study included an analysis into the testing completion times and the location of testing variables in order to determine if there were any predictive value of these variables, which there was none found. With one exception, intercorrelations were lower for this study than the initial validation study (the relationship between Math Skills with AAInfo increased by a value of 0.01). Also, across the board, regression analyses showed higher values in the initial validation study than in this revalidation study. One other difference in the revalidation analyses is that we investigated the possible impact of the Order of Merit List rankings. Although the regression analyses (see Tables 16–20) show significant relationships between many of the criterion variables and the predictor variables in those tables, it is important to observe the "*R* squared" values, as they show that these relationships do not account for much of the total variance in performance.

As a review, we re-address the following questions that were discussed at the beginning of this report, with their findings:

- How well does the SIFT exam total score predict IERW classroom grades and flight grades? **ANSWER**: Better for daily flight grades.
    - The SIFT exam total score provided weak to negligible prediction of IERW classroom grades.
        - Basic Warfighter Skills (r = .12; negligible relationship)
        - Common Core Proficiency (r = .26; weak relationship)
        - Academic Average (r = .16; negligible relationship)
    - The SIFT exam total score provided moderate prediction of IERW daily flight grades (r = .33)

- Which SIFT subscales are the best predictors of IERW classroom grades and flight grades? **ANSWER**: The AAInfo test and AACog composite are moderate predictors of daily flight grades.
    - AAInfo test
        - Weak predictor of IERW Class Grades–Basic Instruments (r = .29)
        - Weak predictor of Order of Merit List ranking (r =.24)
        - Moderate predictor of Daily Flight Grade (r = .37)
    - AACog composite
        - Weak predictor of IERW Class Grades–Common Core Proficiency 2 (r = .20)

        - Weak predictor of Order of Merit List ranking (*r* = .20)
        - Moderate predictor of Daily Flight Grade (*r* = .30)

- Were there differences in the SIFT exams prediction of Commissioned Officers (2LT) and Warrant Officers (WO1) success? **ANSWER**: A few findings stood out.
    - The SIFT exam total score was a better predictor of Academic Average for 2LTs (*r* = .33; moderate relationship) than for WO1s (*r* = .07; negligible relationship)
    - The AAInfo test was a strong predictor for Order of Merit List rankings for 2LTs (*r* = .43)
    - Combined, the subscales of Army Aviation Information test and Spatial Apperception test were strong predictors of IERW Daily Flight Grades
        - WO1s Daily Flight Grade (*r* = .42)
        - 2LTs Daily Flight Grade (*r* = .45)
    - Combined, the subscales of Reading Comprehension, Mechanical Comprehension, and Simple Drawings were strong predictors of IERW Class Grades for 2LTs (*r* = .41)

- Were there differences in SIFT exam predictions among testing sites? **ANSWER**: No differences were found.

Because flight training is expensive and there are a limited number of openings for new students, it is necessary to screen the applicants to ensure that only those persons with the capabilities to succeed in flight school are accepted for training. People who score higher on the SIFT exam total score are generally more successful in flight training than those who score lower.

According to the Organization and Personnel Force Development–SIFT Policy Memo (U.S. Army Aviation Center of Excellence, 2012), the SIFT is intended to be a test that measures those special aptitudes, personality, and background characteristics that are thought to be predictive of success in Army helicopter flight training. It was designed to predict aptitude and success for flight training school. Further, the SIFT exam should be able to assist the IERW cadre in assigning aviators, during the early phases of training, into their designated Go-To-War Aircraft and associated training courses. An initial plan for this study was to determine if the SIFT exam had enough merit to predict a student's placement in the Go-To-War Aircraft that best aligned with their flight training performance. However, based upon the generally weak correlation and regression findings of this revalidation study with regard to predicting a student's initial performance in rotary wing aviation training, the existing procedures for assignment of the Go-To-War Aircraft should continue as is.

