AFCAPS-TR-2020-000X

Validation of the Pilot Candidate Selection Method (PCSM)

USAF Strategic Personnel Research Program

June 2020

John D. Trent

Imelda D. Aguilar, Ph.D.

Prepared for:

Katie Gunther, Ph.D.

AFPC/Strategic Research and Assessment Branch (SRAB)

Air Force Personnel Center
Strategic Research and Assessment
HQ AFPC/DSYX
550 C Street West, Ste 45
Randolph AFB TX 78150-4747

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report was cleared for release by HQ AFPC/DSYX Strategic Research and Assessment Branch (SRAB) and is releasable to the Defense Technical Information Center.

This report is published as received with minor grammatical corrections. The views expressed are those of the authors and not necessarily those of the United States Government, the United States Department of Defense, or the United States Air Force. In the interest of expediting publication of impartial statistical analysis of Air Force tests SRAB does not edit nor revise Contractor assessments appropriate to the private sector which do not apply within military context.

Federal Government agencies and their contractors registered with Defense Technical Information Center should direct request for copies of this report to:

Defense Technical Information Center - http://www.dtic.mil/

Approved for public release, unlimited distribution by AFPC/DSYX Strategic Research and Assessment Branch, Joint Base San Antonio-Randolph AFB, TX 78150-4747 or higher DoD authority. Please contact AFPC/DSYX Strategic Research and Assessment Branch (SRB) with any questions or concerns with the report.

This paper has been reviewed by the Air Force Center for Applied Personnel Studies (AFCAPS) and is approved for publication. AFCAPS members include: Senior Editor Dr. Thomas Carretta AFMC 711 HPW/RHCI and Dr. Imelda Aguilar HQ AFPC/DSYX.

REPORT DOCUMENTATION PAGE

Form Approved

OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
06-17-2020	Technical Report	2009-2020

4. TITLE AND SUBTITLE

Validation of the Pilot Candidate Selection Method (PCSM)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

John D. Trent, Imelda D. Aguilar

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

HQ AFPC/DSYX

8. PERFORMING ORGANIZATION REPORT

AFCAPS-TR-2020-0002

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Personnel Center

Strategic Research and Assessment Branch

10. SPONSOR/MONITOR'S ACRONYM(S)

HQ AFPC/DSYX

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFCAPS-TR-2020-0002

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release. Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

For this initiative, the Pilot Candidate Selection Method (PCSM), which combines several personnel tests weighted to optimally predict pilot training outcomes such as attrition and final course grades, was re-evaluated using updated training data to determine if it is still the most salient predictor of pilot training outcomes out of the available pilot-selection tests within the United States Air Force (USAF). Results suggest that it continues to be a strong predictor of success in completing Specialized Undergraduate Pilot Training (SUPT), as well as other important pilot training outcomes.

15. SUBJECT TERMS

Aviation Psychology, Testing, Adverse Impact, Air Force Officer Qualifying Test, AFOQT, Test of Basic Aviation Skills, TBAS, Pilot Candidate Selection Method, PCSM

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified					Katie Gunther, Ph.D.
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	**U**	**U**	**U**	55	210-565-5245

Standard Form 298 (Rev. 8-98)

Prescribed by ANSI Std. Z39.18

Table of Contents

Executive Summary

The identification of Air Force pilot candidates most likely to excel as Air Force pilots has been a long-established goal (Carretta & Ree, 1994). As more modern, higher-performing aircraft are integrated in the US Air Force (USAF), there is a demand for higher levels of psychomotor coordination and cognitive and perceptual abilities from Air Force pilots. To assess whether pilot candidates possess the cognitive and psychomotor skills to successfully complete pilot training, the USAF operationalized the Pilot Candidate Selection Method (PCSM) in 1993. Although several changes have been made to its components, it is still in use today. The PCSM score is a weighted composite of the Air Force Officer Qualifying Test (AFOQT) Pilot composite, several Test of Basic Aviation Skills (TBAS scores), and a zero- to 9-point pilot flight hour code produce the PCSM score. It is reported as a percentile score with values ranging from 1 to 99. The PCSM weighted composite score is optimized to predict completion of Specialized Undergraduate Pilot Training (SUPT) T-6 training.

Since its implementation, various studies have been conducted to evaluate its effectiveness in predicting pilot training outcomes. PCSM scores have demonstrated validity against various measures of flying performance, flying grades, class rank, and the number of hours required to complete training (Carretta, 2006). High PCSM scores have been found to be associated with an increased probability of completing flight training (Carretta 1992a, 1992b), a decreased number of flying hours required to complete training (Duke & Ree, 1996), and a significant probability of being fighter-qualified (Weeks, Zelenski, & Carretta, 1996). For this initiative, the PCSM composite was re-evaluated using updated training data to determine if it is still the most salient predictor of pilot training outcomes out of the available pilot-selection tests

within the USAF. Results suggest that it continues to be a strong predictor of success in completing SUPT, as well as other important pilot training outcomes.

In addition to a selection procedure being effective in predicting outcomes between selection and performance, it should also minimize discrimination and group differences to the extent possible. Given the current USAF pilot shortage and concerns about limiting diversity in the career field, an evaluation of group differences was also a focus of this initiative. According to Sackett and Ellingson (1997), "Group differences on a predictor or composite of predictors are only meaningful to the extent that they influence selection outcomes," (p.709). Adverse impact is demonstrated when a statistical disparity exists between the selection rates of majority and minority groups. Evidence of adverse impact exists if the minority group is selected at a rate less than 80% of the majority group, unless the test is valid, job-relevant, and other alternatives have been explored (Cascio & Aguinis, 2005). In comparison to the AFOQT Pilot composite alone, the PCSM composite has been found to have less adverse impact for pilot candidate selection for females and racial/ethnic minorities (Carretta, 2006). For this study, group differences are defined as a comparison of majority (e.g., racial majority, male) groups to legally-protected groups (e.g., racial minority, female). Potential alternatives that could reduce group differences were included in the analyses. Results for this study indicate the selection ratio based on gender if the PCSM was the only selection hurdle would be acceptable (80%), the selection ration based on race would be somewhat lower (72%), and there would be no evidence of adverse impact based on ethnicity (90%).

Finally, given the need for more efficient and convenient approaches to pilot selection, tests that would not require expensive and bulky peripheral devices (e.g., rudder pedals, test carrels) were explored. While an alternate test (e.g., the Air Force Multi-Tasking Test) was found

to result in fewer subgroup differences, the resulting algorithms when attempting to replace existing PCSM components were not as predictive of successful pilot training completion as the current model. In addition, attempts to reduce subgroup differences by incorporating individual multi-tasking components were unsuccessful.

Overall, the PCSM composite continues to demonstrate value in predicting multiple manned and unmanned pilot training outcomes. The existing PCSM score is the most valid of all the predictors evaluated in this study as related to trainee completion through SUPT Primary (T-6) training. Although alternatives that do not require the use of the testing carrel, the joystick, and the rudders are available and demonstrate similar validity and reasonably comparable subgroup differences from a scientific/research perspective, the current PCSM score is working as designed and is slightly better than other alternatives in terms of maximizing validity and minimizing subgroup differences to the extent possible.

Background

The Pilot Candidate Selection Method (PCSM) was operationally implemented in 1993 for pilot trainee selection in the United States Air Force (USAF). The current components of the PCSM composite include the Air Force Officer Qualifying Test (AFOQT) Pilot composite, several scores from the Test of Basic Aviation Skills (TBAS), and a zero- to 9-point pilot flight hour code, to be described more fully in the sections that follow. The scoring uses a regression-weighted algorithm optimized to predict flight training completion through USAF Initial Flight Training (IFT) and Primary Specialized Undergraduate Pilot Training (SUPT). In addition, a non-operational, modified version of the SynWin Multi-Tasking Test (Elsemore, 1994; Oswald et al., 2007) has been included in the TBAS battery since 2012 for potential inclusion in the PCSM.

As a best practice in selection test development and maintenance, the tests should be updated and/or revalidated on a regular basis. The AFOQT Form S was developed in 2005 and was replaced by Form T in 2015. The Test of Basic Aviation Skills (TBAS) replaced the Basic Attributes Test (BAT) in 2006. The validity of the PSCM composite for manned aircraft pilot trainees was last formally evaluated in 2011 (Carretta, 2011), and for unmanned aircraft pilot trainees in 2014 (Rose, Barron, Carretta, Arnold, & Howse, 2014). Another validity study examined the predictive validity of the AFOQT pilot and PCSM composites for Specialized Undergraduate Pilot Training (SUPT) and found that both the AFOQT and PCSM composites exhibited good predictive validity (Carretta, 2013).

In addition to test validity, selection tests must also be fair and unbiased. Adverse impact is the negative result that a selection procedure may have on a protected group. This occurs when

there is a statistical difference between the selection rates of the majority and minority groups. Specifically, the potential for adverse impact exists when the selection rate of a minority group is less than 80% of the majority group. When adverse impact exists, the test user must demonstrate that the test is valid, job-relevant, and other alternatives have been explored (Cascio & Aguinis, 2005). While research has found cognitive tests have the highest predictive validity for training and job performance when compared to other personnel selection methods, cognitive tests have also been found to indicate greater group differences in test performance favoring Whites over other racial minorities. As with other cognitive ability tests, the PCSM composite also has the potential to result in adverse impact even though the program strives to avoid adverse impact to the fullest extent possible.

Thus, the purposes of the current study were to1) conduct a revalidation of the current PCSM model using updated training data, 2) evaluate the possibility of modifying the PCSM components to reduce differences between minority and majority applicants, and 3) potentially eliminate the bulky and expensive peripheral devices (e.g., joystick, rudder pedals).

Method

Participants

The study includes samples for both manned and unmanned aircraft. The manned aircraft sample included 888 pilot trainees who attended SUPT training and also had PCSM and SynWin Multi-Tasking test scores on record (see Table 1). The sample was predominantly male (87.0%) and White (86.5%), which differs from the overall officer candidate pool prior to the selection hurdles including the AFOQT Pilot composite requirement of a score of 25 or higher and the PCSM composite requirement of a score of 10 or higher (74% male; 67% White). It should be noted that it is unknown how many of the overall officer applicants were interested in becoming a pilot.

Table 1
Demographic Data for IFT/SUPT Participants

	Total (N = 888)	
	N	Percent
Education Level		
High School/GED	18	2.0
Some College	574	64.6
Bachelor's Degree	265	29.9
Master's Degree	29	3.3
Doctorate	2	0.2
Sex		
Female	115	13.0
Male	772	87.0
Ethnicity		
Hispanic or Latino	86	9.7
Not Hispanic or Latino	802	90.3
Race		
American Indian or Alaska Native	13	1.5
Asian	25	2.8
Black/African American	35	3.9
White	768	86.5
Hawaiian/Pacific Islander	5	0.6
Multiple/Other	42	4.7

Note. Sample sizes vary due to missing data (e.g., for individuals who declined to respond).

The remotely piloted aircraft (RPA) sample included 449 RPA trainees who attended RPA Flight School (RFS) and also had PCSM and SynWin Multi-Tasking test scores on record (see Table 2). The sample was predominantly male (92.0%) and White (87.5%). As with the manned aircraft sample, this sample differs from the overall officer candidate pool prior to the AFOQT and PCSM selection hurdles (74% male; 67% White).

Table 2
Demographic Data for Remotely-Piloted Aircraft Participants

	Total (N = 449)	
	N	Percent
Education Level		
High School/GED	7	1.6
Some College	225	50.1
Bachelor's Degree	194	43.2
Master's Degree	21	4.7
Doctorate	2	0.5
Sex		
Female	36	8.0
Male	412	92.0
Ethnicity		
Hispanic or Latino	49	10.9
Not Hispanic or Latino	399	88.9
Race		
American Indian or Alaska Native	8	1.8
Asian	7	1.6
Black/African American	20	4.5
White	393	87.5
Hawaiian/Pacific Islander	8	1.8
Multiple/Other	13	2.9

Note. Sample sizes vary due to missing data (e.g., for individuals who declined to respond).

Larger samples of varying sizes were used for supplemental analyses including the evaluation of overall SUPT attrition and subgroup differences in test scores.

Predictor Measures

Air Force Officer Qualifying Test (AFOQT) Pilot Composite. The AFOQT Pilot composite measures job-relevant constructs including instrument comprehension, table reading

(i.e., perceptual speed and accuracy), aviation information, and math knowledge. The AFOQT Pilot composite is used to qualify candidates for rated officer pilot training. Candidates must receive a minimum percentile score of 25 to be eligible (AFMAN 36-2664, 2019).

Various studies have examined the relationship between AFOQT scores and future officer and pilot training success. Correlations greater than .21 between predictors and outcome variables are widely considered as evidence that a selection test is likely to be a valid predictor of the outcome (U. S. Department of Labor, 1999). Studies on the AFOQT have demonstrated this level of validity or greater. The most comprehensive study of AFOQT validity (Arth, 1986) found the AFOQT composites have significant relationships with final course grades in most career-field training programs. Other studies also found that the AFOQT predictive validity generalizes to an extensive variety of officer jobs (Carretta, 2009; Hardison, Sims, & Wong, 2010). For example, studies have found a strong and direct relationship between AFOQT and training scores for pilot trainees (Carretta, 2005) demonstrating that using AFOQT for pilot training selection leads to better training outcomes and job performance. Specifically, the AFOQT Pilot composite has been found by Carretta (2005) to be highly predictive of SUPT Primary completion (pass/fail) and final grades (r = .31 and r = .34, respectively). In relation to bias and diversity, results indicate the AFOQT is not biased against females or protected racial minorities (i.e., there are no differences in validity across groups), but it does tend to reduce diversity with larger number of women and minorities being rejected when compared to white and male applicants (Hardison, Sims, & Wong, 2010). In previous research by EASI Consult, Schwartz, and Weissmuller (2008), the AFOQT Pilot composite has been shown to result in substantial group differences (e.g., d = .98 for males and females; d = 1.52 for White and Black officer candidates). In terms of test reliability, coefficient alpha reliability estimates for the four

subtests are α = .91 for Instrument Comprehension, α = .88 for Table Reading, α = .79 for Aviation Information, and α = .84 for Math Knowledge (Aguilar, 2017; Carretta, Rose, & Trent, 2016).

