AFCAPS-TR-2013-0007

Development of Two- and Three-Factor Classification Models for Air Force Battlefield Airmen (BA) and Related AFSs

May 31, 2013

Mark R. Rose
Gregory G. Manley
Johnny J. Weissmuller

Air Force Personnel Center
Strategic Research and Assessment
HQ AFPC/DSYX

Prepared for:

Kenneth L. Schwartz

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

This page intentionally left blank.

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 and is releasable to the Defense Technical Information Center. The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

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/

Available for public release. Distribution Unlimited. Please contact AFPC/DSYX Strategic Research and Assessment 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 R. Carretta AFMC 711 HPW/RHCI, Associate Editor Dr. Gregory Manley HQ AFPC/DSYX, Dr. Lisa Hughes AF/A1PF, Dr. Paul DiTullio AF/A1PF, Kenneth Schwartz HQ AFPC/DSYX, Johnny Weissmuller HQ AFPC/DSYX, Dr. Laura Barron HQ AFPC/DSYX, Dr. Mark Rose HQ AFPC/DSYX, and Brian Chasse 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)***
05-31-2013	Technical	

4. TITLE AND SUBTITLE

Development of Two- and Three-Factor Classification Models for Air Force Battlefield Airmen (BA) and Related AFSs

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Mark R. Rose, Gregory G. Manley, Johnny J. Weissmuller

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Air Force Personnel Center
Strategic Research and Assessment Branch
Randolph AFB TX 78150

8. PERFORMING ORGANIZATION REPORT

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

Air Force Personnel Center
Strategic Research and Assessment Branch
Randolph AFB TX 78150

10. SPONSOR/MONITOR'S ACRONYM(S)

HQ AFPC/DSYX

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

AFCAPS-FR-2013-0007

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release. Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report summarizes development of statistical models for classification of Air Force Battlefield Airmen (BA) and related Air Force Specialties (AFSs), including pararescue (PJ), combat control (CCT), explosive ordnance disposal (EOD), special operations weather (SOWT), survival, evasion, resistance, and escape (SERE), and tactical air control party (TACP). Results generally supported the criterion-related validity of the Tailored Adaptive Personality Assessment System (TAPAS), Armed Services Vocational Aptitude Battery (ASVAB), and Physical Ability and Stamina Test (PAST) for classification of applicants into these AFSs.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: Unclassified			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Kenneth L. Schwartz
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	U	29	**19b. TELEPHONE NUMBER *(include area code)***
U	U	U			210-565-3139

Standard Form 298 (Rev. 8-98)

Contents

List of Tables

List of Figures

Executive Summary

This report summarizes development of statistical models for classification of Air Force Battlefield Airmen (BA) and related Air Force Specialties (AFSs), including pararescue (PJ), combat control (CCT), explosive ordnance disposal (EOD), special operations weather (SOWT), survival, evasion, resistance, and escape (SERE), and tactical air control party (TACP). Results generally supported the criterion-related validity of the Tailored Adaptive Personality Assessment System (TAPAS), Armed Services Vocational Aptitude Battery (ASVAB), and Physical Ability and Stamina Test (PAST) for classification of applicants into these AFSs. Table 1 summarizes model effect size and adverse impact potential by AFS.

The remainder of this report describes the measures and analyses used, statistical results, and recommendations for implementation and ongoing monitoring and evaluation. Methodology used throughout this study was guided by best practices in selection and classification, based on the *Uniform Guidelines on Employee Selection Procedures* (*Guidelines*; Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978), *Principles for the Validation and Use of Personnel Selection Procedures* (*Principles*; Society for Industrial and Organizational Psychology, 2003), and the *Standards for Psychological and Educational Testing* (*Standards*; American Educational Research Institution, American Psychological Association, & National Council on Measurement Education, 1999). Appendix A cross-references information from the current study with requirements for documentation of impact and validity based on the *Guidelines* (1978).

Table 1. *Predictive Validation Summary*

Air Force Specialty (AFS)	*R*	R^2	Cohen's *d* - Ethnicity
Pararescue (PJ)	.497**	.247	.318
Combat Control (CCT)	.483**	.233	.374
Explosive Ordnance Disposal (EOD)	.461**	.213	—
Survival, Evasion, Resistance, and Escape (SERE)	.597**	.356	.110
Special Operations Weather (SOWT)[A]	.264*	.069	-.053
Tactical Air Control Party (TACP)	.487**	.237	-.020

Note. [A]Based on two-factor ASVAB/PAST model only; *d* values calculated only for subgroups with $n>30$; positive *d* values indicate total scores favored majority, and negative *d* values indicate total scores favored minority; $^{*}p<.01$ $^{**}p<.001$.

Development of Classification Models for Battlefield Airmen and Related AFSs

1 Purpose and Overview

Air Force Battlefield Airmen (BA) and related career fields have experienced attrition rates as high as 90% from accession through initial entry training over the last several years. This attrition has generally been driven by (1) the qualifications of candidates selected into training, relative to training demands, and (2) the ability of the training system to develop candidates to meet operational demands. Development of the classification models described in this report was intended to address the qualifications of candidates selected portion of the overall problem, and to provide valuable information to complement training efforts aimed at delivering mission-ready warfighters to the field. Classification models developed incorporated cognitive ability and knowledge, physical ability, and personality trait assessments, and covered the following AFSs:

- Pararescue (PJ) - 1T2X1
- Combat Control (CCT) - 1C2X1
- Explosive Ordnance Disposal (EOD) - 3E8X1
- Survival, Evasion, Resistance, and Escape (SERE) - 1T0X1
- Special Operations Weather (SOWT) - 1W0X2
- Tactical Air Control Party (TACP) - 1C4X1

This report describes the (1) specific measures evaluated, (2) analyses performed, (3) validity and adverse impact potential of measures by AFS, and (4) AFPC/DSYX's recommendations for implementation and ongoing monitoring and evaluation. Appendix A summarizes study procedures and results by cross-referencing them with requirements for documentation of impact and validity based on section 15 of the *Guidelines* (1978).