One of the limitations of aviation research is the skewed focus on cognitive factors versus non-cognitive factors of performance (Paullin et al, 2006). With this said, the revalidation was heavily based on the cognitive outcomes of academic performance, procedural knowledge, and facility of using aircraft systems. The SIFT was not designed to measure non-cognitive factors such as personality. It is beyond the scope of this revalidation study to make any statements with regard to the relationship between non-cognitive factors and IERW performance.

References

Aviation Week. (2008). U.S. Army Aviation accidents rise in deadliness, cost. *Aerospace Daily & Defense Report, 228*, 3.

Bruskiewicz, K. T., Houston, J. S., Paullin, C. J., O'Shea, G., & Damos, D. (2006). Predictor development and pilot testing. In J. S. Houston & K. T. Bruskiewicz (Eds.), *Development and preliminary validation of a Selection Instrument for U.S. Army Flight Training (SIFT): Volume 1* (pp. 81–120). Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (Institute Report 523).

Bruskiewicz, K. T., Houston, J. S., Paullin, C. J., O'Shea, G., & Damos, D. (2007). *Predictor development and pilot testing of a prototype selection instrument for Army flight training.* (Technical Report No. 1195). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Department of the Army (2015). *Army Accident Investigations and Reporting*. (Headquarters Pamphlet 385-40). Fort Belvoir, VA: Director, Army Safety (DACS-SF).

Houston, J. S. (2006). Background and overview. In J. S. Houston & K. T. Bruskiewicz (Eds.), *Development and preliminary validation of a Selection Instrument for U.S. Army Flight Training (SIFT): Volume 1* (pp. 1–2). Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (Institute Report 523).

Houston, J. S., & Bruskiewicz, K. T. (Eds.) (2006). *Development and preliminary validation of a Selection Instrument for U.S. Army Flight Training (SIFT): Volume 1*. Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (Institute Report 523).

Katz, L. C. (2006). *Finding the 'right stuff': Development of an Army aviator selection instrument*. (OMB 0704-0188). Fort Rucker, AL: United States Army Research Institute for the Behavioral and Social Sciences. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA481060.

Kubisiak, C., & Katz, L. C. (2006). *U.S. Army Aviator job analysis* (Technical Report 1189). Arlington, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Maine Army National Guard. (2015). *Becoming a warrant officer*. Retrieved from http://www.me.ngb.army.mil/careers/warrant/becoming/flightprocess.aspx

Martinussen, M. (1996). Psychological measures as predictors of pilot performance. A meta-analysis. *The International Journal of Aviation Psychology, 6*, 1–20.

O'Shea, G., Bruskiewicz, K. T., & Houston, J. S. (2006). Preliminary validation data collection and predictor and criterion analysis. In J. S. Houston & K. T. Bruskiewicz (Eds.), *Development and preliminary validation of a Selection Instrument for U.S. Army Flight Training (SIFT): Volume 1* (pp. 129–166). Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (Institute Report 523).

O'Shea, G., Oppler, S., & Houston, J. S. (2006). Validation analyses. In J. S. Houston & K. T. Bruskiewicz (Eds.), *Development and preliminary validation of a Selection Instrument for U.S. Army Flight Training (SIFT): Volume 1* (pp. 167–197). Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (Institute Report 523).

Paullin, C. J., Bruskiewicz, K. T., Houston, J. S., & Damos, D. (2006). Review of the literature and identification of a selection strategy. In J. S. Houston & K. T. Bruskiewicz (Eds.), *Development and preliminary validation of a Selection Instrument for U.S. Army Flight Training (SIFT): Volume 1* (pp. 3–68). Minneapolis, MN: Personnel Decisions Research Institutes, Inc. (Institute Report 523).

U.S. Army Aviation Center of Excellence. (2012). *Policy memorandum for the Selection Instrument for Flight Training (SIFT)*. Fort Rucker, AL: Director, Organization and Personnel Force Development (OPFD) Directorate (ATZQ-CDF-P).