Test of Basic Aviation Skills (TBAS). The TBAS is a computer-administered battery that consists of job-relevant subtests that assess constructs such as psychomotor skills, spatial orientation, and multitasking. Specifically, the TBAS includes metrics such as keeping the airplane on target (A), measuring the average distance in pixels if the airplane is not on target (AOnTarget and AHOnTarget), measuring rudder control using the average distance in pixels from target (AHROnTarget), and a spatial rotation composite based on the location of an unmanned aerial vehicle in relation to landmarks (UAV Composite). Scores from each section of the TBAS are used as weighted components of the PCSM algorithm. Carretta (2005) performed several analyses to examine the incremental validity of TBAS scores in predicting SUPT T-37 performance criteria and their incremental validity when used with the AFOQT composite scores and previous flying experience. Several subtests from the TBAS indicated predictive validity against T-37 performance, but most of the subtests failed to demonstrate incremental validity "beyond a baseline pilot candidate selection model that included the AFOQT Pilot and AFOQT Quantitative composites and a measure of previous flying experience" (Carretta, 2005, p. 15). In 2011, Carretta evaluated the new PCSM composite, changed in 2005, to assess its predictive validity against the USAF pilot training performance. Statistically significant relationships were found between the TBAS composite and academic average (r = .21), daily flying average (r = .31), check flight average (r = .15), and T-6 average (r = .28).

SynWin Multi-Tasking Test (Modified). To address a gap in research that shows performance on a multitasking battery consisting of primarily cognitive tasks as being more

predictive of pilot performance outcomes than single-task components, Barron and Rose (2017) compared the validity of a pre-employment multitasking assessment (consisting of math, memorization, and monitoring of tasks) to an assessment in which the same tasks were assessed separately. The SynWin Multi-Tasking Test indicated significant positive predictions of academic and flying performance in training. This test, originally developed by the Navy, assesses performance on multiple concurrent tasks of Memorization, Math, Visual Monitoring, and Listening. Although pilot SMEs generally agree that memorization, math, visual monitoring, and listening are important individual skills for effective piloting (with only math skills directly assessed in the current PCSM), part of the intent in the USAF evaluation of SynWin was to potentially assess individual differences in multi-tasking (i.e., O*NET ability of "time sharing") that would not be assessed by performance of serial individual tasks (Paullin, Ingerick, Trippe, & Wasko, 2011). Results revealed that while single-task performance was not significantly related to flying performance in training (*r*s = .00 - .03), a 2-minute multitasking assessment allowed for significant prediction of flying performance across sorties in a 22-week pilot training course (*r*s = .19 - .21) (Barron & Rose, 2017). Results from the study also indicated that when tasks were presented separately as opposed to simultaneously, decrements associated with multitasking were significantly negatively related to flying performance (*r* = -.23) (Barron & Rose, 2017). Coefficient alpha reliability estimates for the SynWin (Modified) are α = .80 for Listening, α = .90 for Math, α = .89 for Memorization, α = .97 for Visual Monitoring, and α = .93 for all four tasks (Barron, 2015).

Flight Hour Code. A component of the PCSM score was determined by FAA logged flying hours using a 9-point interval scale ranging from 0 flying hours to over 200 flying hours. Approximately half of the applicants (19,819 out of 37,506; 52.8%) had no previous flying

experience. Previous research by Carretta (2011) has demonstrated that flying experience is predictive of success in SUPT completion (corrected r = .18).

Grade Point Average (GPA). Although not a component of PCSM, applicants provide the PCSM Program Office with their undergraduate GPA at the time that they take the TBAS. GPA is often viewed as a proxy for overall cognitive ability and considered by rated boards when reviewing the applicant's board package. While GPA is not included in the PCSM composite, it is included in the current study to evaluate if GPA would be beneficial in predicting pilot training outcomes or if inconsistencies in GPA across institutions result in an inaccurate measure of aptitude and therefore a poor predictor of future performance.

Outcome Measures

Primary and Advanced Specialized Undergraduate Pilot Training (SUPT) Performance. SUPT Primary is the phase of training where all trainees learn to fly the same aircraft (i.e., the T-6A airframe). SUPT Advanced training is the phase where trainees are classified into the fighter/bomber track (i.e., the T-38C airframe) or the airlift/tanker track (i.e., the T-1A airframe). Numerous performance criteria were evaluated. These outcomes included SUPT Primary completion/attrition (i.e., the primary outcome variable), as well as daily flying grades, academic grades, and the Merit Assignment Selection System Score (MASS) in SUPT Primary and Advanced phases. Completion of SUPT Primary (i.e., T-6A training) was a dichotomous variable (0 = failed IFT or SUPT Primary; 1 = passed SUPT Primary). Academic grades include performance in classes such as aircraft systems, mission planning, weather, and navigation. Daily flying grades include an average of all procedures and maneuvers during flight. The MASS score is the overall assessment of the student's airmanship and capability based on all indicators of performance (e.g., academic grades, daily grades, Flight Commander's Ranking).

Remotely Piloted Aircraft (RPA) Flight Screening (RFS) Performance and Remotely Piloted Aircraft (RPA) Instrument Qualification (RIQ) Performance. PCSM is also used for RPA selection. To validate the PCSM for unmanned aircraft training, numerous RPA performance criteria were also evaluated in this study. As with manned aircraft, these outcomes included daily flying grades, academic grades, and the MASS.

Control Measures

Socioeconomic Status. Socioeconomic status (SES) has been asked of participants who have taken the AFOQT since 2015. It is a self-report measure included in the demographics:

> "Compared with U.S. families in general, would you say your family income at the time you graduated from high school was…?
>
> A. Much higher than average
> B. Somewhat higher than average
> C. Average
> D. Somewhat lower than average
> E. Much lower than average"

Procedure

All participants were prospective pilot trainees who had previously passed the minimum cut scores for AFOQT Verbal (≥15), Quantitative (≥10) and Pilot (≥25) and a PCSM score or received waivers; and completed the TBAS (with attached SynWin Multi-Tasking Test) under operational and proctored conditions at TBAS stations. TBAS stations are located at 117 active duty United States Air Force Bases and Air Force Reserve Officer Training Corps detachments worldwide, at the United States Air Force Academy, and at all Military Entrance Processing Stations (MEPS). Flight training outcomes were provided by the 19th Air Force (Air Education and Training Command) for the purpose of validating and/or updating the PCSM. All data were evaluated in the aggregate and no individual scores or personally identifiable information (PII)

were shared outside of the Air Force Personnel Center/Strategic Research and Assessment Branch (AFPC/DSYX) research team.

Analyses

Validation of PCSM. Validity coefficients are presented as correlations between the predictor (e.g., the test) and the criterion of interest (e.g., training outcomes). Correlations are values ranging from 0.0 (i.e., no relationship) to 1.0 (i.e., a perfect one-to-one relationship). The U. S. Department of Labor (1999) has recommended the following guidelines for interpreting validity coefficients:

- *r* >.35 “Very Beneficial”
- *r* .21-.35 “Likely to be Useful”
- *r* .11-.20 “Depends on the Circumstances”
- *r* <.11 “Unlikely to be Useful”

For this validity study, zero-order correlations and correlations corrected for multivariate range restriction were generated to evaluate the relationship between the current PCSM score and the various pilot training outcomes. Specifically, a point-biserial correlation was generated between the PCSM score and a dichotomous variable for completion (1) versus attrition (0) through SUPT Primary.

Exploration of Alternatives. Numerous logistic regression models, as recommended by Warner (2013; e.g., forward selection, backward elimination, stepwise), were evaluated using different combinations of the most salient predictors (i.e., SynWin Multi-Tasking Test, TBAS, AFOQT, and Flight Hour Code) of the dichotomous SUPT completion/attrition variable. In addition, an evaluation of alternative coding schemes for FAA-approved flight hours to optimize validity while minimizing adverse impact was conducted.

Group Differences. Mean score group differences were evaluated using Cohen's (1988) *d* effect size. Cohen (1988) recommended the following guidelines for interpretation of *d* values:

- .2 = Small
- .5 = Medium
- .8 = Large

Bias. Mean scores differences across groups are often erroneously interpreted as indicating test bias. However, group differences in mean test scores can represent actual differences in the abilities being measured. In the current study, the potential for test bias was evaluated by comparing the criterion-related validity coefficients across sex and race to evaluate potential differences in validity across groups (i.e., differential validity).

Adverse Impact. Standards provided by the Uniform Guidelines (1978) were used to evaluate adverse impact. Specifically, this study used the 80% (4/5ths) rule where if the minority group is qualified at a rate less than 80% of the majority group, there is evidence of potential adverse impact.

Potential Contribution of Grade Point Average (GPA). The zero-order correlation between GPA and SUPT completion was evaluated. In addition, the validity coefficients of the current PCSM and AFOQT Pilot composite were compared to GPA as a point of reference.

Contribution and Alternatives to Current Flight Hour Code. Several analyses were conducted to evaluate the contribution of the current Flight Hour Code and potential alternatives. Specifically, zero-order correlations between Flight Hour code and multiple pilot-training outcomes were evaluated. In addition, observed and expected frequencies of FAA-approved fight hours were provided by group, as was differential validity to evaluate the potential for bias. Finally, alternatives to the current Flight Hour Code were evaluated for validity and the potential reduction of adverse impact.

Results

Validity

Intercorrelations of Predictors. As shown in Table 3, the intercorrelations among the predictors varied widely. Not surprisingly, tasks most similar to each other were more highly correlated (e.g., TBAS Airplane Redirects and TBAS Distance from Target). In addition, many of the correlations must be interpreted as part-whole correlations (e.g., the PCSM components with the PCSM composite score; the Multitasking components with the Multitasking Overall Score).

Table 3

Intercorrelations of Predictors

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. PCSM Score																	
2. Flight Hour Code	0.76																
3. AFOQT Pilot	0.84	0.40															
4. A (TBAS Airplane Redirects 1)	0.44	0.21	0.31														
5. AOnTarget (TBAS Distance from Target 1)	0.44	0.23	0.32	0.89													
6. AHA (Airplane Redirects 2)	0.45	0.25	0.32	0.70	0.68												
7. AHOnTarget (TBAS Distance from Target 2)	0.47	0.25	0.35	0.76	0.82	0.85											
8. AHROnTarget (TBAS Rudder Task)	0.47	0.23	0.26	0.43	0.43	0.40	0.38										
9. Unmanned Aerial Vehicle Composite	0.48	0.17	0.42	0.24	0.23	0.22	0.23	0.23									
10. Multitasking Trial 1 Mean	0.23	0.03	0.28	0.16	0.16	0.15	0.17	0.18	0.19								
11. Multitasking Trial 2 Mean	0.22	0.02	0.28	0.15	0.15	0.15	0.16	0.17	0.20	0.77							
12. Multitasking Trial 3 Mean	0.22	0.02	0.29	0.16	0.16	0.15	0.17	0.18	0.20	0.75	0.78						
13. Multitasking Trial 4 Mean	0.23	0.01	0.30	0.16	0.15	0.15	0.17	0.18	0.20	0.74	0.78	0.80					
14. Multitasking Listening	0.23	0.14	0.22	0.15	0.16	0.14	0.15	0.17	0.11	0.59	0.60	0.59	0.59				
15. Multitasking Math	0.15	-0.05	0.26	0.09	0.08	0.09	0.10	0.08	0.16	0.64	0.65	0.66	0.66	0.31			
16. Multitasking Memory	0.24	0.01	0.31	0.15	0.13	0.14	0.15	0.19	0.26	0.66	0.68	0.69	0.69	0.37	0.45		
17. Multitasking Visual	0.07	-0.03	0.09	0.10	0.10	0.09	0.11	0.10	0.07	0.57	0.58	0.58	0.57	0.20	0.21	0.25	
18. Multitasking Overall Score	0.25	0.02	0.31	0.17	0.17	0.17	0.18	0.20	0.22	0.89	0.92	0.92	0.91	0.65	0.72	0.75	0.63

N = 14,311

Validation of PCSM. As shown in Table 4, the PCSM model continues to be a salient predictor of completion through SUPT Primary (uncorrected $r = .39$, $p < .001$). The PCSM

validity coefficient is higher than or equal to all other predictors that were evaluated and would be interpreted as “Very Beneficial” according to the Department of Labor Guidelines. The AFOQT Pilot composite validity coefficient is slightly lower (uncorrected $r = .37$, $p < .001$), yet still a “Very Beneficial” of SUPT completion. The coefficients for the TBAS components were all significant with the most salient predictors involving both rudder and joystick components (*r*s ranged from .21 to .24). The TBAS components that were least related to SUPT completion included the UAV composite ($r = .15$) and the rudder task ($r = .19$). Finally, the correlation for flight hour code ($r = .23$) was significant and “Likely to be of Value” even as a stand-alone predictor.

Exploration of Alternatives. Unfortunately, multiple attempts to have other AFOQT composites (e.g., Verbal, Quantitative, Academic) and SynWin Multi-Tasking components enter the logistic regression models were not successful. In addition, forcing SynWin Memory, Math, Visual, and Listening scores to be retained in the “PCSM/Multi-Tasking Model” did not provide any incremental validity above and beyond the PCSM score alone. As will be discussed in the section on group differences, forcing SynWin Multi-Tasking components into the regression model did not reduce subgroup differences based on sex, race, or ethnicity.