2 Description of Measures Evaluated

This study evaluated use of subtest scores from the Tailored Adaptive Personality Assessment System (TAPAS), Armed Services Vocational Aptitude Battery (ASVAB), and the Physical Ability and Stamina Test (PAST), to predict initial course(s) of entry success/failure. All data were collected between July 2008 and May 2013. As TAPAS is the newest of these measures, it is described in detail next, followed by briefer descriptions of ASVAB, PAST, and course graduation/elimination criteria.

2.1 TAPAS

2.1.1 TAPAS Scales and Design

The AF TAPAS is a DOD-owned, non-cognitive/personality measurement system rooted in the Big Five theory of personality, containing 15 scales (see Table 2) designed to assess personality factors related to performance in military specialties. The instrument builds on the Army's Assessment for Individual Motivation (AIM; White & Young, 1998) and incorporates features that address problems associated with more traditional Likert scale measures of personality traits, including faking, limitations of classical test theory (CTT), and test compromise.

To reduce faking, TAPAS uses a forced-choice response format (multidimensional pairwise preference item format; MDPP) that pairs items similar in social desirability but different in measured construct. The MDPP items are developed from pools of precalibrated personality statements that measure construct dimensions relevant to performance in the military (facets). Respondents are instructed to choose the statement in each pair that is "more like me" and must make a choice even if they find it difficult to do so.

To achieve better measurement precision and avoid limitations of CTT, TAPAS utilizes the Generalized Graded Unfolding Model (GGUM; Roberts, Donoghue, & Laughlin, 2000), an IRT method based on ideal point methodology. Whereas CTT models tend to highlight statements in a pool having high item-total correlations and linear factor loadings, ideal point models not only identify those but also discriminating statements that reflect positions of neutrality or moderation (Chernyshenko, Stark, Drasgow, & Roberts, 2007). Consequently the pool of stimuli available

Table 2. *AF Tailored Adaptive Personality Assessment System (TAPAS) Scales*

TAPAS Scale	Description
1. Achievement	High scoring individuals are seen as hard working, ambitious, confident, and resourceful.
2. Adjustment	High scoring individuals are worry free, and handle stress well; low scoring individuals are generally high strung, self-conscious, and apprehensive.
3. Cooperation	High scoring individuals are trusting, cordial, non-critical, and easy to get along with.
4. Dominance	High scoring individuals are domineering, “take charge” and are often referred to by their peers as "natural leaders."
5. Even Tempered	High scoring individuals tend to be calm and stable. They don’t often exhibit anger, hostility, or aggression.
6. Attention Seeking	Individuals scoring high on this facet tend to engage in behaviors that attract social attention; they are loud, talkative, entertaining, and even boastful.
7. Selflessness	High scoring individuals are generous with their time and resources.
8. Intellectual Efficiency	Individuals scoring high on this facet are able to process information quickly and would be described by others as knowledgeable, astute, and intellectual.
9. Non-Delinquency	High scoring individuals tend to comply with rules, customs, norms, and expectations, and they tend not to challenge authority.
10. Order	High scoring individuals tend to organize tasks and activities and desire to maintain neat and clean surroundings.
11. Physical Conditioning	High scoring individuals tend to engage in activities to maintain their physical fitness and are more likely to participate in vigorous sports or exercise.
12. Self Control	Individuals scoring high on this facet tend to be cautious, levelheaded, able to delay gratification, and patient.
13. Sociability	High scoring individuals tend to seek out and initiate social interactions.
14. Tolerance	Individuals scoring high on this facet are interested in other cultures and opinions that may differ from their own.
15. Optimism	High scoring individuals have a positive outlook on life and tend to experience joy and a sense of well-being.

for MDPP test construction is expanded when using an ideal point model for statement calibration, and rank ordering of individuals on traits is improved. GGUM is one of the most flexible ideal point models developed to date and it has been shown to fit data for individual personality statements well in previous investigations (Chernyshenko, Stark, Prewett, Gray, Stilson, & Tuttle, 2009; Stark, Chernyshenko, Drasgow, & Williams, 2006).

To reduce potential for test compromise, and administration time, TAPAS items are administered in an adaptive format. In adaptive testing with MDPP, the goal is to construct items by selecting pairs of statements so that they are highly informative about the respondent's standing on the traits assessed, given the current estimates of his or her trait values. In this way, it is possible to substantially reduce the number of items required for accurate trait estimation, and in return reduce administration time and item exposure. Computerized adaptive testing can also increase test security by imposing “exposure controls” that limit how often individual statements or items are presented to different examinees.

2.1.2 Previous Studies of TAPAS Validity

In addition to the AF-specific findings discussed in this report, Army field study results indicate that TAPAS scales significantly predict a number of criteria of interest, and demonstrate considerable incremental validity for adjustment, graduation, and attrition criteria.