U.S. Army Aviation Center of Excellence. (2014). *USAACE Mission*. Retrieved from http://www.rucker.army.mil/

U.S. Army Aviation Center of Excellence. (2015). *Student information*. Retrieved from http://www.rucker.army.mil/newcomers/students/

Wiener, S. (2005). *Military Flight Aptitude Tests (6th Ed.)*. Lawrenceville, NJ: Thomson Peterson.

Appendix A

SIFT Test Item Examples

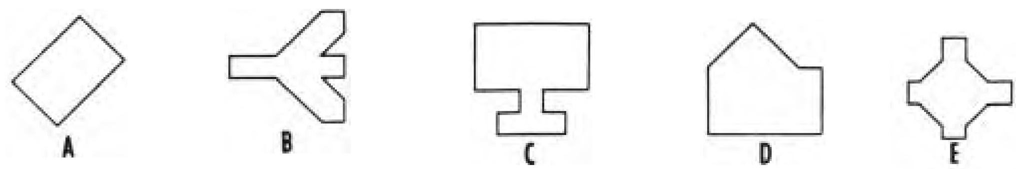
(Adapted from Wiener, 2005)

<u>**PERCEPTUAL SPEED AND ACCURACY—HIDDEN FIGURES**</u>

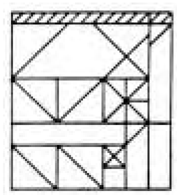
Questions on hidden figures are designed to measure your ability to see simple figures in complex drawings. Although these figures are fairly well camouflaged, proper visualization should enable you to discern them without too much difficulty.

At the top of each section of this subtest are five figures lettered A, B, C, D, and E. Below these on each page are several numbered drawings. You must determine which lettered figure is contained in each of the numbered drawings.

The lettered figures are shown below:



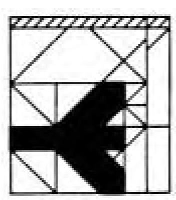
As an example, look at drawing X below:



X

Which one of the five figures is contained in drawing X?

Now look at drawing Y, which is exactly like drawing X except that the outline of figure B has been shaded to show where to look for it. **Thus, B is the answer to sample item X**.



Y

Each numbered drawing contains only one of the lettered figures. The correct figure in each drawing will always be of the same size and in the same position as it appears at the top of the page. Therefore, do not rotate the page in order to find it. Look at each numbered drawing and decide which one of the five lettered figures is contained in it.

PERCEPTUAL SPEED AND ACCURACY—SIMPLE DRAWINGS

This section is pretty straightforward, you are given 5 pictures and you have to decide which one is not like the other.

For example;

B B B b B **("b" would be the correct answer**.)

Or, circle the different item. (**The fourth clock would be the correct answer**).

ARMY AVIATION INFORMATION

(Helicopter parts, types, operations; flight rules, weather)

- The primary purpose of the tail rotor system is to

(A) assist in making a coordinated turn.

(B) maintain heading during forward flight.

(C) counteract the torque effect of the main rotor.

(D) provide additional thrust and lift.

(E) increase maximum speed.

The correct answer is (C). The auxiliary or tail rotor is the anti-torque rotor that produces thrust in the direction opposite to the torque reaction developed by the main rotor.

- During a hover, a helicopter tends to drift in the direction of tail rotor thrust. This movement is called

(A) flapping.

(B) gyroscopic precession.

(C) transverse flow effect.

(D) translating tendency.

(E) Coriolis force.

The correct answer is (D). The entire helicopter has a tendency to move in the direction of tail rotor thrust when hovering. This movement is generally referred to as translating tendency or drift.

- A lighted heliport may be identified by

(A) a flashing yellow light.

(B) a blue lighted square landing area.

(C) white and red lights.

(D) a green, yellow, and white rotating beacon.

(E) blue and red alternating flashes.

The correct answer is (D). The color combination of green, yellow, and white flashed by beacons indicates a lighted heliport.

- The most favorable conditions for helicopter performance are the combination of

(A) low-density altitude, light gross weight, and moderate to strong winds.

