AFRL-RH-WP-TR-2019-0090

THE DEVELOPMENT OF THREE STATIC PERSONALITY RESEARCH FORMS AND ON-LINE SCORING TOOLS FOR THE U.S. AIR FORCE

Oleksandr S. Chernyshenko
Bo Zhang
Fritz Drasgow
Stephen Stark
Christopher D. Nye

Drasgow Consulting Group

November 2019

Interim Report

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
AIRMAN SYSTEMS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2019-0090 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//
THOMAS R. CARRETTA
Work Unit Manager
Collaborative Interfaces and Teaming Branch
Warfighter Interfaces Division

//signature//
TIMOTHY S. WEBB
Chief, Collaborative Interfaces and
Teaming Branch
Warfighter Interfaces Division

//signature//
LOUISE A. CARTER
Chief, Warfighter Interfaces Division
Airman Systems Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
06-11-19	Interim	17 JAN 18 – 1 Nov 19

4. TITLE AND SUBTITLE

The Development of Three Static Personality Research Forms and On-Line Scoring Tools for the U.S. Air Force

5a. CONTRACT NUMBER
FA8650-14-D-6500, Task Order 0007

5b. GRANT NUMBER
Not applicable

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Oleksandr S. Chernyshenko
Bo Zhang
Fritz Drasgow
Stephen Stark
Christopher D. Nye

5d. PROJECT NUMBER
5329

5e. TASK NUMBER

5f. WORK UNIT NUMBER
H0SA (532909TC)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Drasgow Consulting Group (DCG)
3508 Highcross Rd.
Urbana, IL, 6180

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Airman Systems Directorate
Warfighter Interface Division
Collaborative Interfaces and Teaming Branch
Wright-Patterson AFB, OH 45433

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
711 HPW/RHCC

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RH-WP-TR-2019-0090

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited; 88 ABW-2019-5919; Cleared 08 Jan 2020.

13. SUPPLEMENTARY NOTES
Subcontract number FPH02-S021; PO number 180169

14. ABSTRACT
Three versions of the Air Force (AF) Tailored Adaptive Personality Assessment System (TAPAS) test were developed. All assess the same 15 facets underlying the Big Five personality dimensions. The versions include a traditional single statement format with a 4-point Likert response scale, a two-alternative forced choice response format where the pairs of statements are unidimensional, and a two-alternative forced choice format with multidimensional paired statements. Software for scoring all three response formats was also developed. The three forms were administered to a sample of USAF Basic Recruits to assess their psychometric characteristics. Reasonably good cross-form correlations were found, with the multidimensional format the most resistant to faking good. The Likert scale version had the highest correlations with several self-report criterion variables.

15. SUBJECT TERMS
Multidimensional pairwise preference (MDPP) model; ideal point model; Likert scale; personality assessment; Big Five model.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Thomas R. Carretta
Unclassified	Unclassified	Unclassified	SAR	33	19b. TELEPHONE NUMBER *(Include Area Code)* (937) 713-7143

THIS PAGE IS INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF TABLES

Page

PREFACE

The work described in this technical report was performed under the task "Tailored Adaptive Personality Assessment System Item Development," Contract FA8650-14-D-6500, Task Order 7, Enhanced Airman Alignment, Work Unit H0SA (532909TC). We thank John Trent and the Air Force Personnel Center, Strategic Research and Assessment Branch (AFPC/DSYX) for data collection at Lackland AFB, TX.

SUMMARY

The U.S Air Force (USAF) requires static forms of personality inventories for personnel studies. Accordingly, the present study developed three forms that assess the same 15 facets of personality using statements from the Tailored Adaptive Personality Assessment System (TAPAS) statement pool. Form A contains 90 items where each statement is presented individually, and respondents are asked to indicate agreement on a 4-point Likert response scale. Form B and Form C were both 120-item two-alternative forced-choice scales where respondents are presented with a pair of statements and asked to choose the statement that is "more like me." In form B, each item contains two personality statements representing the same personality dimension but differing in extremity. In form C, items are comprised of two personality statements that are similar in extremity and desirability, but represent different personality dimensions. Software to provide scoring for each format was also developed. The forms were administered to a sample of USAF Basic Recruits to assess their psychometric characteristics. The cross-form correlations of facets were reasonably large and the multidimensional forced-choice form was most resistant to faking. The Likert format form had the largest correlations with several self-report criterion measures.