Military evaluation of TAPAS originated from the Army Research Institute’s (ARI) longitudinal research project that began in 2006, which focused on examination of the validity of non-cognitive measures for predicting Army outcomes. The goal of the Army Class (Validating Future Force Performance Measures) research program was to explore the use of several experimental measures for selection and military occupational specialty (MOS) classification. The TAPAS was included in this effort and a version of the TAPAS was administered to new Soldiers in 2007 and 2008. Criterion data were also collected for each individual in the Army Class database. Initial results showed that the TAPAS provided significant incremental validity over the ASVAB for predicting attrition, end of training criteria, and in-unit performance (Knapp & Heffner, 2009; Knapp, Owens, Allen, 2011). This research also showed that the TAPAS provided non-trivial gains in classification efficiency over the ASVAB alone.

The U.S. Army's Expanded Enlistment Eligibility Metrics (EEEM) research project (Knapp & Heffner, 2010), conducted from 2007-2009 in conjunction with ARI's Army Class longitudinal validation, provided additional evidence for TAPAS prediction of important Army criteria. For example, when TAPAS trait scores were added into a regression analysis based on a sample of several hundred Soldiers, the multiple correlation increased by .26 for the prediction of physical fitness, by .16 for the prediction of disciplinary incidents, and by .20 for the prediction of 6-month attrition (Allen, Cheng, Putka, Hunter, & White, 2010). None of these criteria were predicted well by ASVAB cognitive ability scores alone (predictive validity estimates were consistently below .10).

Subsequently, based on results of the Army Class and EEEM research, and unique advantages of TAPAS (e.g., flexibility and resistance to faking), the Army chose to implement TAPAS in an applicant environment. This allowed use of TAPAS as part of an initial Education Tier One Performance Screen (TOPS; aimed at ASVAB Air Force Qualifying Test Category IIIB applicants, and later Category IV applicants) that had promise for selecting highly qualified soldiers with little adverse impact. It also allowed for evaluation of TAPAS' effectiveness as a high stakes selection and classification tool for specific MOS.

Follow-up evaluations using the TOPS data, across the four largest MOS in the dataset (Infantry-11B, Combat Medics-68W, Military Police-31B, and Motor Transport Operators-88M), showed TAPAS scores were useful predictors of can-do, will-do, and attrition outcomes (Nye, et al., 2012). MOS-specific TAPAS composites were correlated with a number of important behaviors such as attrition, job knowledge scores, and disciplinary incidents. In addition, quintile plots showed that use of TAPAS had important implications for reducing attrition. For example, plots of relationships between TAPAS and attrition showed that attrition rates for Soldiers in the bottom TAPAS quintile were approximately 300% higher than for Soldiers in the highest quintile.

Beyond the large and growing body of evidence supporting TAPAS' criterion-related validity for a range of military occupations, reviews of job descriptions (HQ AFPC, 2013), occupational analysis reports and briefings (e.g., Fisk, 2013), and career field education and training plans (USAF, 2006, 2008, 2009, 2010a, 2010b, 2010c) indicated likely relevance for TAPAS in predicting training and job outcomes for BA and related AFSs. Based on these sources, TAPAS dimensions linked to leadership effectiveness, adaptability, and fitness performance (Drasgow, Stark, Chernyshenko, Nye, Hulin, & White, 2012) appeared promising for matching applicants to BA and related career fields.

2.2 ASVAB

The ASVAB was developed specifically for the selection and classification of military personnel (Campbell & Knapp, 2010), and has consistently been observed to predict performance in military jobs (e.g., Ree, Earles, & Teachout, 1994). The ASVAB includes nine subtests with verbal, math, technical knowledge, and spatial content. The tests are General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Mathematics Knowledge (MK), Electronics Information (EI), Auto and Shop Information (AS), Mechanical Comprehension (MC), and Assembling Objects (AO).

Applicants must currently meet a minimum score on the Armed Services Qualification Test (AFQT), a common qualifying exam for all Services based on four ASVAB subtests (WK, PC, AR, MK), to qualify for entry into the Air Force. For qualification into specific careers, including BA and related AFSs, applicants must also meet minimum scores (see Table 3) on one or more of the composites used for classification across the Air Force (Mechanical, Administrative, General, Electronics).

2.3 PAST

PAST components and requirements are established separately for each AFS (see Table 3). PJ, CCT, and SOWT applicants complete a timed swim, a timed run, pull-ups, push-ups, and sit-ups. EOD, SERE, and TACP applicants complete the same subtests with the exception of the timed swim. Failure in any single subtest results in an overall failure to qualify.

Table 3. *ASVAB and PAST Qualifying Scores by AFS*

Air Force Specialty (AFS)	ASVAB MAGE	Pull-ups	Push-ups	Sit-ups	1.5-mile Run	20meter Underwater Swim x 2	0.5k Swim
1T231/PJ	G44	10	52	54	9:47	Pass	10:07
1C231/CCT	M55&G55	8	48	48	10:10	Pass	11:42
3E831/EOD	M60&G64	3	35	50	11:00	--	--
1T031/SERE	G55	8	48	48	11:00	--	--
1W032/SOWT	G66&E50	8	48	48	10:10	Pass	14:00
1C431/TACP	G49	6	40	48	10:47	--	--

Note. M = Mechanical; A = Administrative; G = General; E = Electronics

2.4 Course Graduation/Elimination

Course graduation/elimination was scored as a dichotomous training outcome, with 0 = Elimination and 1 = Graduation. Course graduation rates by AFS were 10.0% (PJ), 47.0% (CCT), 49.1% (EOD), 17.8% (SERE), 43.5% (SOWT), and 67.6% (TACP). Table 4 lists representative course titles, and course locations for each AFS. As additional data mature, evaluations will be conducted on course attrition later in the training pipeline.