An additional objective in this research endeavor was to evaluate the possibility of removing the TBAS components that require expensive and bulky joystick, rudders, and testing carrels. The resulting model included the AFOQT Pilot composite, the Unmanned Aerial Vehicle Composite of the TBAS, and the Flight Hour Code. While the resulting validity coefficient for the “No Joystick or Rudder Model” was comparable to the PCSM ($r = .38$, $p < .001$), it resulted in a slight increase in subgroup differences based on race, as will be discussed in more detail in the section on group differences. The AFOQT Pilot composite was clearly a strong “Very

Beneficial" predictor of SUPT completion (uncorrected $r = .37$, $p < .001$). However, as will be discussed in more detail in the section on subgroup differences, the AFOQT Pilot score alone results in rather substantial subgroup differences by race. Finally, as demonstrated in Table 4, the SynWin Multi-Tasking scores, with the exception of Trial 1 Performance and Memory, were not salient predictors of SUPT completion. The complete set of correlations is provided in Appendix A, Table A.1.

Table 4
Relationship between PCSM, TBAS, and SynWin Multi-Tasking Scores with Completion/Attrition through IFT and SUPT

	Complete through SUPT Primary *N* = 888
	Completion (1) Versus Attrition (0)
	R
Current PCSM Model	**.39*** (.42)**
AFOQT – Pilot	**.37*** (.40)**
No Joystick or Rudder Model	**.38*** **
PCSM/SynWin Multi-Tasking Model	**.39*** **
Flight Hour Code	**.23*** **
TBAS Components	
A (Airplane Redirects 1)	**.23*** **
AOnTarget (Distance from Target 1)	**.23*** **
AHA (Airplane Redirects 2)	**.21*** **
AHOnTarget (Distance from Target 2)	**.24*** **
AHROnTarget (Rudder Task)	**.19*** **
Unmanned Aerial Vehicle Composite	**.15*** **
Multi-Tasking Test Scores	
Overall Trial 1 Performance	**.08***
Overall Trial 2 Performance	.06
Overall Trial 3 Performance	.05
Overall Trial 4 Performance	.07
Multi-Tasking Total Score	.07
Memory Total	**.08***
Math Total	.02
Visual Total	.04
Listening Total	.05

Note. Completion $n = 738$; Attrition $n = 150$. Values in parentheses are corrected for Multivariate Range Restriction.

As demonstrated in Figure 1, the PCSM composite is highly effective in identifying candidates who are likely to attrit versus those who are likely to successfully complete IFT and SUPT Primary. By continuing the current standard of PCSM $\geq$ 10, the USAF is eliminating

candidates who are not likely to succeed (i.e., 63.2% attrit). In contrast, candidates who obtain a PCSM score of 90 or higher are very likely to succeed (i.e., 97.7% complete SUPT).

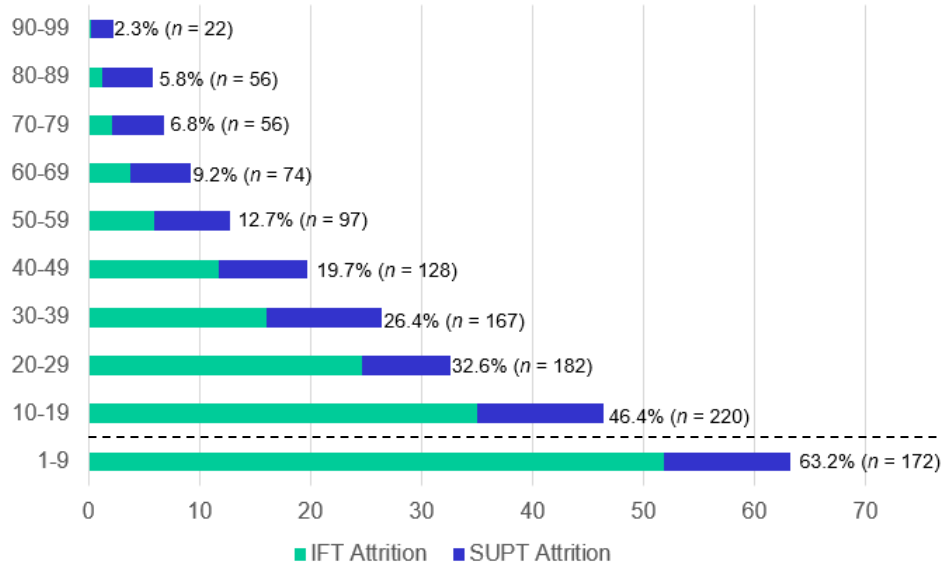

Figure 1. SUPT Attrition by PCSM Score Interval

Note. N = 6,911 (2007-2018); 1,174 IFS or SUPT Primary eliminees and 5,737 SUPT Primary graduates (16.99% attrition rate). All PCSM deciles were based on at least 270 cases. The current policy cut score for the PCSM is >=10. The demographics for those who did not meet the minimum cut score included 8.72% Black/African-American applicants, 19.19% Hispanic applicants, and 17.54% Female applicants. Of those, 4.12% were from USAFA, 8.24% were from OTS, and 85.88% were from ROTC.

As shown in Table 5, the PCSM score is predictive of additional SUPT Primary flight training outcomes such as Daily Flying Scores (uncorrected r = .33, p < .001), Academic Grades (uncorrected r = .24, p < .001), and MASS (uncorrected r = .25, p < .001). The regression-based models to remove the joystick/rudder components had comparable validity coefficients, as did the model which incorporated SynWin Multi-Tasking (see Table 5). The complete set of correlations is provided in Appendix A, Table A.2.

Table 5

Validity of PCSM, TBAS, and SynWin Multi-Tasking Scores as Predictors of Daily Flight Performance, Academic Performance, and Merit Assignment Selection System (MASS) Scores for Manned Aircraft Pilots

	SUPT Primary (T6A) *N* = 756			SUPT Advanced (T1A) *N* = 210			SUPT Advanced (T38C) *N* = 107		
	Daily Flying	**Academic**	**MASS**	**Daily Flying**	**Academic**	**MASS**	**Daily Flying**	**Academic**	**MASS**
Current PCSM Model	**.33*** (.37)**	**.24*** (.31)**	**.25*** (.30)**	**.33*** (.35)**	**.18* (.22)**	**.30*** (.32)**	**.25** (.28)**	.08 (.08)	**.22* (.24)**
AFOQT – Pilot	**.31*** (.47)**	**.31*** (.42)**	**.28*** (.43)**	**.27*** (.46)**	**.20** (.29)**	**.27*** (.44)**	.19 (.28)	.12 (.20)	**.20* (.30)**
No Joystick or Rudder Model	**.33***	**.27***	**.25***	**.33***	**.19**	**.32***	**.22***	.11	.19
PCSM/SynWin Multi-Tasking Model	**.33***	**.25***	**.26***	**.33***	**.18***	**.30***	**.26**	.08	**.22***
Flight Hour Code	**.20***	**.09***	**.11**	**.23***	.09	**.23***	.19	.06	.14
TBAS Components									
A (Airplane Redirects 1)	**.20***	**.14***	**.15***	.13	.08	.11	**.21***	.03	.19
AOnTarget (Distance from Target 1)	**.17***	**.14***	**.14***	.11	.05	.07	**.21***	.13	.17
AHA (Airplane Redirects 2)	**.24***	**.15***	**.20***	.18**	.10	**.17***	**.23***	.14	**.28***
AHOnTarget (Distance from Target 2)	**.24***	**.16***	**.19***	.15*	.08	**.14***	.18	**.19***	.19
AHROnTarget (Rudder Task)	**.19***	.05	**.16***	.19**	.03	.09	**.23***	-.08	.19
Unmanned Aerial Vehicle Composite	**.11***	**.18***	**.12**	.14*	.05	.11	.03	.07	-.03
Multitasking Test Scores									
Overall Trial 1 Performance	**.17***	**.12**	**.17***	.11	.01	.09	.16	.11	.09
Overall Trial 2 Performance	**.16***	**.14***	**.18***	.09	.05	**.14***	**.21***	.12	.15
Overall Trial 3 Performance	**.19***	**.15***	**.20***	.05	-.03	.02	**.20***	.09	.08
Overall Trial 4 Performance	**.17***	**.15***	**.19***	.02	.01	.07	.13	.12	.15
Multi-Tasking Total Score	**.19***	**.15***	**.20***	.07	.01	.09	.19*	.12	.13
Memory Total	**.21***	**.19***	**.22***	.05	.06	.10	**.20***	.02	**.20***
Math Total	.07	**.09***	**.11**	.07	.05	.09	.00	.02	-.06
Visual Total	.06	.03	.05	.09	.00	.03	.12	.15	.08
Listening Total	**.19***	**.13***	**.18***	-.03	-.10	.01	**.26**	.11	**.19***

Note. Values in parentheses are corrected for Multivariate Range Restriction. ** $p < .05$; ** $p < .01$; *** $p < .001$.

In addition to manned aircraft training (i.e., SUPT), the current PCSM is also predictive of remotely piloted aircraft (RPA) training outcomes. Specifically, as shown in Table 6, the PCSM composite was highly related to training outcomes during RPA Initial Flight Screening (RFS), including Daily Flying Grades (uncorrected $r = .31$, $p < .001$), Academic Grades (uncorrected $r = .21$, $p < .001$), and MASS (uncorrected $r = .35$, $p < .001$). The PCSM composite was also related to training outcomes during RPA Instrument Qualification (RIQ), including Daily Flying Grades (uncorrected $r = .22$, $p < .001$), Academic Grades (uncorrected $r = .23$, $p < .001$), and MASS (uncorrected $r = .27$, $p < .001$). Similar results were observed for the alternate regression models (see Table 6). The complete set of correlations is provided in Appendix A, Table A.3.

Table 6

Validity of PCSM, TBAS, and SynWin Multi-Tasking Scores as Predictors of Daily Flight Performance, Academic Performance, and Merit Assignment Selection System (MASS) Scores for Remotely Piloted Aircraft (RPA) Pilots

	RFS *N* = 449			RIQ *N* = 370		
	Daily Flying	**Academic**	**MASS**	**Daily Flying**	**Academic**	**MASS**
Current PCSM Model	**.31*** (.42)**	**.21*** (.28)**	**.35*** (.41)**	**.22*** (.26)**	**.23*** (.30)**	**.27*** (.32)**
AFOQT – Pilot	**.24*** (.36)**	**.24*** (.29)**	**.30*** (.41)**	**.23*** (.31)**	**.24*** (.35)**	**.27*** (.32)**
No Joystick or Rudder Model	**.29***	**.25***	**.35***	**.22***	**.26***	**.28***
PCSM/Multitasking	**.32***	**.21***	**.36***	**.23***	**.23***	**.27***
Flight Hour Code	**.18***	.08	**.19***	.04	.11	**.10***
TBAS Components						
A (Airplane Redirects 1)	**.16***	.00	**.17***	**.18***	.08	**.15****
AOnTarget (Distance from Target 1)	**.15****	-.04	**.12****	**.14****	.06	**.11***
AHA (Airplane Redirects 2)	**.21***	.05	**.18***	**.17****	.08	**.13***
AHOnTarget (Distance from Target 2)	**.21***	.00	**.17***	**.14****	.06	.10
AHROnTarget (Rudder Task)	**.16***	-.01	**.13****	**.15****	.01	.09
Unmanned Aerial Vehicle Composite	**.11***	.09	**.16***	**.15****	.17	**.17****
Multi-Tasking Test Scores						
Overall Trial 1 Performance	**.11***	.07	**.11***	**.18***	.04	**.15****
Overall Trial 2 Performance	**.13***	.06	**.11***	**.18***	.05	**.14****
Overall Trial 3 Performance	**.11***	.03	.12*	**.18***	.05	**.11***
Overall Trial 4 Performance	.07	.03	.05	**.13***	.07	.08
Multi-Tasking Total Score	.12	.05	.11*	**.19***	.06	**.13***
Memory Total	.11	.05	.09	**.17****	.10*	.11
Math Total	.01	.01	.04	**.18***	.04	.10
Visual Total	.05	.00	.02	.01	-.06	-.01
Listening Total	**.17***	.10*	**.17***	**.17****	.09	**.18***

Note. Values in parentheses are corrected for Multivariate Range Restriction. ** $p < .05$; ** $p < .01$; *** $p < .001$.

Group Differences

Male-Female mean score differences were evaluated for a sample of N = 14,214 examinees who completed the AFOQT, TBAS, and SynWin Multi-Tasking Test (see Table 7). Results suggest that males in the training applicant sample generally score higher on the PCSM composite with an observed moderate to large effect size (Cohen's d = .65). Moderate to large effect sizes favoring males were also observed for the TBAS components (Cohen's ds ranged from .32 for the Rudder Average Distance from Target task [AHROnTarget] to 1.52 for the Airplane on Target task [AOnTarget]). While small mean score differences were observed for the SynWin Multi-Tasking components, incorporating them into the PCSM algorithm did not

reduce the effect size (Cohen's *d* remained at .65). The complete set of group mean differences is provided in Appendix A, Table A.4.