1.0 INTRODUCTION

Interest in temperament/personality as a predictor of performance has increased dramatically over the past two decades. Much of this interest was galvanized by empirical evidence showing that temperament constructs predict performance across a diverse array of civilian and military occupations (e.g., Barrick & Mount, 1991; Campbell & Knapp, 2001) and provide incremental validity beyond general cognitive ability (Schmidt & Hunter, 1998). Today, there is little doubt that personality traits, defined either as broad factors (e.g., Big Five), facets (e.g., Achievement or Dominance), or compound traits (e.g., Self-Efficacy), predict a multitude of job outcomes including task performance (Tett, Steele, & Beauregard, 2003; Judge et al., 2007), organizational citizenship behaviors (Chiaburu et al, 2011), counter-productivity (Grijalva & Newman, 2015; Van Iddenkinge et al., 2012;), leadership (Judge et al., 2002), and adaptive performance (Huang et al., 2014).

Military researchers have been at the forefront of this personality assessment renaissance. In particular, the U.S. military had funded the development and validation research for a number of inventories including the Trait Self-Description Inventory (TSDI) (Christal, Barucky, Driskill, & Collins,1997), the Assessment of Background and Life Experiences (ABLE) Questionnaire (White, Nord, Mael, & Young, 1993), the Navy Computer Adaptive Personality Scales (NCAPS) (Houston et al., 2005), the Army's Assessment of Individual Motivation (AIM) (White & Young, 1998), and the TAPAS (Drasgow et al., 2012). As an example, the TSDI that was originally developed by the U.S. Air Force has also been used by British, Australian, and Canadian militaries to predict performance in basic training and well as in officer cadet schools (e.g., Allender, 2005). In 2014, back-to-back special issues of the *Military Psychology* journal published several papers that summarized extant research and argued for incorporating personality scores when making personnel selection and classification decisions.

The use of temperament variables for selection and classification brings attention to the quality of their measurement. Besides identifying key personality traits to measure, a large body of recent research has focused on ways to mitigate various response distortions and biases commonly associated with self-reports (e.g., White & Young, 1998; Brown & Maydeu-Olivares, 2013). In particular, forced-choice item formats, which ask test takers to choose among two or more statements rather than to provide Likert ratings of individual statements, have been found to be reasonably effective in dealing with response distortion problems (e.g., Brown, Inceolglu, & Lin, 2017; Cao & Drasgow 2019; Christiansen, Burns, & Montgomery, 2005; Martin, Bowen, & Hunt, 2002; Pavlov, Maydeu-Olivares, & Fairchild, 2018). The added resistance to response biases comes at a price, however, as forced-choice measures may take longer to answer, are more cognitively demanding, and require more advanced psychometrics models for scoring. For these reasons, many personality testing programs are now developing their measures in multiple formats so they can customize their products to meet the diverse needs of their clients.

2.0 METHODS, ASSUMPTONS, AND PROCEDURES

This section of the report briefly describes the development of three personality research forms to be used by the Air Force in their personnel studies. The three forms were designed to measure the same 15 narrow personality dimensions (facets), but differ in terms of the item response format. Each form is based on the TAPAS personality statement pool that had been developed by the Drasgow Consulting Group in 2007 and pretested using U.S. Army recruits in 2007-2008. The unique feature of the TAPAS statement pool is that it was developed under the ideal point response process assumptions, so, in addition to commonly used positive or negative statements, it also contains personality statements representing *moderate/neutral standings* on a trait continuum (for more detail see Chernyshenko et al., 2007).

3.0 FORMS AND SCORING

3.1 Three Forms of Personality Inventory

3.1.1 Form A: 90-item Likert Form

Form A utilizes the traditional, single statement response format, where each statement is presented individually, and respondents are asked to indicate agreement on a 4-point Likert scale (strongly disagree, disagree, agree, strongly agree). The Likert format has been used widely by personality researchers to specify the hierarchical structure of personality traits and to estimate criterion related validities in educational, health, and employment contexts. Example personality measures utilizing the Likert format include the Neuroticism-Extraversion-Openness Personality Inventory-Revised (NEO-PI-R) (Costa & McCrae, 2008), TSDI, and ABLE.

Form A consists of 90 personality single-statement items that are distributed evenly across the 15 personality facets (see Table 1 below). To create 6-item scales for each facet, statements having large discrimination parameters and describing either positive or negative standings on the trait continuum (a.k.a., positively/negatively worded) were selected from the TAPAS research statement pool. Negatively worded statements were included because they help to combat the acquiescence bias typically associated with the Likert format (i.e., a tendency to agree with all items in a scale regardless of their content). All statements were then randomly ordered and a standard set of instructions was added.

Table 1. Personality Facets Assessed by the Three Air Force Forms

Personality Facet Name	Brief Description
Achievement	High scoring individuals are seen as hard working, ambitious, confident, and resourceful.
Adjustment	High scoring individuals are well adjusted, worry free, and handle stress well.
Attention Seeking	High scoring individuals tend to engage in behaviors that attract social attention. They are loud, loquacious, entertaining, and even boastful.
Cooperation	High scoring individuals are pleasant, trusting, cordial, non-critical, and easy to get along with.
Dominance	High scoring individuals are domineering, “take charge,” and are often referred to by their peers as "natural leaders."