Table 4. *Predictive Validation Criteria by AFS*

Air Force Specialty (AFS)	Representative Course Titles	Course Location(s)
PJ	Pararescue Development Pararescue Indoctrination	Lackland AFB, TX
CCT	Combat Control Selection	Lackland AFB, TX
EOD	Explosive Ordnance Disposal Preliminary	Sheppard AFB, TX
SERE	SERE Specialist Selection	Lackland AFB, TX
SOWT	Special Operations Weather Team Selection	Lackland AFB, TX
TACP	Terminal Attack Control Party Preparatory	Lackland AFB, TX

3 Analyses

Analyses focused on the relation of TAPAS, ASVAB, and PAST scores to course graduation/elimination. Analyses were conducted separately for each AFS, beginning with examination of descriptive statistics, distribution shapes, and outliers for all variables. Given that the outcome variable was binary (course graduation/elimination), discriminant function analyses were used to develop prediction models. To facilitate implementation and interpretation, the following variables were excluded from discriminant analyses:

- TAPAS subtests not administered as part of AF TAPAS (Version 5)
- PAST subtests not administered for a corresponding AFS (i.e., timed swim scores for EOD, SERE, and TACP applicants)
- ASVAB composite scores (MAGE, AFQT, and in cases where PC or WK was used, Verbal Expression)
- Variables with missing data that would decrease listwise total sample size by 20% or more
- Variables with zero-order correlations where $p > .15$

Two-factor models using ASVAB and PAST variables were generated first, followed by three-factor models composed of ASVAB, PAST, and TAPAS. ASVAB and PAST variables used in the two- and three-factor analyses were generated using the two-factor datasets, which were larger than the three-factor datasets for all AFSs (see *N*s in Tables 5 and 6).

Significance levels for individual predictors were set at .15, slightly higher than the conventional standard of .05. This criterion was used to achieve an appropriate balance between the need to maximize prediction for the overall equation while retaining defensibility of the individual predictors.

Predicted probabilities derived from the discriminant analyses were used to determine classification accuracy at cut scores ranging from the 20^{th} to 80^{th} percentile, by decile. Predicted probabilities also were correlated with actual training outcomes (course graduation/elimination) to estimate criterion-related validity. These results were further corrected for dichotomization of the criterion (Cohen, 1983).

Cross-validation was conducted for each set of discriminant results using the U-method, based on the “leave-one-out” principle (Stone, 1974). In the U-Method, a discriminant function is constructed by taking a single observation out of the data set, and the function is used to classify the case left out. This process is repeated for each case in the dataset, thus reclassifying every data point as if it were a new unknown observation. This procedure provides a method for evaluating the stability of estimates based on the original samples.

Adverse impact potential of prediction models was evaluated using standardized mean differences, or Cohen’s *d* (Cohen, 1988) values. Only subgroups with sample sizes of 30 or more were included in these analyses.

4 Criterion-related Validity and Impact on Attrition Rates by AFS

For each AFS, Tables 5 and 6 present ratios of sample size to number of predictors tested, and criterion-related validities for the two- and three-factor models, respectively. Samples sizes for both the two- and three-factor models were large relative to the number of variables evaluated for each model, exceeding ratios (e.g., 20:1) generally considered best practice for discriminant analysis (Stevens, 2010). Further, prediction of training completion was statistically significant for all models and AFSs. Table 7 presents a direct comparison of R^2 for the two- and three-factor models evaluated, using common data. As shown, R^2 values improved between 7.9 and 63.4% for the two- versus three-factor models evaluated. Overall, the results provide evidence that the proposed models are generally likely to improve the qualification rates of applicants selected in each of the respective AFSs.

Table 5. *Summary of Predictive Validities*: *Two-Factor ASVAB/PAST Models*

Air Force Specialty (AFS)	***N***	**Ratio of *N*:Variables**	***R***	**R^2**
PJ	1,565	112:1	.384**	.148
CCT	867	79:1	.410**	.168
EOD	472	36:1	.459**	.210
SERE	706	59:1	.399**	.159
SOWT	223	45:1	.264*	.069
TACP	800	200:1	.270**	.073

**p*<.01, ***p*<.001

Table 6. *Summary of Predictive Validities*: *Three-Factor ASVAB/PAST/TAPAS Models*

Air Force Specialty (AFS)	***N***	**Ratio of *N*:Variables**	***R***	**R^2**
PJ	560	70:1	.497**	.247
CCT	332	47:1	.483**	.233
EOD	234	33:1	.461**	.213
SERE	241	34:1	.597**	.356
TACP	284	28:1	.487**	.237

Note. No analysis was conducted for three-factor SOWT model due to insufficient sample size.
**p*<.01, ***p*<.001

Table 7. *Comparison of Effect Size (R^2) for Two- versus Three-Factor Models Using Common Data*

Air Force Specialty (AFS)	***N***	**R^2: Two-Factor Model = ASVAB + PAST**	**R^2: Three-Factor Model = ASVAB + PAST + TAPAS**	**ΔR^2: Three- versus Two-Factor Model**
PJ	560	.168**	.247**	.079
CCT	332	.216**	.233**	.017
EOD	234	.185**	.213**	.028
SERE	241	.260**	.356**	.096
TACP	284	.145**	.237**	.092

Note. **p*<.01, ***p*<.001

Figures 1 through 6 show pass rates by quintile, for the proposed models by AFS. Pass rates were on average 41 percentile points higher for the highest versus lowest quintile across the AFSs evaluated. Results provide evidence that the proposed models are likely to reduce attrition and AETC training costs, through initial selection of better qualified candidates.