Table 7
Male-Female Effect Size Differences across PCSM, TBAS, and Multi-Tasking Scores for Training Applicant Sample

	Male (*N* = 12,451)		Female (*N* = 1,763)		*d*
	Mean	SD	Mean	SD	
Existing PCSM Score	45.07	27.34	27.63	24.31	0.65
AFOQT Pilot	71.35	21.89	56.53	23.19	0.67
No Joystick or Rudder Model	1.39	1.30	0.61	1.31	0.60
PCSM/MTT Model	0.77	0.15	0.67	0.14	0.65
Flight Hour Code	2.12	2.91	1.65	2.54	0.16
TBAS Components					
A (Airplane Redirects 1)	10.17	4.43	4.34	3.28	1.35
AOnTarget (Distance from Target 1)	648.06	141.54	431.03	154.58	1.52
AHA (Airplane Redirects 2)	7.88	5.25	2.48	2.74	1.08
AHOnTarget (Distance from Target 2)	906.43	287.16	494.96	264.68	1.45
AHROnTarget (Rudder Task)	1087.17	219.29	1016.36	242.99	0.32
Unmanned Aerial Vehicle Composite	0.28	1.01	-0.17	1.13	0.43
Multi-Tasking Test Scores					
Overall Trial 1 Performance	0.02	0.64	-0.11	0.66	0.19
Overall Trial 2 Performance	0.01	0.66	-0.10	0.67	0.17
Overall Trial 3 Performance	0.01	0.66	-0.10	0.67	0.17
Overall Trial 4 Performance	0.01	0.66	-0.10	0.68	0.17
Multi-Tasking Total Score	0.01	0.60	-0.10	0.61	0.19
Memory Total	0.01	0.77	-0.06	0.84	0.09
Math Total	0.01	0.88	-0.06	0.92	0.08
Visual Total	0.01	0.87	-0.05	0.87	0.07
Listening Total	0.03	0.94	-0.22	1.00	0.27

Cohen's (1988) *d* Interpretation: .2 = Small; .5 = Medium; .8 = Large.

Group differences by socioeconomic status (SES) were evaluated in a sample of *N* = 9,382 and were less remarkable than the differences by sex (see Table 8). Specifically, the overall PCSM score resulted in a small effect size (Cohen's *d* = .28). The Flight Hour Code and the AFOQT Pilot composite also resulted in small effect sizes based on mean score differences (Cohen's *d* = .21 and .28, respectively). The TBAS components resulted in almost no practical mean score differences (Cohen's *d*s = .02 to .11). Finally, the SynWin Multi-Tasking components resulted in near-zero mean-score differences by SES. The complete set of group mean differences is provided in Appendix A, Table A.5.

Table 8

Socioeconomic Status (Average or Higher versus Low) Effect Size Differences across PCSM, TBAS, and Multi-Tasking Scores for Training Applicant Sample

	Average or Higher (*N* = 7,641)		Low (*N* = 1,741)		*d*
	Mean	SD	Mean	SD	
Existing PCSM Score	45.92	27.45	38.17	26.51	0.28
AFOQT Pilot	72.15	21.82	66.06	22.92	0.28
No Rudder or Joystick Model	1.45	1.30	1.07	1.30	0.30
PCSM/MTT Model	0.77	0.15	0.73	0.15	0.28
Flight Hour Code	2.31	2.94	1.70	2.65	0.21
TBAS Components					
A (Airplane Redirects 1)	9.44	4.67	9.32	4.74	0.02
AOnTarget (Distance from Target 1)	623.08	158.83	614.97	161.88	0.05
AHA (Airplane Redirects 2)	7.32	5.29	6.76	5.12	0.11
AHOnTarget (Distance from Target 2)	864.84	312.04	831.13	313.49	0.11
AHROnTarget (Rudder Task)	1084.07	218.06	1059.21	219.46	0.11
UAVComposite	0.25	1.01	0.15	1.044	0.09
Multi-Tasking Test Scores					
Overall Trial 1 Performance	0.01	0.64	-0.02	0.64	0.05
Overall Trial 2 Performance	0.02	0.65	-0.02	0.65	0.06
Overall Trial 3 Performance	0.02	0.65	-0.02	0.66	0.06
Overall Trial 4 Performance	0.01	0.66	-0.02	0.69	0.05
Multi-Tasking Total Score	0.02	0.59	-0.02	0.60	0.06
Memory Total	0.01	0.86	-0.02	0.88	0.04
Math Total	0.03	0.87	-0.05	0.88	0.09
Visual Total	0.03	0.93	0.00	0.97	0.03
Listening Total	0.00	0.77	0.01	0.79	0.00

Cohen's (1988) *d* Interpretation: .2 = Small; .5 = Medium; .8 = Large.

Mean score differences by race (White versus Black/African-American) were evaluated for a sample of *N* = 12,430 examinees who completed the AFOQT, TBAS, and SynWin Multi-Tasking Test (see Table 9). Results suggest that White applicants generally score higher on the PCSM composite with an observed moderate effect size (Cohen's *d* = .75). The AFOQT Pilot composite resulted in a mean score difference of one standard deviation favoring White applicants (Cohen's *d* = 1.01). Flight hour code had a small effect size based on the mean score difference favoring White applicants (Cohen's *d* = .21).

Moderate effect sizes favoring White applicants were observed for the TBAS components (Cohen's *d*s ranged from .32 for the Rudder Average Distance from Target task [AHROnTarget] to 1.52 for the Airplane on Target task [AOnTarget]). As with other group differences, small mean score differences were observed for the SynWin Multi-Tasking components. However, incorporating them into the PCSM algorithm did not reduce the effect

size (Cohen's *d* remained at .75). The complete set of group mean differences is provided in Appendix A, Table A.5.

Table 9
Race Effect Size Differences across PCSM, TBAS, and Multi-Tasking Scores for Training Applicant Sample: White versus Black/African-American

	White (*N* = 11,803)		Black (*N* = 627)		*d*
	Mean	SD	Mean	SD	
Existing PCSM Score	44.93	27.32	24.51	25.42	0.75
AFOQT Pilot	71.34	21.67	49.24	24.27	1.01
No Joystick or Rudder Model	1.40	1.29	0.25	1.40	0.89
PCSM/MTT Model	0.76	0.15	0.65	0.14	0.75
Flight Hour Code	2.17	2.92	1.57	2.70	0.21
TBAS Components					
A (Airplane Redirects 1)	9.67	4.67	7.86	4.60	0.39
AOnTarget (Distance from Target 1)	628.84	156.93	564.95	170.70	0.41
AHA (Airplane Redirects 2)	7.46	5.35	5.33	4.53	0.40
AHOnTarget (Distance from Target 2)	869.98	312.15	729.75	310.41	0.45
AHROnTarget (Rudder Task)	1087.19	221.40	1007.92	220.84	0.36
Unmanned Aerial Vehicle Composite	0.26	1.01	-0.41	1.21	0.65
Multi-Tasking Test Scores					
Overall Trial 1 Performance	0.00	0.64	-0.16	0.61	0.27
Overall Trial 2 Performance	0.00	0.66	-0.19	0.63	0.29
Overall Trial 3 Performance	0.01	0.66	-0.19	0.64	0.30
Overall Trial 4 Performance	0.01	0.66	-0.21	0.66	0.34
Multi-Tasking Total Score	0.01	0.59	-0.19	0.58	0.33
Memory Total	0.01	0.86	-0.34	0.87	0.41
Math Total	0.00	0.88	-0.26	0.91	0.30
Visual Total	0.01	0.95	-0.14	0.92	0.15
Listening Total	0.00	0.77	-0.02	0.80	0.03

Cohen's (1988) *d* Interpretation: .2 = Small; .5 = Medium; .8 = Large.

Mean score differences by ethnicity (Hispanic versus Non-Hispanic) were evaluated for a sample of *N* = 14,124 examinees who completed the AFOQT, TBAS, and SynWin Multi-Tasking Test (see Table 10). Results suggest that Non-Hispanic applicants generally score slightly higher on the PCSM with a relatively small effect size (Cohen's *d* = .33). The AFOQT Pilot composite resulted in a slightly higher mean score difference (Cohen's *d* = .38). Flight hour code had a very small effect size based on the mean score difference favoring Non-Hispanic applicants (Cohen's *d* = .16).

Virtually no score differences based on ethnicity were observed for the TBAS components (Cohen's *d*s ranged from .07 for Airplane on Target tasks [AOnTarget and AHOnTarget] to .15 for the Unmanned Aerial Vehicle task (UAV Composite]). As with other

group differences, small mean score differences were observed for the SynWin Multi-Tasking components. However, incorporating them into the PCSM algorithm did not reduce the effect size (Cohen's *d* remained at .33). The complete set of group mean differences is provided in Appendix A, Table A.6.

Table 10

Ethnicity Effect Size Differences across PCSM, TBAS, and Multi-Tasking Scores for Training Applicant Sample: Non-Hispanic versus Hispanic

	Non-Hispanic (*N* = 12,529)		Hispanic (*N* = 1,595)		
	Mean	SD	Mean	SD	*d*
Existing PCSM Score	43.98	27.56	34.99	26.38	0.33
AFOQT Pilot	70.54	22.25	61.93	23.52	0.38
No Joystick or Rudder Model	1.35	1.31	0.88	1.33	0.36
PCSM/MTT Model	0.76	0.16	0.71	0.15	0.33
Flight Hour Code	2.12	2.90	1.67	2.62	0.16
TBAS Components					
A (Airplane Redirects 1)	9.50	4.71	9.05	4.70	0.09
AOnTarget (Distance from Target 1)	622.38	159.29	611.28	164.84	0.07
AHA (Airplane Redirects 2)	7.28	5.35	6.70	5.10	0.11
AHOnTarget (Distance from Target 2)	858.15	314.42	835.69	321.90	0.07
AHROnTarget (Rudder Task)	1081.69	223.46	1054.04	223.93	0.12
Unmanned Aerial Vehicle Composite	0.24	1.03	0.08	1.06	0.15
Multi-Tasking Test Scores					
Overall Trial 1 Performance	0.01	0.64	-0.07	0.64	0.13
Overall Trial 2 Performance	0.01	0.65	-0.08	0.68	0.15
Overall Trial 3 Performance	0.01	0.66	-0.09	0.68	0.16
Overall Trial 4 Performance	0.02	0.67	-0.11	0.68	0.18
Multi-Tasking Total Score	0.01	0.60	-0.09	0.61	0.17
Memory Total	0.02	0.87	-0.10	0.87	0.14
Math Total	0.02	0.88	-0.13	0.90	0.17
Visual Total	0.01	0.95	-0.09	0.99	0.10
Listening Total	0.01	0.77	-0.04	0.84	0.06

Cohen's (1988) *d* Interpretation: .2 = Small; .5 = Medium; .8 = Large.

Evaluation of Bias

"A lower mean test score in one group compared to another is not by itself evidence of bias" (Guion, 1998, p. 436). Group differences in mean test scores can represent actual differences in the abilities being measured. A more appropriate indicator of whether a test is biased is to evaluate whether the test is unequally predictive of job performance based on group membership (i.e., differential prediction).

Results of this study indicated that the criterion-related validity coefficients for the female sample were comparable to those obtained for the male sample. As shown in Table 11, a

similarly strong relationship between PCSM scores and pilot training completion was observed in both the sample of females (uncorrected $r = .383$, $p < .001$) and the sample of males (uncorrected $r = .382$, $p < .001$). In addition, a similarly strong relationship between AFOQT Pilot Scores (i.e., a primary component in the PCSM algorithm) and pilot training completion was observed in both the sample of females (uncorrected $r = .354$, $p < .001$) and the sample of males (uncorrected $r = .351$, $p < .001$).

Table 11
Criterion-Related Validity of PCSM Based on Sex

	***r* with SUPT Primary Completion**	**Fisher *Z***	***Significance of Difference***
PCSM Score			
Female	.383***	.404	*Ns*
Male	.382***	.402	
AFOQT Pilot Score			
Female	.354***	.370	*Ns*
Male	.351***	.367	

Note. Female $n = 540$; Male $n = 6{,}304$; All correlations are uncorrected. *** $p < .001$.

Similar results were observed based on race. The criterion-related validity coefficients for the Black/African-American sample were comparable to, or even slightly higher than those obtained for the White sample. As shown in Table 12, a similarly strong relationship between PCSM scores and pilot training completion was observed in both the sample of Black/African-American examinees (uncorrected $r = .383$, $p < .001$) and the sample of White examinees (uncorrected $r = .380$, $p < .001$). In addition, a similarly strong relationship between AFOQT Pilot composite scores and pilot training completion was observed in both the sample of African-American examinees (uncorrected $r = .391$, $p < .001$) and the sample of White examinees (uncorrected $r = .348$, $p < .001$).

Table 12
Criterion-Related Validity of PCSM Based on Race

	***r* with SUPT Primary Completion**	**Fisher *Z***	***Significance of Difference***
PCSM Score			
Black/African-American	.383***	.404	*Ns*
White	.380***	.400	
AFOQT Pilot Score			
Black/African-American	.391***	.413	*Ns*
White	.348***	.363	

Note. Black/African-American n = 164; White n = 6,244; All correlations are uncorrected. *** p <.001.

Similar results were observed based on ethnicity. The criterion-related validity coefficients for Hispanic applicants were higher than those obtained for Non-Hispanics. As shown in Table 13, a similarly strong relationship between PCSM scores and pilot training completion was observed in both the sample of Hispanic applicants (uncorrected r = .445, p < .001) and the sample of Non-Hispanic applicants (uncorrected r = .378, p < .001). In addition, a similarly strong relationship between AFOQT Pilot composite scores and pilot training completion was observed for both Hispanic applicants (uncorrected r = .389, p < .001) and Non-Hispanic applicants (uncorrected r = .347, *ns*).

Table 13
Criterion-Related Validity of PCSM Based on Ethnicity

	***r* with SUPT Primary Completion**	**Fisher *Z***	***Significance of Difference***
PCSM Score			
Hispanic	.445***	.478	p < .05
Non-Hispanic	.378***	.398	
AFOQT Pilot Score			
Hispanic	.389***	.411	*ns*
Non-Hispanic	.347***	.362	

Note. Hispanic n = 548; White n = 6,359; All correlations are uncorrected. *** p <.001.