Even-Tempered	High scoring individuals tend to be calm and stable. They don't often exhibit anger, hostility, or aggression.
Intellectual Efficiency	High scoring individuals believe they process information and make decisions quickly; they see themselves (and they may be perceived by others) as knowledgeable, astute, or intellectual.
Non-Delinquency	High scoring individuals tend to comply with rules, customs, norms, and expectations, and they tend not to challenge authority.
Optimism	High scoring individuals have a positive outlook on life and tend to experience joy and a sense of well-being.
Order	High scoring individuals tend to organize tasks and activities and desire to maintain neat and clean surroundings.
Physical Condition	High scoring individuals tend to engage in activities to maintain their physical fitness and are more likely to participate in vigorous sports or exercise.
Self-Control	High scoring individuals tend to be cautious, levelheaded, able to delay gratification, and patient.
Selflessness	High scoring individuals are generous with their time and resources.
Sociability	High scoring individuals tend to seek out and initiate social interactions.
Tolerance	High scoring individuals are interested in other cultures and opinions that may differ from their own.

3.1.2 Form B: 120-item Unidimensional Pairwise Preference (UPP) Form

Form B utilizes the UPP item format. In this format, each item contains two personality statements representing the same personality dimension but differing in extremity. Respondents are asked to choose the statement in each pair that is "more like me." An example of an inventory utilizing the UPP format is the NCAPS.

Form B consists of 120 UPP items with 8 items measuring each of the 15 personality facets. UPP items were constructed so that the two statements are at least two units apart (see below) in extremity. Pairing statements too close to each other negatively impacts item discrimination, which, in turn, reduces scales reliability. Extremity parameters were derived by rescaling existing TAPAS statement pool location parameters to fit into a -3 to +3 standard normal metric with negatively worded statements receiving extremity ratings in the -3 to -1 range, neutrally worded statements receiving extremity parameter values in the -1 to +1 range, and positively worded statements receiving values in the +1 to +3 range.

3.1.3 Form C: 120-item Multidimensional Pairwise Preference (MDPP) Form

Form C utilizes the MDPP item format. In this format, items ware comprised of two personality statements that are similar in extremity and desirability, but represent different personality dimensions. A small number of unidimensional pairs were also added to facilitate scoring accuracy and to improve examinee reactions. Specifically, there were a total of 12 unidimensional pairs (10% of the test length), one pair per dimensions except Achievement, Physical Conditioning, and Self Control; the latter three dimensions had no unidimensional pairs. Similar to the UPP form, respondents are asked to choose one statement in each pair that is "more like me." An example of an inventory utilizing the MDPP format is TAPAS.

Form C consists of 120 pairwise preference items with most items being multidimensional. Each personality facet is assessed by 12-16 statements that were selected from the TAPAS statement pool. MDPP items were constructed by matching statements based on extremity (statement location) and desirability. UPP items were constructed so that the two statements were at least 2 units apart in terms of their locations. A statement could be repeated once, but had to be paired with a different statement.

3.2 Scoring Procedures for the Three Personality Research Forms

The three personality research forms are designed for paper-and-pencil test administration. To score each form, an on-line scoring tool was developed. This was necessary because pairwise preference forms (Form B and Form C) need to be scored using item response theory (IRT) methods containing complex mathematical routines. To obtain scores for each of the three forms, examinee item response data must be submitted in an excel format (.cvs); resulting test scores are also outputted in this excel format. We briefly describe each scoring routine below.

3.2.1 Scoring Form A: Likert Form

The scoring routine for *Form A* is straightforward and utilizes the classical test theory approach. Each examinee's item responses must be coded as A = "strongly disagree", B = "disagree", C = "agree", and D = "strongly agree"; missing responses must be left as blanks. After receiving examinee responses, the scoring routine first reverse scores negatively worded items and only then recodes ABCD letters into 1234 integers. Then, for each personality scale, the routine computes the average score across all endorsed items belonging to that scale; if some of the six items are not endorsed, they do not affect the computation of that scale average. Finally, the average is multiplied by 6 to produce the final personality facet score and the scores are outputted in the .csv format.

3.2.2 Scoring Form B: UPP Form

The scoring ro utine for *Form B* is based on IRT methodology and utilizes the posteriori (Expected A Posteriori (EAP)) estimation method to derive scores for the 15 personality facets. The EAP estimate of theta (Bock & Mislevy, 1982) is a Bayesian estimator derived by finding the mean of the posterior trait distribution given the item parameters and responses for items comprising a particular personality scale (coded 0,1). The posterior distribution is computed as the conditional probability of the response pattern multiplied by a prior distribution function (normal with a mean of 0 and a variance of 1).