Figure 1. *PJ Pass Rate by ASVAB/PAST/TAPAS Quintile*

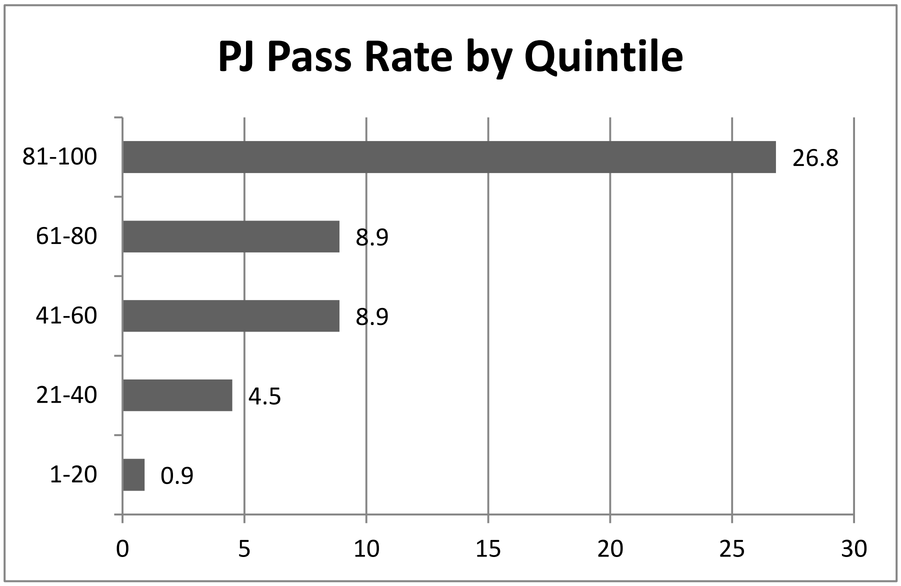


Figure 2. *CCT Pass Rate by ASVAB/PAST/TAPAS Quintile*

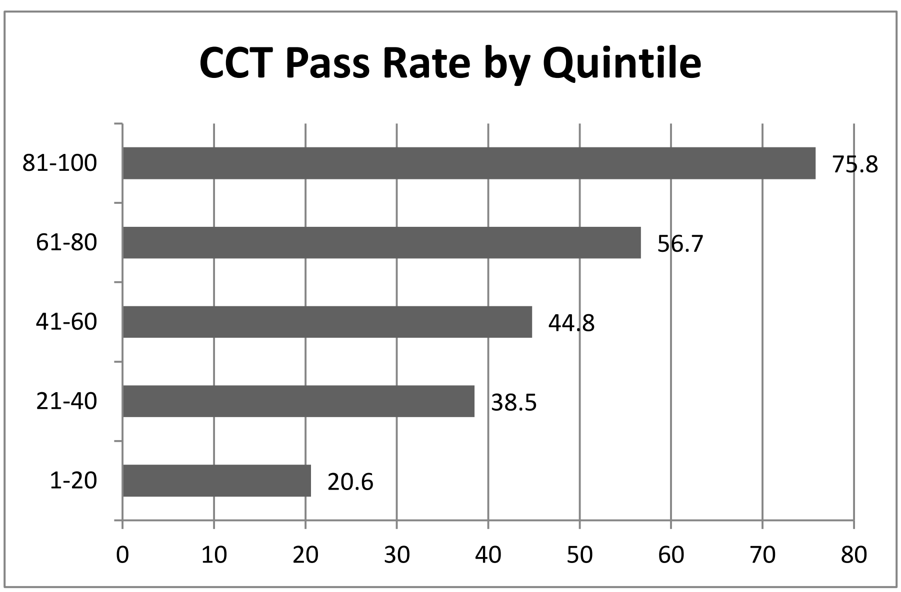


Figure 3. *EOD Pass Rate by ASVAB/PAST/TAPAS Quintile*

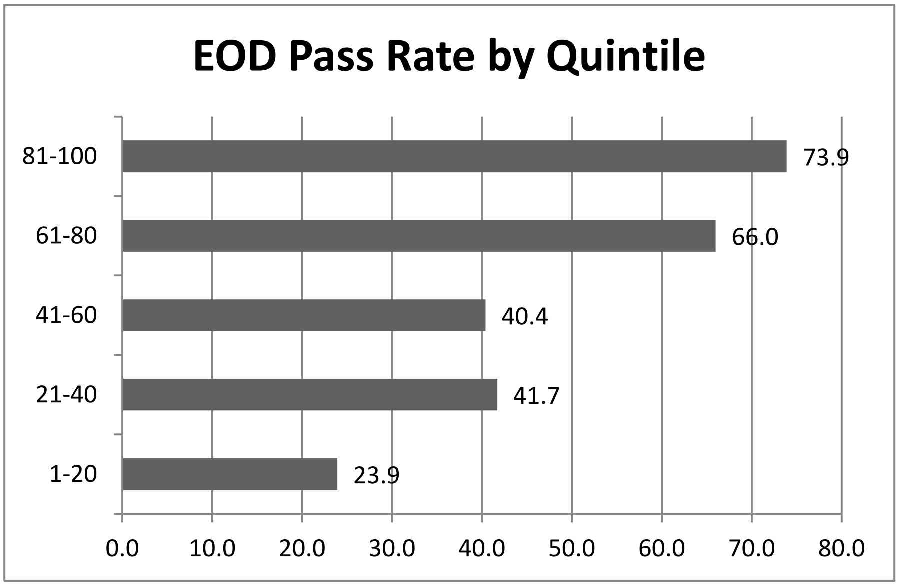
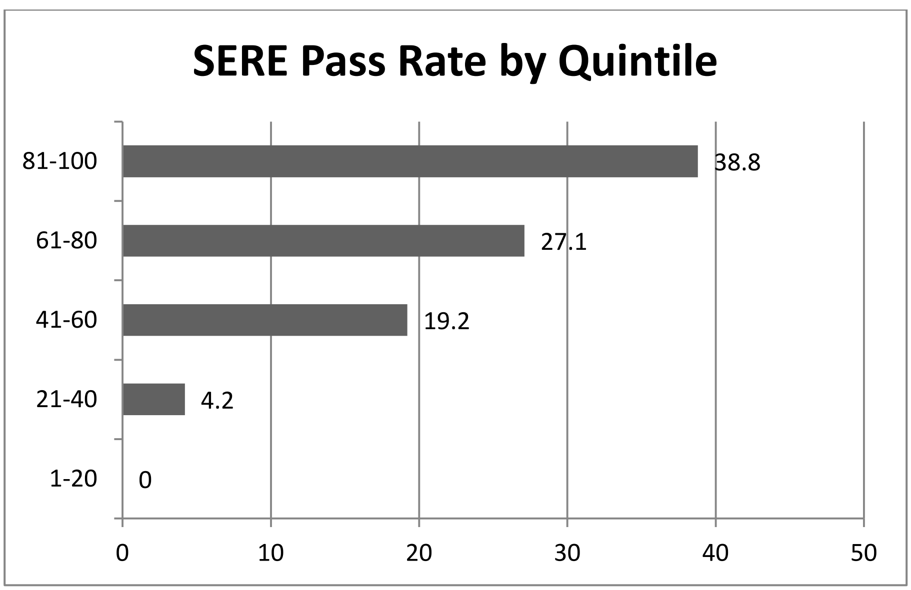


Figure 4. *SERE Pass Rate by ASVAB/PAST/TAPAS Quintile*



Note. No SERE candidates scoring in 1st through 19th percentile passed training.