Adverse Impact

As previously mentioned, differences between groups on a predictor or composite of predictors become significant when they influence selection outcomes (Sackett & Elligson,

1997). Adverse impact is demonstrated by a statistical disparity between the selection rates of majority and minority groups. The Uniform Guidelines (1978) recommend using an 80% (4/5ths) rule where if the minority group is selected at a rate less than 80% of the majority group, there is evidence of potential adverse impact. While measures of cognitive ability normally have been found to have the highest predictive validity for training and job performance when compared to other personnel selection methods commonly used (Jensen, 1980, 1998; Ree & Carretta, 2002; Carretta, 2006), they also result in group differences in test performance where Whites are more likely to score higher than African Americans and Hispanics (Carretta, 2006). As with any aptitude test, the PCSM composite has the potential to result in adverse impact. While adverse impact does not violate the law as long as the test is valid, job-relevant, and other alternatives have been explored (Cascio & Aguinis, 2005), the PCSM program strives to avoid adverse impact to the extent possible. In a study conducted by Carretta (2006), the PCSM composite had less adverse impact for pilot candidate selection for females and racial/ethnic minorities when compared to the AFOQT alone. These results were based on evaluating varying minimum PCSM qualifying scores on training qualifications (e.g., PCSM $\geq$ 25; PCSM $\geq$ 50).

For this study, in terms of potential adverse impact for the PCSM composite, the selection ratio, if the PCSM were the only hurdle for selection, based on gender would be acceptable based on the 10^{th} percentile cut score (Adverse Impact [AI] Ratio = .80). The adverse impact ratio based on race (White versus Black/African-American) would be somewhat lower (AI = .72). There was no evidence that there would be adverse impact based on ethnicity (AI = .91). It is important to note that these AI values may be inflated because multiple hurdles have already been passed prior to the PCSM being used as a selection tool (e.g., AFOQT Verbal >=15; AFOQT Quantitative >=10; AFQT Pilot >= 25). It also does not account for the numerous

other factors (e.g., weighting of the PCSM in the board process, other less-objective sources of influence such as GPA) in the pilot-selection process that play a role in the selection ratios of the entire process.

Grade Point Average (GPA)

Although grade point average (GPA) is collected via self-report during the TBAS test-administration process, it is not used in the PCSM algorithm. One contributing factor in the rationale to disregard GPA is that it does not necessarily have the same meaning across schools with differing performance/grading standards. Relative to the AFOQT Pilot composite score and the PCSM score, GPA is relatively ineffective in predicting SUPT completion (see Table 14). Specifically, using the self-reported GPA values, a weak relationship was observed between GPA and pilot training completion (uncorrected $r = .06$, $p < .001$). In contrast to the PCSM score (uncorrected $r = .39$, $p < .001$) and AFOQT Pilot score (uncorrected $r = .36$, $p < .001$), GPA is of little value in the pilot candidate selection process.

Table 14
Criterion-Related Validity of GPA

	***r* with SUPT Primary Completion**
Grade Point Average	.06***
PCSM Score	.39***
AFOQT Pilot Score	.36***

Note. N = 6,907. All correlations are uncorrected. *** p <.001.

Contribution and Implications of Including Flight Hour Code

Previous flying experience, both actual and simulated, has been well-documented as a strong predictor of pilot training performance (Carretta & Ree, 1994; Darr, 2009; Deitcher & Johnston, 2004; Johnston & Catano, 2013; Woycheshin, 2001). Caretta and Ree (1994) found that previous flying experience was almost as predictive of pilot training success as cognitive

ability, with the combination of the two being the best overall predictor. Additional studies have demonstrated that previous flying experience is associated with fewer safety incidents (Ison, 2015).

Results of this study were consistent with previous research in that the flight hour code included in the PCSM algorithm is a modest but significant predictor (i.e., Department of Labor Interpretation = "Likely to be Useful") of completion through IFT and SUPT Primary (uncorrected $r = .23$, $p < .001$). It also provides incremental validity to the PCSM score as evidenced by a reduction in the existing PCSM validity coefficient (uncorrected $r = .39$, $p < .001$) when removed from the algorithm (uncorrected $r = .37$, $p < .001$).

While the predictive validity of previous flying experience is unquestionable (see validity section), an evaluation of the impact on diversity of using flight hours for pilot training selection is warranted. As demonstrated previously, the practical differences were small based on mean flight hour codes across groups. Specifically, small practical differences were observed based on sex (Cohen's $d = .16$), socioeconomic status ($d = .21$), race ($d = .21$), and ethnicity ($d = .16$). A more stringent evaluation would be to evaluate the number of applicants who have previous flight experience based on group membership. Using crosstabs with expected versus observed frequencies, a series of analyses were conducted to further elucidate the impact of flight hours on diverse groups. Table 15 shows the observed frequency (the number of people with flying hours within a gender category) versus expected frequencies (the number of people within a gender category we would expect based on their ratio in the larger applicant population) of applicants with previous flying experience based on sex. Based on a chi square test, there was a significantly higher proportion of male applicants with previous flight hours than female applicants (48.2% versus 39.5%, respectively).

Table 15
Previous Flight Experience Based on Sex

		# of Applicants with Previous Flying Experience	Percent of Applicants (Within Group) with Previous Flying Experience
Female	Observed	1,726	39.5% (*N* = 4,371)
	Expected	2,062	
Male	Observed	15,795	48.2% (*N* = 32,763)
	Expected	15,495	

$\chi^2(1) = 117.73$, $p < .0001$

Table 16 shows the observed versus expected frequencies of applicants with and without previous flying experience based on race. Based on a chi square test, there was a significantly higher proportion of White applicants with previous flight hours than Black/African-American applicants (48.9% versus 36.4%, respectively).

Table 16
Previous Flight Experience Based on Race

		# of Applicants with Previous Flying Experience	Percent of Applicants (Within Group) with Previous Flying Experience
Black African – American	Observed	536	36.4% (*N* = 1,474)
	Expected	712	
White	Observed	15,446	48.9% (*N* = 31,601)
	Expected	15,270	

$\chi^2(1) = 88.34$, $p < .0001$

Finally, Table 17 shows the observed versus expected frequencies of applicants with and without previous flying experience based on ethnicity. While the difference was less remarkable than for sex or race, a chi square test still revealed that there was a significantly higher proportion of Non-Hispanic applicants with previous flight hours than Hispanic applicants (47.8% versus 41.4%, respectively).

Table 17
Previous Flight Experience Based on Ethnicity

		# of Applicants with Previous Flying Experience	Percent of Applicants (Within Group) with Previous Flying Experience
Hispanic	Observed	1,553	41.4% (*N* = 3,753)
	Expected	1,770	
Non-Hispanic	Observed	15,995	47.8% (*N* = 33,453)
	Expected	15,778	

$\chi^2(1) = 56.04$, $p < .0001$

Differential Analysis Based on Flight Hour Code

Additional results of this study indicated that the criterion-related validity coefficients based on flight hour code for the female sample were comparable to those obtained for the male sample. As shown in Table 18, a similar relationship between flight hour code and pilot training completion was observed in both the f female (uncorrected $r = .283$, $p < .001$) and male (uncorrected $r = .265$, $p < .001$) samples. Similar results were observed based on race. The criterion-related validity coefficients for a Black/African-American applicants were comparable to, or even slightly higher than those obtained for White applicants. As shown in Table 18, a similar relationship between flight hour code and pilot training completion was observed for both Black/African-American applicants (uncorrected $r = .273$, $p < .001$) and White applicants (uncorrected $r = .262$, $p < .001$). Similar results were observed based on ethnicity. The criterion-related validity coefficient for the Hispanic applicants was significantly higher than those obtained for the Non-Hispanic applicants. As shown in Table 18, a similar relationship between Flight hour code and pilot training completion was observed for both Hispanic applicants (uncorrected $r = .389$, $p < .001$) and Non-Hispanic applicants (uncorrected $r = .260$, $p < .001$).

Table 18

Criterion-Related Validity of Flight Hour Code Based on Sex, Race, and Ethnicity.

	***r* with SUPT Primary Completion**	**Fisher *Z***	***Significance of Difference***
Female (*n* = 540)	.283	.291	*ns*
Male (*n* = 6,304)	.265	.271	
Black/African-American (*n* = 164)	.273	.280	*ns*
White (*n* = 6,244)	.262	.268	
Hispanic (*n* = 548)	.389	.342	$p < .05$
Non-Hispanic (*n* = 6,359)	.260	.266	

Note. All correlations are uncorrected and significant at $p < .001$.

Adverse Impact Based on Flight Hour Code

As previously addressed, adverse impact is demonstrated by a disparity between the selection rates of majority and minority groups. Analyses were run to determine the potential adverse impact for the PCSM with and without flight hour code, if the PCSM were the only selection hurdle. As shown in Table 19, the selection ratio based on gender would be acceptable based on the PCSM 10th percentile cut score with flight hour code (Adverse Impact [AI] Ratio = .80) and slightly less optimal without flight hour code (AI = .78). Removing flight hour code would have resulted in 29 fewer female selections over a 12-year period. The adverse impact ratio based on race (Black/African-American versus White) would be somewhat better with flight hour code (AI = .72) than without flight hour code (AI = .71). Removing flight hour code would not have resulted in any fewer Black/African-American selectees over the 12-year period. There was no evidence that adverse impact would exist based on ethnicity with (AI = .91) or without (AI = .92) flight hour code. The removal of flight hour code from the PCSM algorithm would have increased the number of Hispanic trainees by 58 over the 12-year period. As stated previously, it is important to note that these AI values may be inflated because multiple hurdles have already been passed prior to the PCSM being used as a selection tool (e.g., AFOQT Verbal >=15; AFOQT Quantitative >=10; AFQT Pilot >= 25). It also does not account for the numerous other factors (e.g., weighting of the PCSM in the board process, other less-objective sources of influence such as GPA) in the pilot-selection process that play a role in the selection ratios of the entire process.

Table 19

Adverse Impact of PCSM with and without Flight Hour Code

	Current PCSM with Flight Hour Code	PCSM without Flight Hour Code	Difference in # of URG Selectees (2007-2019)
Female/Male	0.80 (*N* = 3,181)	0.78 (*N* =3,152)	29 fewer
Black African-American/White	0.72 (*N* = 963)	0.71 (*N* = 963)	0
Hispanic/Non-Hispanic	0.91 (*N* = 3,074)	0.92 (*N* =3,132)	58 more

Flight-Hour Code Alternatives

In an effort to quantify the contribution of the 10-category flight hour code, an evaluation of the predicted probability of success through SUPT by flight hour code was conducted. As shown in Figure 2, the predicted probability of success increases substantially from 0 hours to 41-60 hours, then the probability of success demonstrates a relative plateau. It should also be noted that 41-60 hours of FAA-approved flight hours is the range typically required for a private pilot's license.

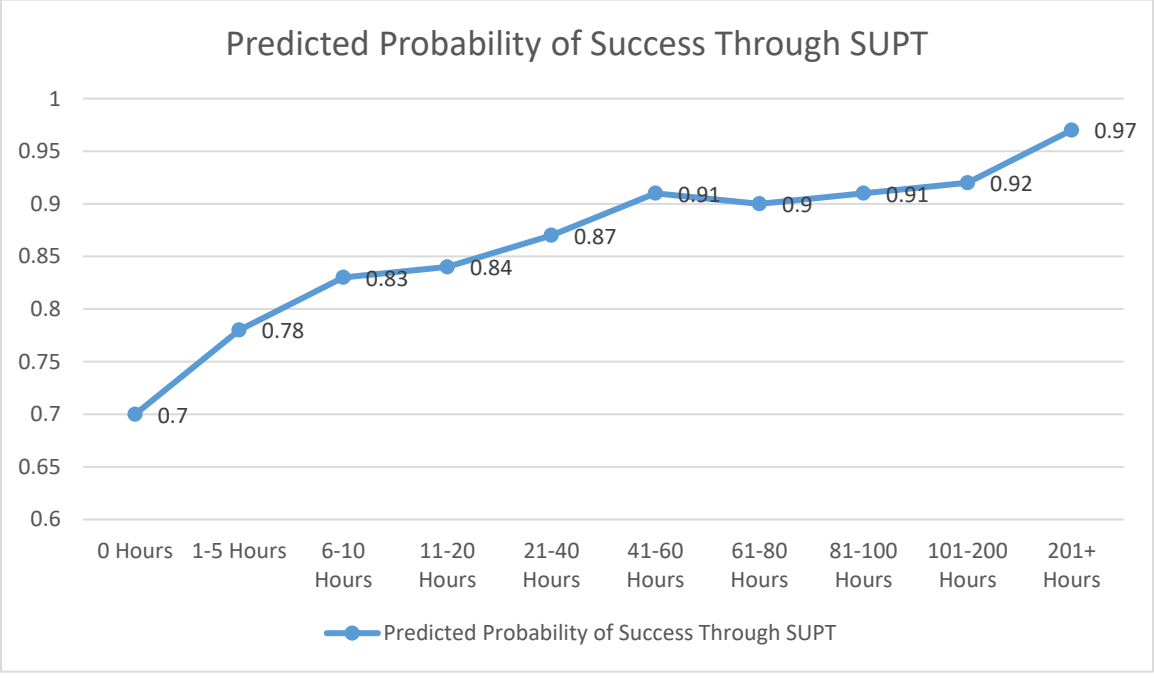

Figure 2. Predicted Probably of Success through SUPT by Flight Hour Code

Based on the evidence presented in Figure 2, the impact of reducing the number of flight hour codes was investigated further. Such a reduction would dis-incentivize paying for additional flight hours to improve the PCSM score if there is little empirical evidence that a high number (>60) of flight hours are of benefit to the Air Force. Thus, to determine if reducing flight hour code to 6 categories instead of 10 would have an impact on validity, a correlational analysis was conducted to determine if the PCSM would continue to predict the most important training outcomes from SUPT (e.g., Completion, Daily Ratings). As shown in Table 20, reducing the flight hour code to 6 categories had no impact on the criterion-related validity of PCSM.