EAP estimation proceeds as follows. First, the latent trait continuum is divided into 80 equally spaced discrete points called quadrature nodes (Q_r) on the interval [-3, +3]. Next, the item parameters and examinee responses are used to compute the conditional likelihood of a response pattern, $L(Q_r)$, as shown:

$$L(\mathbf{u}|Q_r,\beta_1,...,\beta_n)=\prod_i P_i(Q_r)^{u_i}[1-P_i(Q_r)]^{1-u_i} \quad , \tag{1}$$

where $\mathbf{u}$= <u1, u2, … un> is a vector of item responses, $\beta_i = <\mu_s,\mu_t>$ is a vector of item parameters, and P_i is the probability of preferring statement *s* to statement *t* in item *i* computed as follows:

$$\mathrm{P_{st}}(\theta)=1-\Phi(\mathrm{a_{st}})-\Phi(\mathrm{b_{st}})+2\Phi(\mathrm{a_{st}})\Phi(\mathrm{b_{st}}), \text{ where} \tag{2}$$

$$\mathrm{a_{st}}=(2\theta-\mu_s-\mu_t)/\sqrt{3}, \tag{3}$$

$$\mathrm{b_{st}}=\mu_s-\mu_t, \text{ where} \tag{4}$$

θ represents the respondent's ideal point, μ_s and μ_t represent the locations of the respective statements on the trait continuum, and $\Phi(\mathrm{a_{st}})$ and $\Phi(\mathrm{b_{st}})$ are cumulative standard normal density functions evaluated at $\mathrm{a_{st}}$ and $\mathrm{b_{st}}$, respectively.

The conditional likelihood at each node is then multiplied by weights $W(Q_r)$ corresponding to the prior distribution, and the products are summed to obtain the marginal distribution (see the denominator of Equation 5). The EAP estimator of theta is then computed as:

$$\theta_{EAP}=\frac{\sum_{r=1}^{80} Q_r * L(Q_r) * W(Q_r)}{\sum_{r=1}^{80} L(Q_r) * W(Q_r)} \quad . \tag{5}$$

The EAP estimate of theta represents an examinee's score on a particular personality construct and, because Form B assesses 15 personality constructs, the scoring routine computes 15 EAP scores for each 120-item response pattern.

3.2.3 Scoring Form C: MDPP Form

Form C scoring also uses IRT theory methodology, but the underlying model and scoring approach differs from Form B. The IRT model is the MDPP (Stark, Chernyshenko, & Drasgow, 2005) which specifies the probability of endorsing statement *s* over a statement *t* as

$$P_{(s>t)_i}(\theta_{d_s},\theta_{d_t})=\frac{P_{st}\{1,0\}}{P_{st}\{1,0\}+P_{st}\{0,1\}}\approx\frac{P_s\{1\}P_t\{0\}}{P_s\{1\}P_t\{0\}+P_s\{0\}P_t\{1\}},$$

where:

i = index for items, consisting of pairs of statements, where *i* = 1 to I,

$d =$ index for dimensions, where d = 1, …, D,
$s, t =$ indices for the first and second statements, respectively, in an item,
$\theta_{d_s}, \theta_{d_t} =$ latent trait values for a respondent on dimensions d_s and d_t respectively,
$P_s\{1\}, P_s\{0\} =$ probability of endorsing/not endorsing statement s at θ_{d_s},
$P_t\{1\}, P_t\{0\} =$ probability of endorsing/not endorsing statement t at θ_{d_t},
$P_{st}\{1,0\} =$ joint probability of endorsing statement s, and not endorsing statement t at $(\theta_{d_s}, \theta_{d_t})$,
$P_{st}\{0,1\} =$ joint probability of not endorsing statement s, and endorsing statement t at $(\theta_{d_s}, \theta_{d_t})$,
and
$P_{(s>t)_i}(\theta_{d_s}, \theta_{d_t}) =$ probability of respondent j preferring statement s to statement t in pairwise preference item i.

Note that the probabilities of endorsing/not endorsing a stimulus in a pairwise preference item is computed using the Generalized Graded Unfolding Model (GGUM); (Roberts, Donoghue, & Laughlin, 2000); GGUM parameters for the TAPAS research pool were estimated using samples of U.S. military recruits undergoing their basic training.