Figure 5. *SOWT Pass Rate by ASVAB/PAST/TAPAS Quintile*

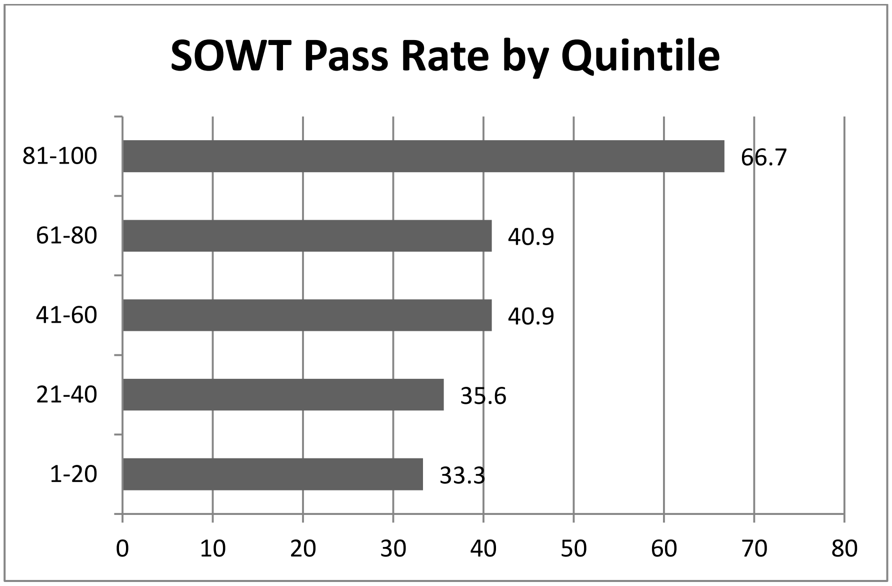
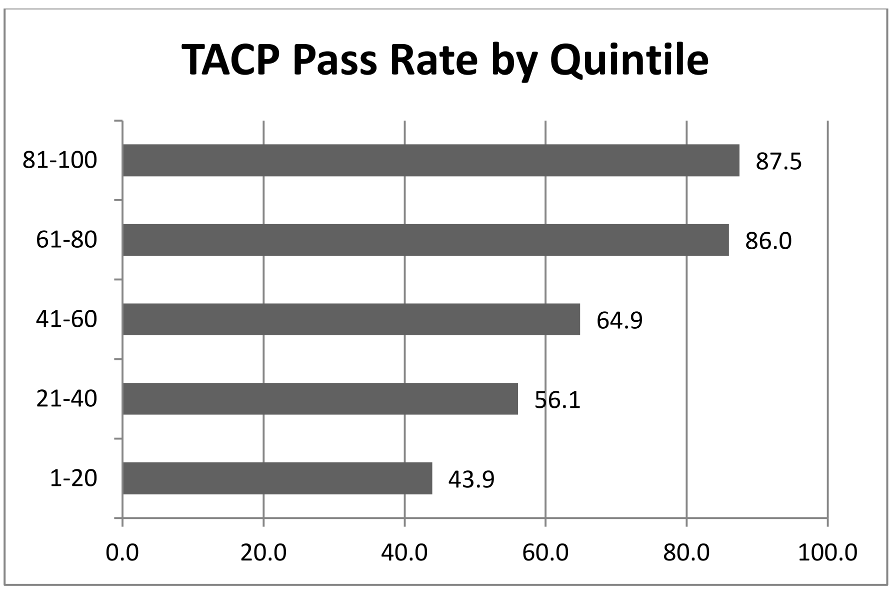


Figure 6. *TACP Pass Rate by ASVAB/PAST/TAPAS Quintile*



Appendices B and C present additional evidence supporting the proposed models, including validities by AFS based on cross-validation (Appendix B) and projected model impacts on attrition by AFS at selected cut scores (Appendix C).