Table 20
Criterion-Related Validity of PCSM with 6-Category Flight Hour Code Based on Sex, Race, and Ethnicity.

	***r* with SUPT Primary Completion**	***r* with SUPT Daily Ratings**
Current PCSM with 10-Category Flight Hour Code	.39	.34
PCSM with 6-Category Flight Hour Code	.40	.34

Note. All correlations are uncorrected and significant at $p < .001$.

As shown in Table 21, the selection ratio using the PCSM (assuming that it was the only selection hurdle) with the 6-category Flight hour code based on gender would still be acceptable based on the 10th percentile cut score with flight hour code (Adverse Impact [AI] Ratio = .80). Using the 6-category flight hour code would have resulted in 47 more female selections over a 12-year period. The adverse impact ratio based on race (Black/African-American versus White) would be slightly better with the 6-category Flight hour code (AI = .73) than the 10-category code (AI = .71), and would have resulted in 26 more Black/African-American selections. As with the 10-category code, there was no evidence that adverse impact would exist based on ethnicity using the 6-category code (AI = .92). The 6-hour code would have increased the number of Hispanic trainees by 69 over the 12-year period. It should be noted that the additional

selectees over the 12-year period would have been selected at the low end of ability as measured by the PCSM (~10th percentile), thus adding risk to the Air Force of selecting candidates who may not be as likely to succeed in pilot training as their higher-ability counterparts.

Table 21
Adverse Impact of Current PCSM and PCSM with 6-Category Flight Hour Code

	Current PCSM with Flight Hour Code	**PCSM with 6-Category Flight Hour Code**	**Difference in # of URG Selectees (2007-2019)**
Female/Male	0.80 (*N* = 3,181)	0.80 (*N* =3,152)	47 more
Black African-American/White	0.72 (*N* = 963)	0.73 (*N* = 989)	26 more
Hispanic/Non-Hispanic	0.91 (*N* = 3,074)	0.92 (*N* =3,143)	69 more

Discussion

The PCSM composite continues to demonstrate value in predicting multiple manned and unmanned pilot training outcomes. Most importantly, as it was designed to do, the existing PCSM composite is the most valid of all the predictors evaluated in this study as related to trainee completion through SUPT Primary. Alternatives that do not require the use of the testing carrel, the joystick, and the rudders are available and demonstrate similar validity and reasonably comparable group differences. Financial and logistical considerations can drive determinations regarding modifications to the components and scoring algorithm, but from a scientific/research perspective, the current PCSM is working as designed and is slightly better than other alternatives in terms of maximizing validity and minimizing group differences to the extent possible.

Conclusion/Recommendations

There is nothing noted in the data or the results of this study to suggest that immediate changes to the PCSM are critically needed, with the possible exception of reducing the number of flight hour categories. This validation effort also did not reveal changes in PCSM components or the scoring algorithm that would vastly improve validity and/or dramatically reduce group differences. That said, the comprehensive evaluation of flight hour codes demonstrated smaller group differences than what popular sentiment suggests and smaller group differences than some other components of the PCSM (e.g., TBAS, AFOQT-Pilot). Flight hour code demonstrated validity coefficients that are considered “Likely to be Useful” by U.S. Department of Labor (1999) standards within each group, and did not show differences in validity across groups. In addition, there was some evidence that the complete removal of flight hour code from the PCSM

algorithm would result in a reduction in the accuracy of predicting attrition through SUPT. Given that fewer female applicants and racial/ethnic minority applicants have previous flight experience upon application as compared to their male and white counterparts, reducing the flight hour code to 6 categories instead of 10 could result in slight improvements in adverse impact ratios and a slight improvement in the selection of members of diverse groups with no negative impact on the criterion-related validity of PCSM. Thus, a transition to the 6-category code is worth consideration.

In addition, the components/graphics of the TBAS appear somewhat dated and the rudders are not always well-received by examinees or subject matter experts (SMEs). Thus, modifying or updating the test to appear more modern and face-valid may be warranted.

Finally, all analyses and recommendations provided in this study were based on pilot training outcomes, particularly attrition/completion given the associated training costs. However, actual post-training job performance data may provide additional insight into how the PCSM predicts long-term performance.

Additional Recommendations for Future Research

Numerous additional tests are currently being evaluated by Georgia Tech Research Institute for the potential of predicting performance with RPA pilots (Ackermann, 2018). If these tests demonstrate solid psychometric characteristics and evidence of validity in predicting remotely piloted aircraft training outcomes, they may also be evaluated for predicting performance in manned aircraft pilot training. In addition, tests such as the Self-Description Inventory (SDI), which is a personality test administered as part of the AFOQT, and the AFOQT Situational Judgment Test (SJT), which measures leadership and judgment, have now been administered long enough to provide sufficient samples to evaluate how they may contribute to

the predictive power of PCSM while reducing adverse impact against minority groups. Some of the components of PSCM that demonstrate the most adverse impact (e.g., the AFOQT Pilot Composite) should also be evaluated for revision.

Finally, while the SynWin Multi-Tasking Test did not show much promise in predicting pilot training outcomes, preliminary results with Combat Systems Operators (CSO) training outcomes suggests that it might be worthy of additional research. For example, the Multi-Tasking Total Score was related to CSO Primary MASS scores ($r = .19$, $p < .001$; $N = 356$) and CSO Advanced MASS scores ($r = .33$, $p < .001$; $N = 234$). Thus, further exploration with CSO training outcomes is warranted.

References

Ackerman, P. L. (2018). *Selection and classification for UAS personnel.* Unpublished manuscript. Arlington, VA: Office of Naval Research.

Arth, T. O. (1986). *Validation of the AFOQT for non-rated officers.* File 11-2-1. Brooks, AFB, TX: Manpower and Personnel Division.

Aguilar, I. D. (2017). *Comparison of the Air Force Officer Qualifying Test Form T and Form S: Initial Item- and Subtest-Level Analyses.* (Tech. Rep. No. AFCAPS-TR-2017-0002). Randolph AFB, TX: Air Force Personnel Center.

Barron, L. G. (2015). *Preliminary evaluation of the USAF Multi-Tasking Test: Original SynWin and modified Test of Basic Aviation Skills (TBAS) Versions*. Unpublished Manuscript. Randolph AFB, TX: Air Force Personnel Center.

Barron, L, G., & Rose, M. R. (2017). Multitasking as a predictor of pilot performance: Validity beyond serial single-task assessments, *Military Psychology*, 29:4, 316-326.

Carretta, T. R. (1992a). Recent developments in U. S. Air Force pilot candidate selection and classification. *Aviation, Space, and Environmental Medicine*, 63, 1112-1114.

Carretta, T. R. (1992b). Understanding the relations between selection factors and pilot training performance: Does the criterion make a difference? *The International Journal of Aviation Psychology*, 2, 2, 95-105.

Carretta, T. R. (2005). *Development and validation of the Test of Basic Aviation Skills (TBAS)* (Tech. Rep. No. AFRL-HEWPTR-2005-0172). Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate.

Carretta, T. R. (2006). *Evaluation of Adverse Impact for US Air Force Officer and Aircrew Selection Tests*. (Tech. Rep. No. AFRL-HE-WP-TR-2006-0078). Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate.

Carretta. T. R. (2009). *Predictive validity of the Air Force Officer Qualifying Test (AFOQT) for non-rated officer specialties,* AFRL-RH-WP-TR-2010-0065. Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate, Crew Systems Interface Division, Supervisory Control Interfaces Branch.

Carretta, T. R. (2011). Pilot Candidate Selection Method: Still an effective predictor of US Air Force pilot training performance. *Aviation Psychology and Applied Human Factors, 1,* 3–8.

Carretta, T. R. (2013). Predictive validity of pilot selection instruments for remotely piloted aircraft training outcome. *Aviation, Space, and Environmental Medicine, 84*(1), 47-53. https://doi.org/10.3357/asem.3441.2013

Carretta, T. R., & Ree, M. J. (1994). Pilot candidate selection method (PCSM): Sources of validity. *International Journal of Aviation Psychology, 4,* 103-117.

Carretta, T. O., Rose, M. R., & Trent, J. D. (2016). *Air Force Officer Qualifying Test Form T: Initial item-, test-, factor-, and composite-level analyses.* AFRL-RH-WP-TR-2016-0093. Wright-Patterson AFB, OH: Air Force Research Laboratory, 711 Human Performance Wing, Airman Systems Directorate.

Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management* (6th ed.). Upper Saddle River, NJ: Pearson.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Darr, W. A. (2009). *A psychometric examination of the Canadian Automated Pilot Selection System* (Tech. Memo. No. 2009-024). Ottawa, ON, Canada: Director General Military Personnel Research and Analysis, National Defence Headquarters.

Deitcher, J., & Johnston, P. J. (2004). *Validation of the Canadian Forces Pilot Selection System.* Ottowa, ON, Canada: Director of Human Resources Research and Evaluation, National Defence Headquarters.

Duke, A. P., & Ree, M. J. (1996). Better candidates fly fewer training hours: Another time testing pays off. *International Journal of Selection and Assessment, 4,* 115-121.

EASI Consult, Schwartz, K. L., and Weissmuller, J. J. (2008). *Air Force Officer Qualifying Test (AFOQT) composite structure validation: Subgroup qualification rates,* Deliverable #2, FA3089-07-F-0483 (FMLO-FR-2009-0001), Randolph AFB, Tex.: Force Management Liaison Office, HQ Air Force Personnel Center, 2008.

Elsmore, T. F. (1994). SYNWORK: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instrumentation, and Computers, 26,* 421-426.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions.* Mahwah, NJ: Erlbaum.

Hardison, C. M., Sims, C. S., & Wong, E. C. (2010*). The Air Force Officer Qualifying Test: Validity, Fairness, and Bias*. Santa Monica, Ca: Rand Corporation, Project Air Force

Ison, D. C. (2015). Comparative analysis of accident and non-accident pilots. *Journal of Aviation Technology and Engineering, 4*(2). 20-31.

Jensen, A. A. (1980). *Bias in mental testing.* New York City, NY: The Free Press.

Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.

Johnston, P. J., & Catano, V. M. (2013). Investigating the validity of previous flying experience, both actual and simulated, in predicting initial and advanced military pilot training performance. *The International Journal of Aviation Psychology*, *23*(3), 227-244.

Oswald, F. L., Hambrick, D. Z., & Jones, L. A. (2007). Keeping all the plates spinning: Understanding and predicting multitasking performance. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems*, pp. 77-96. New York: Lawrence Erlbaum.

Paullin, C., Ingerick, M., Trippe, D. M., & Wasko, L. (2011). *Identifying best bet entry-level selection measures for US Air Force remotely piloted aircraft (RPA) pilot and sensor operator (SO) occupations* (AFCAPS-FR-2011-0013). Randolph AFB, TX: Strategic Research & Assessment Branch (AFPC/DSYX).

Ree, M. J., & Carretta, T. R. (2002). g2K. *Human Performance, 15,* 3-23.

Rose, M. R., Barron, L. G., Carretta, T. R., Arnold, R. D. and Howse, W. R., (2014). Early identification of unmanned aircraft pilots using measures of personality and aptitude. *The International Journal of Aviation Psychology, 24*(1), 36-52.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50*(3), 707-721.

U.S. Department of Labor. (1999). Testing and assessment: An employer's guide to good practices. Washington, DC: Author.

Weeks, J. L., Zelenski, W. E., & Carretta, T. R. (1996). *Advances in USAF pilot selection. Selection and Training Advances in Aviation* (AGARD-CP-588) (pp. l - 11). Prague, Czech Republic: Advisory Group for Aerospace Research and Development.

Woycheshin, D. (2001). *Validation of the Canadian Automated Pilot Selection System (CAPSS) against primary flying training results* (Rep. No. D AIR PG&T 3-6. Ottawa, ON, Canada: Chief of the Air Staff, National Defence Headquarters.

Appendix A

Table A.1

Relationship betweenPCSM, TBAS, SynWin Multiple-Task, and SynWin Constituent Single-Task Scores with Completion/Attrition through SUPT

	Complete through SUPT Primary *N* = 888
	Completion (1) Versus Attrition (0)
	r
Current PCSM Model	**.39*** (.42)**
AFOQT – Pilot	**.37*** (.40)**
No Joystick or Rudder Model	**.38*** **
PCSM/Multi-Tasking Model	**.39*** **
Flight Hour Code	**.23*** **
TBAS Components	
A (Airplane Redirects 1)	**.23*** **
AOnTarget (Distance from Target 1)	**.23*** **
AHA (Airplane Redirects 2)	**.21*** **
AHOnTarget (Distance from Target 2)	**.24*** **
AHROnTarget (Rudder Task)	**.19*** **
Unmanned Aerial Vehicle Composite	**.15*** **
Multi-Tasking Test Scores	
Single Task Memory (Practice)	.02
Single Task Math (Practice)	-.04
Single Task Visual (Practice)	.00
Single Task Listening (Practice)	**.11*** **
Overall Multi-Task Practice Performance	.05
Multi-Task Memory (Practice)	.05
Multi-Task Math (Practice)	.00
Multi-Task Visual (Practice)	.00
Multi-Task Listening (Practice)	**.08***
Overall Trial 1 Performance	**.08***
Trial 1 Memory	**.10****
Trial 1 Math	.00
Trial 1 Visual	.03
Trial 1 Listening	**.08***
Overall Trial 2 Performance	.06
Trial 2 Memory	.07*
Trial 2 Math	.02
Trial 2 Visual	.03
Trial 2 Listening	-.01
Overall Trial 3 Performance	.05
Trial 3 Memory	.04
Trial 3 Math	.05
Trial 3 Visual	.05
Trial 3 Listening	.05
Overall Trial 4 Performance	.07
Trial 4 Memory	**.08***
Trial 4 Math	.02
Trial 4 Visual	.04
Trial 4 Listening	.05
Multi-Tasking Total Score	.07
Memory Total	**.08***
Math Total	.02
Visual Total	.04
Listening Total	.05

Note. Completion n = 738; Attrition n = 150. Values in parentheses are corrected for Multivariate Range Restriction.