The scoring of Form C response patterns is accomplished via the Bayes modal estimation approach. For a vector of latent trait scores, $\tilde{\theta} = (\theta_{d'=1}, \theta_{d'=2}, \ldots, \theta_{d'=D})$, $\tilde{\boldsymbol{\theta}} = (\theta_{d'=1}, \theta_{d'=2}, ..., \theta_{d'=D})$, this involves maximizing:

$$L(\tilde{\mathbf{u}}, \tilde{\boldsymbol{\theta}}) = \{\prod_{i=1}^{n} [P_{(s>t)_i}]^{u_i} [1 - P_{(s>t)_i}]^{1-u_i}\} * f(\tilde{\boldsymbol{\theta}}),$$

where $\tilde{\mathbf{u}}$ represents an examinee's item response pattern, u_i is the dichotomous response to item i, $P_{(s>t)_i}$ is the probability of preferring statement s to statement t in item i, and $f(\tilde{\boldsymbol{\theta}})$ is a D-dimensional prior density function, which, for simplicity, is assumed to be the product of independent normals,

$$\prod_{d'=1}^{D} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\theta_{d'}^2}{2\sigma^2}\right).$$

Taking the natural log, for convenience, the above equation can be rewritten as:

$$\ln L(\tilde{\mathbf{u}}, \tilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} [u_i \ln P_{(s>t)_i} + (1 - u_i) \ln(1 - P_{(s>t)_i})] + \sum_{d'=1}^{D} \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\theta_{d'}^2}{2\sigma^2} \right] \quad ,$$

leaving the following set of equations to be solved numerically:

$$\frac{\partial \ln L}{\partial \tilde{\boldsymbol{\theta}}} = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta_{d'=1}} \\ \frac{\partial \ln L}{\partial \theta_{d'=2}} \\ \ldots \\ \frac{\partial \ln L}{\partial \theta_{d'=D}} \end{bmatrix} = 0.$$

In total, for each 120-item response pattern submitted, the scoring routine outputs 15 Bayes modal estimates (one per personality construct). Similar to Form A and Form B, the scores are saved in the .csv format.

4.0 METHOD

To evaluate the psychometric characteristics of the instruments, the three forms were administered to a sample of USAF Basic Recruits. The forms were administered in two orders: (1) Likert, multidimensional forced-choice, and unidimensional forced-choice; and (2) unidimensional forced-choice, Likert, and multidimensional forced choice. The Basic Recruits were first asked to answer honest for research purposes, and then complete the forms a second time where they were asked to do their best to "convince the Air Force that you would make a good Airman" (i.e., fake good).

The Likert format scales were scored by computing the mean response. Item response theory scoring, described in the preceding section, was used for the forced-choice scales.

Five self-report scales were also included to serve as criterion variables. They utilized a five-point Likert rating format. Situational Decision-Making was assessed by 11 items. Its coefficient alpha reliability was .83 for the data collected in the Honest condition. Communication was assessed by six items taken from Lentz et al. (2009) and had a reliability of .81 in the present study. Decision-Making and Management was assessed by eight items taken from Lentz (2009). In this study, its reliability was .84. Leading Others was assessed by five items from Lentz (2009) with a reliability of .77. Displaying Professionalism was assessed by six items, also from Lentz (2009), with a reliability of .78.

5.0 RESULTS

5.1 Data Cleaning

Data were obtained from 148 respondents in the honest condition and 150 in the faking condition in administration order 1 and 201 respondents in the honest condition and 190 in the faking condition for administration order 2.

5.2 Order Effects

To examine whether the order of administration of the forms had a material effect, scores for the "respond honestly" were compared. Across 45 comparisons (three formats, each with 15 facets), 13 mean differences were small (Cohen's *d* between .2 and .39), five mean differences were moderate (Cohen's *d* between .4 and .65), and none were large (Cohen's *d* greater than .65); the remaining differences were smaller than "small" (Cohen's *d* of .2). As the overall impact of order does not appear substantial, the data from the two forms was merged and subsequent analyses were conducted on the entire sample.

5.3 Descriptive Statistics

Table 2 presents the descriptive statistics for the 15 facets from the Likert format scales. Scores were computed as the mean response and, in this table, reliability is coefficient alpha. Reliabilities in the Honest condition are all acceptable, ranging from .70 for Self-Control to .85 for Dominance, with a mean of .77.

Note that means in the Faking condition are elevated from the Honest condition. In fact, all of the means, except for Attention Seeking, are significantly higher. The mean effect size (Cohen's *d*) was .57, which is consistent with Viswesvaran and Ones's (1999) meta-analysis of instructed faking studies.