(B) high-density altitude, heavy gross weight, and calm or no wind.

(C) low-density altitude, light gross weight, and calm or no wind.

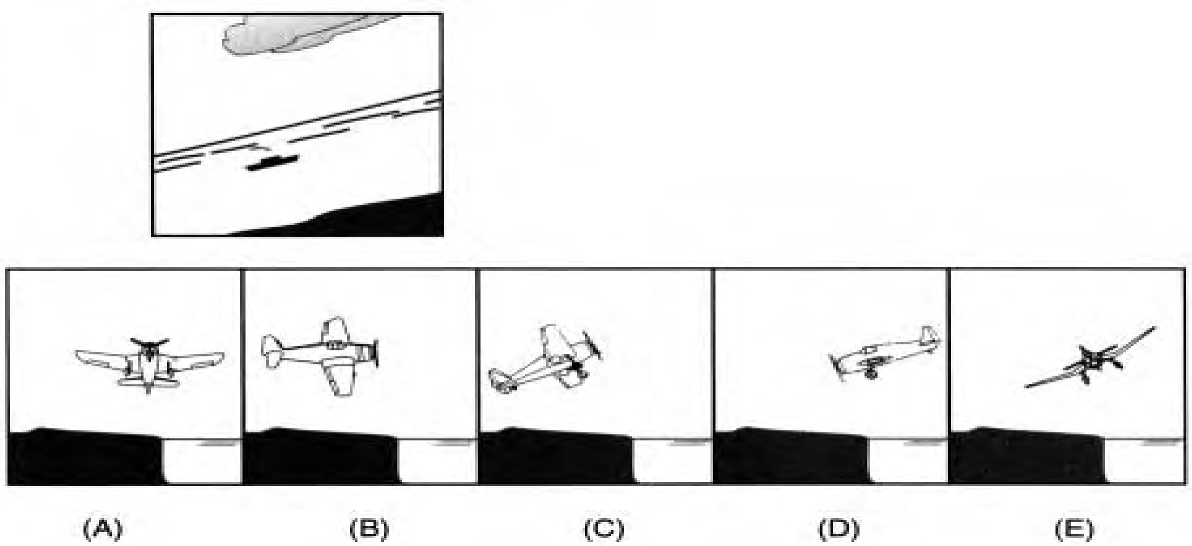
(D) high-density altitude, light gross weight, and moderate to strong winds.

(E) low-density altitude, heavy gross weight, and moderate to strong winds.

The correct answer is (A). The most favorable conditions for helicopter performance are the combination of a low-density altitude, light gross weight, and moderate to strong winds. The most adverse conditions are the combination of a high-density altitude, heavy gross weight, and calm or no wind. Any other combination of density altitude, gross weight, and wind conditions fall somewhere between the most adverse conditions and the most favorable conditions.

SPATIAL APPERCEPTION

Each problem in this test consists of six pictures: an aerial view at the upper left and five pictured choices below labeled (A), (B), (C), (D), and (E). Each pictured choice shows a plane in flight. The picture at the upper left shows the view that the pilot would have looking straight ahead from the cockpit of one of the five pictured planes. Determine which of the five lettered sketches most nearly represents the position or attitude of the plane and the direction of flight from which the view would have been seen.



The correct answer is (B). The plane is shown in the position from which the pilot would have seen through the windshield of the cockpit—the view shown in the upper left aerial view. The plane is shown on a level flight, banking right, and flying out to sea.

COGNITIVE ABILITIES TEST—READING COMPREHENSION

The rates of vibration perceived by the ears as musical tones lie between fairly well-defined limits. In the ear, as in the eye, there are individual variations. However, variations are more marked in the ear, since its range of perception is greater.

The paragraph best supports the statement that the ear

(A) is limited by the nature of its variations.

(B) is the most sensitive of the auditory organs.

(C) differs from the eye in its broader range of perception.

(D) is sensitive to a great range of musical tones.

(E) depends for its sense on the rate of vibration of a limited range of sound waves.