Table A.2

Validity of PCSM, TBAS, and SynWin Multiple-Task, and Constituent Single-Task Scores as Predictors of Daily Flight Performance, Academic Performance, and Merit Assignment Selection System (MASS) Scores for Manned Aircraft Pilots

	SUPT Primary (T6A) *N* = 756			SUPT Advanced (T1A) *N* = 210			SUPT Advanced (T38C) *N* = 107		
	Daily Flying	**Academic**	**MASS**	**Daily Flying**	**Academic**	**MASS**	**Daily Flying**	**Academic**	**MASS**
Current PCSM Model	**.33*** (.37)**	**.24*** (.31)**	**.25*** (.30)**	**.33*** (.35)**	**.18* (.22)**	**.30*** (.32)**	**.25** (.28)**	.08 (.08)	**.22* (.24)**
AFOQT – Pilot	**.31*** (.47)**	**.31*** (.42)**	**.28*** (.43)**	**.27*** (.46)**	**.20** (.29)**	**.27*** (.44)**	.19 (.28)	.12 (.20)	**.20* (.30)**
No Joystick or Rudder Model	**.33***	**.27***	**.25***	**.33***	**.19**	**.32***	**.22***	.11	.19
PCSM/Multi-Tasking Model	**.33***	**.25***	**.26***	**.33***	**.18***	**.30***	**.26**	.08	**.22***
Flight Hour Code	**.20***	**.09***	**.11**	**.23***	.09	**.23***	.19	.06	.14
TBAS Components									
A (Airplane Redirects 1)	**.20***	**.14***	**.15***	.13	.08	.11	**.21***	.03	.19
AOnTarget (Distance from Target 1)	**.17***	**.14***	**.14***	.11	.05	.07	**.21***	.13	.17
AHA (Airplane Redirects 2)	**.24***	**.15***	**.20***	.18**	.10	**.17***	**.23***	.14	**.28***
AHOnTarget (Distance from Target 2)	**.24***	**.16***	**.19***	.15*	.08	**.14***	.18	**.19***	.19
AHROnTarget (Rudder Task)	**.19***	.05	**.16***	.19**	.03	.09	**.23***	-.08	.19
Unmanned Aerial Vehicle Composite	**.11***	**.18***	**.12**	.14*	.05	.11	.03	.07	-.03
Multitasking Test Scores									
Single Task Memory (Practice)	.07	**.09**	.04	.04	**.14***	.04	-.06	.12	-.10
Single Task Math (Practice)	.02	.07	.04	-.04	-.03	-.05	.12	.02	-.02
Single Task Visual (Practice)	.03	-.01	.04	.09	.02	.06	.10	.17	.14
Single Task Listening (Practice)	.04	.04	.00	.09	.04	.07	-.02	.00	.11
Overall Multi-Task Practice Performance	**.13***	**.10**	**.15***	.04	-.04	.03	.18	.09	.15
Multi-Task Memory (Practice)	**.10**	**.09**	**.10**	.01	.01	.06	**.20***	.04	**.28****
Multi-Task Math (Practice)	.01	.08	.05	-.02	.01	.01	-.07	.01	-.07
Multi-Task Visual (Practice)	.05	.00	.06	.05	-.07	-.01	.05	.06	.02
Multi-Task Listening (Practice)	**.14***	.05	**.12***	.07	-.04	.01	.18	.10	.07
Overall Trial 1 Performance	**.17***	**.12**	**.17***	.11	.01	.09	.16	.11	.09
Trial 1 Memory	**.19***	**.16***	**.20***	.05	.09	.11	.11	.03	.12
Trial 1 Math	.06	.06	**.08***	.08	.00	.07	.02	-.03	-.04
Trial 1 Visual	.04	.01	.02	.06	-.01	-.01	.09	.11	.02
Trial 1 Listening	**.14***	.07	**.12***	.07	-.06	.07	.19	.18	.16
Overall Trial 2 Performance	**.16***	**.14***	**.18***	.09	.05	**.14***	**.21***	.12	.15
Trial 2 Memory	**.13***	**.14***	**.14***	.02	.04	.12	.17	-.01	.11
Trial 2 Math	.05	.08*	.11**	.08	.09	.14	-.01	.06	.00
Trial 2 Visual	.07	.04	.06	.10	.03	.05	.13	.12	.08
Trial 2 Listening	**.16***	**.11**	**.14***	.03	-.06	.05	**.24***	.11	**.21***
Overall Trial 3 Performance	**.19***	**.15***	**.20***	.05	-.03	.02	**.20***	.09	.08
Trial 3 Memory	**.20***	**.17***	**.20***	.01	.01	.02	**.20***	.05	.19
Trial 3 Math	.08*	.08*	.10**	.07	.03	.07	.01	.06	-.11
Trial 3 Visual	.06	.03	.05	.09	.00	.03	.15	.19	.13
Trial 3 Listening	**.14***	**.11**	**.15***	-.04	-.10	-.05	.19	-.13	.03
Overall Trial 4 Performance	**.17***	**.15***	**.19***	.02	.01	.07	.13	.12	.15
Trial 4 Memory	**.21***	**.18***	**.22***	.09	.07	.10	**.20***	.02	**.26****
Trial 4 Math	.05	**.08***	**.10**	.03	.05	.05	-.03	-.01	-.04
Trial 4 Visual	.06	.03	.06	.10	-.03	.05	.10	.17	.06
Trial 4 Listening	**.13***	**.09***	**.13***	**-.16***	-.08	-.02	.08	.15	.12
Multi-Tasking Total Score	**.19***	**.15***	**.20***	.07	.01	.09	.19*	.12	.13
Memory Total	**.21***	**.19***	**.22***	.05	.06	.10	**.20***	.02	**.20***
Math Total	.07	**.09***	**.11**	.07	.05	.09	.00	.02	-.06
Visual Total	.06	.03	.05	.09	.00	.03	.12	.15	.08
Listening Total	**.19***	**.13***	**.18***	-.03	-.10	.01	**.26****	.11	**.19***

Note. Values in parentheses are corrected for Multivariate Range Restriction. ** $p < .05$; ** $p < .01$; *** $p < .001$.

Table A.3

Validity of PCSM, TBAS, and Synwin Multiple-Task, and Constituent Single-Task Scores as Predictors of Daily Flight Performance, Academic Performance, and Merit Assignment Selection System (MASS) Scores for Remotely Piloted Aircraft (RPA) Pilots

	RFS *N* = 449			RIQ *N* = 370		
	Daily Flying	**Academic**	**MASS**	**Daily Flying**	**Academic**	**MASS**
Current PCSM Model	**.31*** (.42)**	**.21*** (.28)**	**.35*** (.41)**	**.22*** (.26)**	**.23*** (.30)**	**.27*** (.32)**
AFOQT – Pilot	**.24*** (.36)**	**.24*** (.29)**	**.30*** (.41)**	**.23*** (.31)**	**.24*** (.35)**	**.27*** (.32)**
No Joystick or Rudder Model	**.29*** **	**.25*** **	**.35*** **	**.22*** **	**.26*** **	**.28*** **
PCSM/Multi-Tasking Model	**.32*** **	**.21*** **	**.36*** **	**.23*** **	**.23*** **	**.27*** **
Flight Hour Code	**.18*** **	.08	**.19*** **	.04	.11	**.10***
TBAS Components						
A (Airplane Redirects 1)	**.16*** **	.00	**.17*** **	**.18*** **	.08	**.15****
AOnTarget (Distance from Target 1)	**.15****	-.04	**.12****	**.14****	.06	**.11***
AHA (Airplane Redirects 2)	**.21*** **	.05	**.18*** **	**.17****	.08	**.13***
AHOnTarget (Distance from Target 2)	**.21*** **	.00	**.17*** **	**.14****	.06	.10
AHROnTarget (Rudder Task)	**.16*** **	-.01	**.13****	**.15****	.01	.09
Unmanned Aerial Vehicle Composite	**.11***	.09	**.16*** **	**.15****	.17	**.17****
Multi-Tasking Test Scores						
Single Task Memory (Practice)	.04	-.01	.01	**.12***	.02	.03
Single Task Math (Practice)	.07	-.04	.06	**.15****	.04	.12
Single Task Visual (Practice)	.01	-.02	.02	.05	-.05	.02
Single Task Listening (Practice)	.03	.07	.06	.07	**.13***	.10
Overall Multi-Task Practice Performance	.08	.03	.08	**.16****	.05	.10
Multi-Task Memory (Practice)	.07	.08	.08	**.12***	**.13***	.09
Multi-Task Math (Practice)	.06	.03	.04	**.19*** **	.06	**.12***
Multi-Task Visual (Practice)	.00	.03	-.01	.01	-.08	-.03
Multi-Task Listening (Practice)	.05	-.01	**.18*** **	.05	.01	.04
Overall Trial 1 Performance	**.11***	.07	**.11***	**.18*** **	.04	**.15****
Trial 1 Memory	.09	.05	.06	**.15****	.03	.10
Trial 1 Math	.01	.01	.03	**.17*** **	.01	.10*
Trial 1 Visual	.04	-.01	.02	.02	-.04	.00
Trial 1 Listening	**.16*** **	**.12***	.14	**.13***	.10*	**.18*** **
Overall Trial 2 Performance	**.13***	.06	**.11***	**.18*** **	.05	**.14****
Trial 2 Memory	.08	.04	.07	**.15****	.13*	**.14****
Trial 2 Math	.03	.04	.04	**.17****	.07	.09
Trial 2 Visual	.07	.01	.03	.01	-.07	.00
Trial 2 Listening	**.17*** **	.07	**.14****	**.17****	.02	**.16****
Overall Trial 3 Performance	**.11***	.03	.12*	**.18*** **	.05	**.11***
Trial 3 Memory	.13	.06	**.14****	**.16****	.10*	.10
Trial 3 Math	.03	-.03	.06	**.18*** **	.05	.09
Trial 3 Visual	.05	.00	.02	.02	-.06	-.01
Trial 3 Listening	.08	.06	**.11***	**.11***	.04	**.10***
Overall Trial 4 Performance	.07	.03	.05	**.13***	.07	.08
Trial 4 Memory	.07	.02	.03	**.11***	.09	.06
Trial 4 Math	-.03	.01	.00	**.12***	.02	.06
Trial 4 Visual	.04	.01	.02	.00	-.04	-.01
Trial 4 Listening	**.10***	.04	**.09***	**.11***	.12*	.09
Multi-Tasking Total Score	.12	.05	.11*	**.19*** **	.06	**.13***
Memory Total	.11	.05	.09	**.17****	.10*	.11
Math Total	.01	.01	.04	**.18*** **	.04	.10
Visual Total	.05	.00	.02	.01	-.06	-.01
Listening Total	**.17*** **	.10*	**.17*** **	**.17****	.09	**.18*** **

Note. Values in parentheses are corrected for Multivariate Range Restriction. ** $p < .05$; ** $p < .01$; *** $p < .001$.

Table A.4

Male-Female Effect Size Differences across PCSM and Multi-Tasking Scores for Training Applicant Sample

	Male (*N* = 12,451)		Female (*N* = 1,763)		*d*
	Mean	**SD**	**Mean**	**SD**	
Existing PCSM Score	45.07	27.34	27.63	24.31	0.65
AFOQT Pilot	71.35	21.89	56.53	23.19	0.67
No Joystick or Rudder Model	1.39	1.30	0.61	1.31	0.60
PCSM/Multi-Tasking Model	0.77	0.15	0.67	0.14	0.65
Flight Hour Code	2.12	2.91	1.65	2.54	0.16
TBAS Components					
A (Airplane Redirects 1)	10.17	4.43	4.34	3.28	1.35
AOnTarget (Distance from Target 1)	648.06	141.54	431.03	154.58	1.52
AHA (Airplane Redirects 2)	7.88	5.25	2.48	2.74	1.08
AHOnTarget (Distance from Target 2)	906.43	287.16	494.96	264.68	1.45
AHROnTarget (Rudder Task)	1087.17	219.29	1016.36	242.99	0.32
Unmanned Aerial Vehicle Composite	0.28	1.01	-0.17	1.13	0.43
Multi-Tasking Test Scores					
Single Task Memory (Practice)	-0.01	1.01	0.06	0.92	-0.07
Single Task Math (Practice)	0.01	0.99	-0.05	1.04	0.06
Single Task Visual (Practice)	0.02	0.99	-0.10	1.05	0.11
Single Task Listening (Practice)	0.00	0.99	-0.01	1.05	0.01
Overall Multi-Task Practice Performance	0.01	0.59	-0.10	0.61	0.20
Multi-Task Memory (Practice)	0.03	0.99	-0.20	1.08	0.23
Multi-Task Math (Practice)	0.01	1.00	-0.08	0.99	0.09
Multi-Task Visual (Practice)	0.00	1.00	0.01	1.02	-0.01
Multi-Task Listening (Practice)	0.02	0.99	-0.14	1.05	0.16
Overall Trial 1 Performance	0.02	0.64	-0.11	0.66	0.19
Trial 1 Memory	0.02	0.98	-0.13	1.10	0.15
Trial 1 Math	0.01	0.99	-0.06	1.05	0.07
Trial 1 Visual	0.00	1.00	-0.02	0.99	0.02
Trial 1 Listening	0.03	0.99	-0.21	1.05	0.25
Overall Trial 2 Performance	0.01	0.66	-0.10	0.67	0.17
Trial 2 Memory	0.01	0.99	-0.06	1.05	0.06
Trial 2 Math	0.01	1.00	-0.06	1.03	0.07
Trial 2 Visual	0.01	1.00	-0.06	1.00	0.07
Trial 2 Listening	0.03	0.99	-0.22	1.03	0.26
Overall Trial 3 Performance	0.01	0.66	-0.10	0.67	0.17
Trial 3 Memory	0.01	0.99	-0.05	1.05	0.05
Trial 3 Math	0.01	0.99	-0.06	1.04	0.07
Trial 3 Visual	0.01	1.00	-0.06	1.01	0.07
Trial 3 Listening	0.03	0.99	-0.22	1.04	0.25
Overall Trial 4 Performance	0.01	0.66	-0.10	0.68	0.17
Trial 4 Memory	0.00	0.99	-0.03	1.04	0.03
Trial 4 Math	0.01	0.99	-0.06	1.04	0.07
Trial 4 Visual	0.01	1.00	-0.08	1.02	0.09
Trial 4 Listening	0.03	0.99	-0.23	1.05	0.26
Multi-Tasking Total Score	0.01	0.60	-0.10	0.61	0.19
Memory Total	0.01	0.77	-0.06	0.84	0.09
Math Total	0.01	0.88	-0.06	0.92	0.08
Visual Total	0.01	0.87	-0.05	0.87	0.07
Listening Total	0.03	0.94	-0.22	1.00	0.27

Cohen's (1988) *d* Interpretation: .2 = Small; .5 = Medium; .8 = Large.