Table 2. Descriptive Statistics for the Single Statement Likert AF TAPAS

Variables	Data Merged from Two Admistration Orders								
	Honest (n=349)			Faking (n=340)			t	p	d
	M	SD	reliability	M	SD	reliability			
SS_Achievement	3.34	0.46	0.84	3.61	0.45	0.80	6.22	0.00	0.58
SS_Adjustment	2.75	0.54	0.77	3.12	0.52	0.72	7.49	0.00	0.70
SS_Attention Seeking	2.28	0.63	0.85	2.34	0.57	0.73	1.07	0.28	0.10
SS_Cooperation	3.24	0.43	0.74	3.45	0.43	0.71	5.12	0.00	0.48
SS_Dominance	2.77	0.58	0.85	3.21	0.58	0.82	8.03	0.00	0.75
SS_Even-Tempered	3.10	0.51	0.76	3.37	0.50	0.72	5.85	0.00	0.55
SS_Intellectual Efficiency	2.94	0.49	0.72	3.30	0.54	0.79	7.52	0.00	0.71
SS_Non-Delinquency	3.04	0.48	0.77	3.28	0.50	0.74	5.46	0.00	0.51
SS_Optimism	3.16	0.51	0.77	3.48	0.48	0.78	6.91	0.00	0.64
SS_Order	3.00	0.52	0.74	3.18	0.48	0.61	4.01	0.00	0.37
SS_Physical Condition	2.98	0.59	0.81	3.31	0.55	0.78	6.30	0.00	0.59
SS_SelfControl	3.04	0.44	0.70	3.41	0.49	0.78	8.53	0.00	0.80
SS_Selflessness	3.06	0.45	0.71	3.36	0.48	0.70	6.86	0.00	0.64
SS_Sociability	2.63	0.62	0.80	3.07	0.57	0.77	7.96	0.00	0.74
SS_Tolerance	3.04	0.51	0.75	3.27	0.50	0.71	4.73	0.00	0.44

Coefficient alpha cannot be computed for the unidimensional and multidimensional forced-choice items, so IRT marginal reliability was used as the estimate of reliability,

$$\text{Marginal reliability} = 1 - \frac{\bar{\sigma}_e^2}{\sigma_{\hat{\theta}}^2},$$

where $\bar{\sigma}_e^2$ is the average squared standard error of the latent trait estimate $\hat{\theta}$ and $\sigma_{\hat{\theta}}^2$ is the variance of the $\hat{\theta}$ values.

Table 3 provides summary statistics for the unidimensional force-choice scales. The means are the mean $\hat{\theta}$ value. The item parameters used for estimation originate from when the statements were originally calibrated (several years ago) using data from samples of Army recruits. Because the scale of a latent trait is arbitrary, item parameter estimation proceeded with the assumption that the latent trait distributions were standard normal. In sum, the means of approximately zero in Table 3 are about what should be expected. The standard deviations are noticeably less than

unity because the $\hat{\theta}s$ were estimated by Bayesian methods and therefore tend to be pulled back toward the mean (zero) of the prior.

With forced-choice instruments, single-subject response consistency error inflates reliability less than for single statement assessments, so the lower reliabilities seen in this table are not too surprising. In any case, IRT marginal reliabilities ranged from a disappointing .34 for Cooperation to .73 for dominance, with a mean of .56.

The results for fakability are largely consistent with the single statement results, with all but Attention Seeking showing statistically significant score increases. The mean effect size was .49, just slightly lower than .57 for the Likert format scales.

Table 3. Descriptive Statistics for the Unidimensional Forced-Choice AF TAPAS

Variables	Data Merged from Two Admistration Orders								
	Honest (n=349)			Faking (n=340)			t	p	d
	M	SD	reliability	M	SD	reliability			
UFC_Achievement	0.75	0.65	0.47	1.08	0.52	0.08	5.98	0.00	0.56
UFC_Adjustment	0.86	0.86	0.71	1.34	0.59	0.46	6.98	0.00	0.68
UFC_Attention Seeking	-0.68	0.68	0.55	-0.81	0.62	0.45	-2.10	0.04	-0.20
UFC_Cooperation	0.94	0.57	0.34	1.18	0.50	0.06	4.84	0.00	0.45
UFC_Dominance	0.23	0.92	0.73	0.77	0.70	0.46	7.15	0.00	0.68
UFC_Even-Tempered	0.81	0.68	0.52	1.13	0.56	0.20	5.55	0.00	0.52
UFC_Intellectual Efficiency	0.73	0.69	0.55	0.98	0.64	0.44	4.02	0.00	0.37
UFC_Non-Delinquency	0.34	0.53	0.39	0.63	0.52	0.33	5.99	0.00	0.56
UFC_Optimism	0.77	0.74	0.58	1.21	0.55	0.19	7.16	0.00	0.68
UFC_Order	0.85	0.78	0.66	1.26	0.63	0.47	6.33	0.00	0.60
UFC_Physical Condition	0.86	0.77	0.65	1.24	0.59	0.36	5.98	0.00	0.57
UFC_SelfControl	0.84	0.71	0.53	1.07	0.65	0.41	3.66	0.00	0.34
UFC_Selflessness	0.95	0.64	0.40	1.19	0.53	0.07	4.44	0.00	0.42
UFC_Sociability	0.26	0.88	0.72	0.77	0.67	0.51	7.10	0.00	0.68
UFC_Tolerance	0.99	0.66	0.52	1.27	0.53	0.21	5.09	0.00	0.48

Table 4 presents the results for the multidimensional forced-choice scales. Again, as expected, the means are approximately zero with standard deviations of scale scores noticeably less than one. IRT marginal reliabilities ranged from .33 for Selflessness to .79 for Physical Conditioning, with a mean of .62.