The correct answer is (C). The passage makes the point that individual differences in auditory range are greater than individual differences in visual range because the total range of auditory perception is greater. Although the statements made by choices (D) and (E) are both correct, neither expresses the main point of the reading passage.

COGNITIVE ABILITIES TEST—MATHEMATICAL SKILLS

The numerical value of 4! is

(A) 8

(B) 12

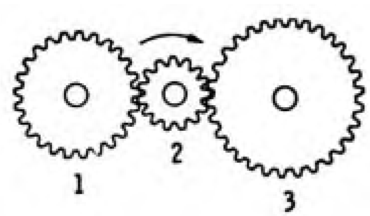
(C) 16

(D) 20

(E) 24

The correct answer is (E). The factorial of a natural number is the product of that number and all the natural numbers less than it. 4! = 4 x 3 x 2 x 1 = 24.

COGNITIVE ABILITIES TEST—MECHANICAL COMPREHENSION



Which of the other gears is moving in the same direction as gear 2?

(A) Gear 1

(B) Gear 3

(C) Neither of the gears

(D) Both of the gears

(E) It cannot be determined.

The correct answer is (C). Gear 2 is moving clockwise and is causing both gear 1 and gear 3 to move counterclockwise.

Appendix B

Courses by Training Stage—Phase 1

Training Stage	**Courses**
Stage 1: Pre-flight	Theory of Rotary Wing Flight Navigation Weather Flight Support Subjects TH-67 Systems, Parts I & II Aviation Medicine
Stage 2: Primary	Instruments, Part I
Stage 3: Instruments	Instruments, Part II, III, & IV Crew Coordination Tactics Fundamentals Terrain Flight Operations
Stage 4: Basic Warfighting Skills	Target Identification Basic Fire Support Aerial Artillery Adjustment Aircraft Survivability Equipment Identification, Friend or Foe Army Airspace Command and Control Single-Channel Ground and Airborne Radio System

Appendix C

SIFT Subscale Correlations

Initial Validation Study (Bruskiewicz, Houston, Paullin, O'Shea, & Damos, 2006).

	Simple Drawings Subtest	**Hidden Figures Subtest**	**Army Aviation Information**	**Mechanical Comprehension Subtest**	**Math Skills Subtest**	**Reading Comprehension Subtest**	**Spatial Apperception Subtest**
Simple Drawings Subscale	—						
Hidden Figures Subscale	**.42****	—					
Army Aviation Information Test	**.16***	**.17***	—				
Mechanical Comprehension Subscale	**.31****	**.27****	**.41****	—			
Math Skills Subscale	**.33****	**.37****	**.22****	**.58****	—		
Reading Comprehension Subscale	**.19****	**.26****	**.32****	**.46****	**.45****	—	
Spatial Apperception Subscale	**.10**	**.24****	**.48****	**.26****	**.15***	**.22****	—

Note. *N*s between 231 and 240. **p* < .05, one tailed. ***p* < .01, one tailed. Numbers rounded to nearest hundredth.

SIFT Subscale Inter-correlations—Present Revalidation Study

	Simple Drawings Subtest	Hidden Figures Subtest	Army Aviation Information	Mechanical Comprehension Subtest	Math Skills Subtest	Reading Comprehension Subtest	Spatial Apperception Subtest
Simple Drawings Subscale	**—**						
Hidden Figures Subscale	**.32****	**—**					
Army Aviation Information Test	**-.01**	**.06**	**—**				
Mechanical Comprehension Subscale	**.12***	**.24****	**.31****	**—**			
Math Skills Subscale	**.18****	**.32****	**.08**	**.51****	**—**		
Reading Comprehension Subscale	**.13****	**.21****	**.17****	**.38****	**.34****	**—**	
Spatial Apperception Subscale	**.03**	**.19****	**.30****	**.22****	**.09***	**.10***	**—**

Note. N=463. *$p < .05$, one tailed. **$p < .01$, one tailed. Numbers rounded to nearest hundredth.