Table A.5

Socioeconomic Status (Average or Higher versus Low) Effect Size Differences across PCSM and Multi-Tasking Scores

	Average or Higher (*N* = 7,641)		Low (*N* = 1,741)		*d*
	Mean	SD	Mean	SD	
Existing PCSM Score	45.92	27.45	38.17	26.51	0.28
AFOQT Pilot	72.15	21.82	66.06	22.92	0.28
No Rudder or Joystick Model	1.45	1.30	1.07	1.30	0.30
PCSM/Multi-Tasking Model	0.77	0.15	0.73	0.15	0.28
Flight Hour Code	2.31	2.94	1.70	2.65	0.21
TBAS Components					
A (Airplane Redirects 1)	9.44	4.67	9.32	4.74	0.02
AOnTarget (Distance from Target 1)	623.08	158.83	614.97	161.88	0.05
AHA (Airplane Redirects 2)	7.32	5.29	6.76	5.12	0.11
AHOnTarget (Distance from Target 2)	864.84	312.04	831.13	313.49	0.11
AHROnTarget (Rudder Task)	1084.07	218.06	1059.21	219.46	0.11
Unmanned Aerial Vehicle Composite	0.25	1.01	0.15	1.044	0.09
Multi-Tasking Test Scores					
Single Task Memory (Practice)	0.01	0.97	-0.03	1.10	0.05
Single Task Math (Practice)	0.02	1.00	-0.05	0.99	0.07
Single Task Visual (Practice)	0.01	0.98	-0.01	1.03	0.02
Single Task Listening (Practice)	0.01	0.96	-0.08	1.12	0.09
Overall Multi-Task Practice Performance	0.01	0.59	-0.01	0.58	0.04
Multi-Task Memory (Practice)	0.02	1.00	-0.02	0.95	0.04
Multi-Task Math (Practice)	0.03	0.99	-0.04	1.00	0.07
Multi-Task Visual (Practice)	0.01	1.00	-0.01	1.00	0.02
Multi-Task Listening (Practice)	0.00	1.00	0.03	1.00	-0.03
Overall Trial 1 Performance	0.01	0.64	-0.02	0.64	0.05
Trial 1 Memory	0.02	0.99	-0.04	0.99	0.06
Trial 1 Math	0.02	0.99	-0.06	0.98	0.08
Trial 1 Visual	0.03	0.98	0.00	1.02	0.02
Trial 1 Listening	-0.01	1.00	0.02	1.02	-0.03
Overall Trial 2 Performance	0.02	0.65	-0.02	0.65	0.06
Trial 2 Memory	0.02	0.99	-0.01	0.99	0.03
Trial 2 Math	0.03	0.99	-0.05	1.00	0.08
Trial 2 Visual	0.03	0.97	0.00	1.00	0.03
Trial 2 Listening	0.01	1.00	-0.01	1.00	0.02
Overall Trial 3 Performance	0.02	0.65	-0.02	0.66	0.06
Trial 3 Memory	0.01	0.99	-0.03	0.99	0.04
Trial 3 Math	0.03	0.99	-0.04	1.00	0.07
Trial 3 Visual	0.03	0.97	0.00	0.99	0.03
Trial 3 Listening	0.01	0.99	0.00	0.98	0.00
Overall Trial 4 Performance	0.01	0.66	-0.02	0.69	0.05
Trial 4 Memory	0.01	0.99	-0.02	1.01	0.03
Trial 4 Math	0.02	0.98	-0.05	1.01	0.08
Trial 4 Visual	0.03	0.98	-0.01	1.06	0.04
Trial 4 Listening	0.00	0.99	0.01	1.02	-0.01
Multi-Tasking Total Score	0.02	0.59	-0.02	0.60	0.06
Memory Total	0.01	0.86	-0.02	0.88	0.04
Math Total	0.03	0.87	-0.05	0.88	0.09
Visual Total	0.03	0.93	0.00	0.97	0.03
Listening Total	0.00	0.77	0.01	0.79	0.00

Cohen's (1988) *d* Interpretation: .2 = Small; .5 = Medium; .8 = Large.

Table A.6

Race Effect Size Differences across PCSM and Multi-Tasking Scores: White versus Black/African-American

	White (*N* = 11,803)		Black (*N* = 627)		*d*
	Mean	SD	Mean	SD	
Existing PCSM Score	44.93	27.32	24.51	25.42	0.75
AFOQT Pilot	71.34	21.67	49.24	24.27	1.01
No Joystick or Rudder Model	1.40	1.29	0.25	1.40	0.89
PCSM/Multi-Tasking Model	0.76	0.15	0.65	0.14	0.75
Flight Hour Code	2.17	2.92	1.57	2.70	0.21
TBAS Components					
A (Airplane Redirects 1)	9.67	4.67	7.86	4.60	0.39
AOnTarget (Distance from Target 1)	628.84	156.93	564.95	170.70	0.41
AHA (Airplane Redirects 2)	7.46	5.35	5.33	4.53	0.40
AHOnTarget (Distance from Target 2)	869.98	312.15	729.75	310.41	0.45
AHROnTarget (Rudder Task)	1087.19	221.40	1007.92	220.84	0.36
Unmanned Aerial Vehicle Composite	0.26	1.01	-0.41	1.21	0.65
Multi-Tasking Test Scores					
Single Task Memory (Practice)	0.01	0.98	-0.09	1.06	0.10
Single Task Math (Practice)	0.00	1.00	-0.15	1.00	0.15
Single Task Visual (Practice)	0.01	1.00	-0.10	1.06	0.11
Single Task Listening (Practice)	0.03	0.93	-0.19	1.33	0.23
Overall Multi-Task Practice Performance	0.01	0.59	-0.15	0.57	0.27
Multi-Task Memory (Practice)	0.02	1.00	-0.26	0.96	0.28
Multi-Task Math (Practice)	0.00	1.00	-0.21	0.98	0.21
Multi-Task Visual (Practice)	0.00	1.00	-0.12	1.05	0.13
Multi-Task Listening (Practice)	0.01	0.99	-0.01	1.04	0.02
Overall Trial 1 Performance	0.00	0.64	-0.16	0.61	0.27
Trial 1 Memory	0.01	0.99	-0.29	0.96	0.30
Trial 1 Math	0.00	1.00	-0.22	0.99	0.22
Trial 1 Visual	0.01	1.00	-0.15	0.99	0.15
Trial 1 Listening	0.01	0.99	0.00	1.07	0.01
Overall Trial 2 Performance	0.00	0.66	-0.19	0.63	0.29
Trial 2 Memory	0.01	1.00	-0.31	0.99	0.32
Trial 2 Math	0.00	1.00	-0.30	1.03	0.30
Trial 2 Visual	0.01	1.00	-0.14	0.97	0.15
Trial 2 Listening	0.00	0.99	0.00	1.01	0.00
Overall Trial 3 Performance	0.01	0.66	-0.19	0.64	0.30
Trial 3 Memory	0.02	0.99	-0.38	1.05	0.40
Trial 3 Math	0.00	1.00	-0.24	1.03	0.23
Trial 3 Visual	0.01	0.99	-0.13	0.95	0.14
Trial 3 Listening	0.00	0.99	-0.03	1.03	0.03
Overall Trial 4 Performance	0.01	0.66	-0.21	0.66	0.34
Trial 4 Memory	0.01	0.99	-0.39	1.06	0.40
Trial 4 Math	0.00	0.99	-0.29	1.02	0.30
Trial 4 Visual	0.01	0.99	-0.13	0.95	0.14
Trial 4 Listening	0.01	0.99	-0.04	1.01	0.05
Multi-Tasking Total Score	0.01	0.59	-0.19	0.58	0.33
Memory Total	0.01	0.86	-0.34	0.87	0.41
Math Total	0.00	0.88	-0.26	0.91	0.30
Visual Total	0.01	0.95	-0.14	0.92	0.15
Listening Total	0.00	0.77	-0.02	0.80	0.03

Cohen's (1988) *d* Interpretation: .2 = Small; .5 = Medium; .8 = Large.

Table A.7

Ethnicity Effect Size Differences across PCSM and Multi-Tasking Scores: Non-Hispanic versus Hispanic

	Non-Hispanic (*N* = 12,529)		Hispanic (*N* = 1,595)		*d*
	Mean	SD	Mean	SD	
Existing PCSM Score	43.98	27.56	34.99	26.38	0.33
AFOQT Pilot	70.54	22.25	61.93	23.52	0.38
No Joystick or Rudder Model	1.35	1.31	0.88	1.33	0.36
PCSM/Multi-Tasking Model	0.76	0.16	0.71	0.15	0.33
Flight Hour Code	2.12	2.90	1.67	2.62	0.16
TBAS Components					
A (Airplane Redirects 1)	9.50	4.71	9.05	4.70	0.09
AOnTarget (Distance from Target 1)	622.38	159.29	611.28	164.84	0.07
AHA (Airplane Redirects 2)	7.28	5.35	6.70	5.10	0.11
AHOnTarget (Distance from Target 2)	858.15	314.42	835.69	321.90	0.07
AHROnTarget (Rudder Task)	1081.69	223.46	1054.04	223.93	0.12
Unmanned Aerial Vehicle Composite	0.24	1.03	0.08	1.06	0.15
Multi-Tasking Test Scores					
Single Task Memory (Practice)	0.01	0.98	-0.07	1.10	0.08
Single Task Math (Practice)	0.02	1.00	-0.11	0.97	0.12
Single Task Visual (Practice)	0.01	0.99	-0.06	1.03	0.07
Single Task Listening (Practice)	0.01	0.98	-0.09	1.12	0.11
Overall Multi-Task Practice Performance	0.01	0.59	-0.07	0.61	0.14
Multi-Task Memory (Practice)	0.01	1.00	-0.08	1.01	0.09
Multi-Task Math (Practice)	0.01	1.00	-0.11	1.01	0.12
Multi-Task Visual (Practice)	0.01	0.99	-0.06	1.05	0.07
Multi-Task Listening (Practice)	0.01	1.00	-0.03	1.03	0.03
Overall Trial 1 Performance	0.01	0.64	-0.07	0.64	0.13
Trial 1 Memory	0.01	1.00	-0.08	0.98	0.10
Trial 1 Math	0.02	1.00	-0.12	0.99	0.14
Trial 1 Visual	0.01	0.99	-0.07	1.04	0.09
Trial 1 Listening	0.00	1.00	-0.02	1.02	0.02
Overall Trial 2 Performance	0.01	0.65	-0.08	0.68	0.15
Trial 2 Memory	0.02	1.00	-0.10	1.00	0.12
Trial 2 Math	0.02	1.00	-0.12	1.02	0.14
Trial 2 Visual	0.01	1.00	-0.07	1.03	0.09
Trial 2 Listening	0.01	0.99	-0.04	1.07	0.04
Overall Trial 3 Performance	0.01	0.66	-0.09	0.68	0.16
Trial 3 Memory	0.02	1.00	-0.12	1.02	0.14
Trial 3 Math	0.02	1.00	-0.13	1.00	0.15
Trial 3 Visual	0.01	1.00	-0.09	1.03	0.11
Trial 3 Listening	0.01	0.99	-0.03	1.06	0.04
Overall Trial 4 Performance	0.02	0.67	-0.11	0.68	0.18
Trial 4 Memory	0.02	1.00	-0.11	1.01	0.12
Trial 4 Math	0.02	1.00	-0.15	1.01	0.18
Trial 4 Visual	0.02	1.00	-0.10	1.03	0.12
Trial 4 Listening	0.01	0.99	-0.06	1.06	0.07
Multi-Tasking Total Score	0.01	0.60	-0.09	0.61	0.17
Memory Total	0.02	0.87	-0.10	0.87	0.14
Math Total	0.02	0.88	-0.13	0.90	0.17
Visual Total	0.01	0.95	-0.09	0.99	0.10
Listening Total	0.01	0.77	-0.04	0.84	0.06

Cohen's (1988) *d* Interpretation: .2 = Small; .5 = Medium; .8 = Large.