Only 7 of the 15 scales showed statistically significant score differences between the Honest and Faking conditions. The mean effect size was d = .20, which is substantially lower than that observed for the Likert and unidimensional forced-choice scales.

Table 4. Descriptive Statistics for the Multidimensional Forced-Choice AF TAPAS

Variables	Data Merged from Two Admistration Orders								
	Honest (n=349)			Faking (n=340)			t	p	d
	M	SD	reliability	M	SD	reliability			
MFC_Achievement	0.13	0.52	0.61	0.46	0.56	0.67	6.64	0.00	0.62
MFC_Adjustment	-0.05	0.60	0.60	0.08	0.54	0.52	2.30	0.02	0.21
MFC_Attention Seeking	-0.31	0.51	0.65	-0.17	0.44	0.54	3.14	0.00	0.29
MFC_Cooperation	-0.01	0.45	0.59	0.01	0.41	0.54	0.50	0.61	0.05
MFC_Dominance	-0.34	0.61	0.74	-0.08	0.52	0.65	4.97	0.00	0.46
MFC_Even-Tempered	0.23	0.44	0.53	0.30	0.40	0.45	1.77	0.08	0.16
MFC_Intellectual Efficienc	-0.22	0.57	0.73	-0.19	0.47	0.60	0.60	0.55	0.06
MFC_Non-Delinquency	0.18	0.47	0.50	0.26	0.40	0.32	1.81	0.07	0.17
MFC_Optimism	0.24	0.54	0.65	0.30	0.41	0.42	1.31	0.19	0.12
MFC_Order	-0.11	0.55	0.71	-0.08	0.43	0.57	0.72	0.47	0.07
MFC_Physical Condition	-0.03	0.65	0.79	-0.11	0.50	0.67	-1.45	0.15	-0.14
MFC_SelfControl	0.11	0.52	0.45	0.25	0.51	0.43	3.06	0.00	0.28
MFC_Selflessness	-0.10	0.43	0.33	0.04	0.39	0.17	3.74	0.00	0.35
MFC_Sociability	-0.35	0.64	0.69	-0.35	0.48	0.45	0.13	0.90	0.01
MFC_Tolerance	0.05	0.64	0.74	0.20	0.47	0.52	2.85	0.00	0.27

5.4 Cross-Format Correlations

Table 5 presents the convergent validity correlations for responses obtained in the Honest condition. For example, the SS-MFC correlation for Achievement is the correlation between the single statement Likert Achievement scale and the multidimensional forced-choice Achievement scale. This correlation was .55, but rose to .76 when disattenuated for measurement error. In Table 5, the correlations of the multidimensional forced-choice scales with either the Likert or unidimensional forced-choice scales tend to be somewhat lower than the Likert-unidimensional correlations.

Table 5 also shows the correlations after disattenuating for measurement error. Interestingly, the single statement Likert-unidimensional forced-choice corrected correlations are all nearly perfect, with an average of .98. The multidimensional forced-choice-unidimensional forced-choice corrected correlations are also very large, with an average of .88. The single statement

Likert-multidimensional forced-choice corrected correlations are somewhat lower, with an average of .75. These results might be understood as a result of there being two differences between the Likert scales and the multidimensional forced-choice scales: the forced-choice format and the multidimensional comparison. In contrast, the Likert scales and the unidimensional forced-choice scales only differ in the response format; both involve judgments concerning only one dimension.

Table 5. Cross-Format Correlations Obtained in the Honest Condition

	Observed Correlation			Disattenuated Correlations		
Scale	SS-MFC	SS-UFC	MFC-UFC	SS-MFC	SS-UFC	MFC-UFC
Achievement	0.55	0.61	0.57	0.76	0.97	1.00
Adjustment	0.53	0.72	0.53	0.77	0.97	0.81
Attention Seeking	0.62	0.65	0.60	0.84	0.94	1.00
Cooperation	0.28	0.50	0.29	0.42	0.99	0.64
Dominance	0.69	0.75	0.68	0.87	0.95	0.93
Even-Tempered	0.57	0.67	0.50	0.89	1.00	0.94
Intellectual Efficiency	0.51	0.60	0.56	0.70	0.95	0.88
Non-Delinquency	0.38	0.61	0.45	0.60	1.00	1.00
Optimism	0.55	0.69	0.53	0.78	1.00	0.85
Order	0.60	0.69	0.60	0.82	0.99	0.88
Physical Condition	0.58	0.75	0.55	0.72	1.00	0.77
SelfControl	0.32	0.58	0.37	0.57	0.95	0.76
Selflessness	0.53	0.58	0.50	1.00	1.00	1.00
Sociability	0.58	0.73	0.57	0.78	0.97	0.80
Tolerance	0.59	0.70	0.60	0.79	1.00	0.96
Mean	0.52	0.66	0.53	0.75	0.98	0.88

Note: SS = single statement; MFC = multidimensional forced-choice; UFC = unidimensional forced-choice.