5 AFPC/DSYX Recommendations for Implementation

To properly validate a selection or classification process for operational use by the Air Force, a test must go through five specific levels of validation known as the Selection and Classification Test Acquisition Process. This study focused on Level 4 validation, which is used to show that the test or measure can improve selection or classification (reduce or solve an identified problem or need) for the applicant population, and to develop formal models and weights for operational use. Prior AF activities since 2008 were conducted to validate TAPAS at Levels 1 (development of test or measure to meet mission need), 2 (concept exploration or proof of concept research using experimental or operational samples), and 3 (ensuring predictive validity and unique contribution of test relative to other measures) of the Selection and Classification Test Acquisition Process.

Next steps should focus on Level 5 Validation, which involves production, fielding/deployment, operational support, and ongoing monitoring. This level means the test or measure is now in operational use and personnel decisions can be made based upon the test or measure. Level 5 validation is a continuous process as long as the test or measure is used operationally to ensure external influences do not erode the effectiveness of the test. In line with the Level 5 Validation process, AFPC/DSYX recommends the following activities for implementation and ongoing monitoring and evaluation:

1. **Establish passing scores for qualification into each AFS**. Passing or cut scores should be based on a combination of AFS-specific student training requirements (STR), predicted classification accuracy, expected false negative rejection rates, the recruiting environment, and the costs of training, recruiting, and testing.
2. **Periodically (e.g., every six months) reassess criterion-related validity, adverse impact potential, and cut scores for all components of the selection system**. Studies similar to the current one should be used for this purpose. These studies also should

evaluate adverse impact for protected subgroups (e.g., race) in addition to those based on ethnicity, and consider adverse impact potential in setting of cut scores, provided sufficient subgroup sample sizes (*n*s > 30) are available.

6 REFERENCES

Allen, M.T., Cheng, Y.A., Putka, D.J., Hunter, A., & White, L. (2010). Analysis and findings. In D.J. Knapp & T.S. Heffner (Eds.). *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.

Campbell, J.P., & Knapp, D.J. (2010). Project A: 12 years of R&D. In J.L. Farr & N.T. Tippins (Eds.), *Handbook of employee selection* (pp. 865-886). New York: Routledge.

Chernyshenko, O.S., Stark, S., Drasgow, F., & Roberts, B.W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19*, 88-106.

Chernyshenko, O.S., & Stark, S., Prewett, M. S., Gray, A. A., Stilson, F. R., & Tuttle, M.D. (2009). Normative scoring of multidimensional pairwise preference personality scales using IRT: Empirical comparisons with other formats. *Human Performance, 22*, 105-127.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249–253.

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

Drasgow, F., Stark, S., Chernyshenko, O.S., Nye, C.D., Hulin, C.L., & White, L.A. (2012). *Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions* (Technical Report 1311). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43* (166), 38290–39315.

Fisk, M. (2013). *Occupational Analysis Report, AFSC 1T2X1 Pararescue, OSSN 2942*. Randolph AFB, TX: HQ AETC A2/3/10-OAA.

HQ AFPC/DPSIDC (2013, April). *Air Force enlisted classification directory (AFECD): The official guide to the Air Force enlisted classification codes*. Randolph AFB, TX: Author.

Knapp, D. J., & Heffner, T. S. (Eds.). (2009). *Predicting future force performance (Army Class): End of training longitudinal evaluation* (Technical Report 1257). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Heffner, T. S. (Eds.). (2010). *Expanded Enlistment Eligibility Metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Owens, K. S., & Allen, M. T. (Eds.). (2011). *Validating Future Force Performance Measures (Army Class): First In-Unit Performance Longitudinal Evaluation* (Technical Report 1293). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Manacapilli, T., Matthies, C.F., Miller, L.W., Howe, P., Perez, P.J., Hardison, C.M.,…Sims, C.S. (2012). *Reducing attrition in selected Air Force training pipelines*. Santa Monica, CA: Rand Corporation.

Nye, C.D., Drasgow, F., Chernyshenko, O.S., Stark, S., Kubisiak, U.C., White, L.A., & Jose, I. (2012). *Assessing the Tailored Adaptive Personality Assessment System as an MOS qualification instrument* (Technical Report 1312). Fort Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ree, M.J., Earles, J.A., & Teachout, M.S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology, 79*, 518-524.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general model for unfolding unidimensional polytomous responses using item response theory. *Applied Psychological Measurement, 24,* 2-32.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

Stark, S., Chernyshenko, O.S., Drasgow, F., & Williams, B.A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91*, 25-39.

Stevens, J.P. (2010). *Applied multivariate statistics for the social sciences*. New York, NY: Routledge.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, *36*, 111-147.

White, L.A., & Young, M.C. (1998, August). *Development and validation of the Assessment of Individual Motivation (AIM).* Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

United States Air Force (2006). *1C2X1: Combat control: Career field education and training plan.* Washington D.C.: Department of the Air Force.

United States Air Force (2008). *1T2X1: Pararescue specialty: Career field education and training plan.* Washington D.C.: Department of the Air Force.

United States Air Force (2009). *1C4X1: Tactical air control party: Career field education and training plan.* Washington D.C.: Department of the Air Force.

United States Air Force (2010a). *1T0X1: Survival, evasion, resistance, and escape specialist: Career field education and training plan.* Washington D.C.: Department of the Air Force.

United States Air Force (2010b). *1W0XX: Change 2 CFETP.* Washington D.C.: Department of the Air Force.

United States Air Force (2010c). *3E8X1: Explosive ordnance disposal: Career field education and training plan.* Washington D.C.: Department of the Air Force.

7 APPENDIX A: Uniform Guidelines Documentation of Impact and Validity Evidence

This tables cross-references information from the current study with requirements for documentation of impact and validity based on section 15 of the Uniform Guidelines on Employee Selection Procedures. Location of relevant information in the current report is identified by section, and where appropriate, further details are provided.

Uniform Guidelines Documentation Requirement (§ 1607.15)	Current Study
A2: *Information on Impact*	See Table 1 for Cohen's *d* values based on ethnicity; Section 4.2 describes need for ongoing evaluation of impact as sample sizes for additional protected subgroups increase.
B1: *User(s), location(s), and date(s) of study*	Section 2, *Description of Measures Evaluated*, describes the time frame (July 2008 through May 2013) for collection of data on selection procedures, and representative course titles and locations by AFS.
B2: *Problem and setting*	See Section 1, *Purpose and Overview* for definition of the purpose of the study. Section 2.2, *ASVAB*, and 2.3, *PAST*, describe existing selection procedures.
B3: *Job analysis or review of job information*	Various sources of job information were reviewed for each AFS including job descriptions (HQ AFPC, 2013), occupational analysis reports and briefings (e.g., Fisk, 2013), and career field education and training plans (USAF, 2006, 2008, 2009, 2010a, 2010b, 2011). Relevant technical reports (e.g., Manacapilli, et al., 2012) also were reviewed.
B4: *Job titles and codes*	See *Purpose and Overview* for list of Air Force Specialty Codes covered.
B5: *Criterion measures*	See section 2.4, *Course Graduation/Elimination*, for description of criterion measures.
B6: *Sample description*	Tables 5 and 6 present sample sizes by AFS. The majority of participants reported race as White (79.8%-93.2%), ethnicity as Non-Hispanic (75.2%-94.5%), and gender as Male (98.2%-100%).
B7: *Description of selection procedures*	See Section 2, *Description of Measures Evaluated*.
B8: *Techniques and results*	For a description of methods used in analyzing data, See Section 3, *Analyses*. For reports of results, see Section 4, *Criterion-related Validity and Impact on Attrition Rates by AFS*, Table 1, and Appendices B and C.
B9: *Alternative procedures investigated*	Criterion-related validity and adverse impact potential were compared for two- (ASVAB and PAST) versus three-factor (ASVAB, PAST, and TAPAS) models for five of six AFSs (e.g., see validity summaries in Tables 5 and 6). Validities were generally higher for three- versus two-factor models with approximately equal or less adverse impact

	potential. Previous studies focused on the pararescue career field also examined the viability of using an alternative non-cognitive test for predicting success in initial courses of entry. These studies found that TAPAS had greater validity than the alternative, although both non-cognitive tests reduced adverse impact potential for the total battery, which also included ASVAB and PAST components.
B10: ***Uses and applications***	Models will be used with cut scores for selection and classification. Evidence of the validity and utility of the procedure, as it is to be used (pre-accession), is provided in Tables 1, 5 -7, Figures 1 – 6, and Appendices B and C.
B11: ***Source Data***	Source data are being maintained in accordance with security requirements for facility storing of Federal data, as set forth in the Electronic Government Act Title III, also known as the Federal Information Security Management Act (FISMA).
B12: ***Contact person***	Title page includes name (HQ AFPC/DSYX) and mailing address of the organization to contact for additional information about this study.
B13: ***Accuracy and completeness***	Accuracy of data was ensured through examination of descriptive statistics, distribution shapes, and outliers for all variables, and appropriate recoding of values representing missing data. Complete analysis and reporting of results was ensured through close adherence to analysis and documentation principles established by the Uniform Guidelines on Employee Selection Procedures (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978), Principles for the Validation and Use of Personnel Selection Procedures (Society for Industrial and Organizational Psychology, 2003), and the Standards for Psychological and Educational Testing (American Educational Research Institution, American Psychological Association, & National Council on Measurement Education, 1999).

8 APPENDIX B: Cross-Validation Results By AFS

Air Force Specialty (AFS)	*R* Original	*R* Cross-Validation	Δ*R*
PJ (n=560)	.497**	.419**	-.078
CCT (n=332)	.483**	.429**	-.054
EOD (n=234)	.461**	.368**	-.093
SERE (n=241)	.597**	.482**	-.115
SOWT[A] (n=224)	.264*	.180*	-.084
TACP (n=284)	.487**	.401**	-.086

Note. [A]Based on two-factor ASVAB/PAST model only; *$p<.01$ **$p<.001$

9 APPENDIX C: Reductions in Attrition Rates at Selected Cut Scores By AFS

Air Force Specialty (AFS)	Sample Attrition Rate	Set Percentile Cut	Model Attrition Rate	∆Attrition Rate	Model False Reject Rate
PJ	90.0%	60th%ile	82.1%	7.9%	2.9%
CCT	53.0%	40th%ile	41.1%	11.9%	11.4%
EOD	50.9%	30th%ile	42.4%	8.5%	9.0%
SERE	82.2%	50th%ile	67.7%	14.5%	1.7%
SOWT	56.5%	30th%ile	50.0%	6.5%	8.5%
TACP	32.4%	30th%ile	24.3%	8.1%	14.8%

Note. **p*<.01 ***p*<.001