For comparison, Table 6 presents the convergent validity correlations obtained under the Faking condition. The observed correlations, as one might expect, are substantially smaller than those observed in the Honest condition. Many of the disattenuated correlations are large, mainly because the scale reliabilities obtained in the Faking condition were very low.

Table 6. Cross-Format Correlations Obtained in the Faking Condition

	Observed Correlation			Disattenuated Correlation		
Scale	SS-MFC	SS-UFC	MFC-UFC	SS-MFC	SS-UFC	MFC-UFC
Achievement	0.27	0.51	0.31	0.37	1.00	1.00
Adjustment	0.29	0.48	0.30	0.47	0.83	0.62
Attention Seeking	0.48	0.48	0.39	0.77	0.83	0.79
Cooperation	0.16	0.41	0.24	0.25	1.00	1.00
Dominance	0.51	0.56	0.54	0.69	0.91	0.98
Even-Tempered	0.25	0.44	0.17	0.43	1.00	0.58
Intellectual Efficiency	0.31	0.47	0.43	0.45	0.81	0.84
Non-Delinquency	0.32	0.43	0.31	0.66	0.87	0.98
Optimism	0.40	0.36	0.27	0.70	0.93	0.95
Order	0.31	0.46	0.41	0.52	0.86	0.80
Physical Condition	0.34	0.53	0.23	0.47	0.98	0.47
SelfControl	0.32	0.37	0.28	0.55	0.66	0.67
Selflessness	0.40	0.39	0.43	1.00	1.00	1.00
Sociability	0.35	0.57	0.25	0.59	0.91	0.52
Tolerance	0.42	0.44	0.40	0.68	1.00	1.00
Mean	0.34	0.46	0.33	0.57	0.91	0.81

5.5. Correlations with Criterion Variables

In the final set of analyses, the scales were correlated with the five criterion variables. These correlations appear in Tables 7 through 11. To provide an overall index of the prediction of a criterion from each of the three sets of scales, the criterion was regressed on the set of scales and the adjusted R^2 was computed. For the data obtained under the Honest condition, these tables also present correlations disattenuated for measurement error in the scales and criterion variables. Reliability estimates from Tables 2, 3, and 4 were used for the AF TAPAS scales. Due to concerns about overcorrection when reliability estimates are low (Zimmerman & Williams, 1997), .60 was used in the disattenuation formula when a reliability estimate was less than this value.

Perhaps the most salient feature of these tables is that scale scores computed from responses in the Faking condition do not predict the criterion variables. None of the adjusted R^2 values was larger than .17, and many were less than .10. In contrast, scale scores

from the Honest condition had adjusted R^2 values usually in the .2 to .4 range, which indicates fairly good prediction.

The Likert format scales generally had the highest correlations with the criterion variables. Their average adjusted R^2 was .38, which is noticeably higher than the adjusted R^2 values of the unidimensional forced-choice scales, .27, and multidimensional forced-choice scales, .22. One possible explanation for this pattern of results is that the criterion variables were assessed via Likert response scales, so the Likert AF TAPAS may have shared some mono-method response consistency error variance.

Table 7. Scale Correlations with Situational Decision-Making

	Situational Decision-Making									
	Honest				Honest Disattenuated				Faking	
Scale	SS	MFC	UFC		SS	MFC	UFC		SS	MFC	UFC
Achievement	-0.43	-0.32	-0.43		-0.52	-0.45	-0.61		-0.25	-0.04	-0.15
Adjustment	-0.20	0.00	-0.16		-0.25	0.00	-0.20		-0.17	0.01	-0.09
Attention Seeking	-0.11	-0.18	0.00		-0.13	-0.24	-0.01		0.10	0.00	0.09
Cooperation	-0.33	-0.11	-0.23		-0.42	-0.15	-0.33		-0.29	0.03	-0.17
Dominance	-0.28	-0.12	-0.24		-0.33	-0.15	-0.31		-0.21	-0.08	-0.14
Even-Tempered	-0.30	-0.03	-0.27		-0.38	-0.05	-0.39		-0.24	-0.05	-0.16
Intellectual Efficiency	-0.22	-0.14	-0.18		-0.28	-0.18	-0.26		-0.21	0.00	-0.09
Non-Delinquency	-0.43	-0.11	-0.39		-0.54	-0.16	-0.56		-0.29	-0.05	-0.16
Optimism	-0.33	-0.17	-0.31		-0.41	-0.23	-0.45		-0.28	-0.16	-0.09
Order	-0.28	-0.08	-0.27		-0.36	-0.10	-0.37		-0.19	-0.05	-0.16
Physical Condition	-0.13	0.03	-0.12		-0.16	0.03	-0.17		-0.20	-0.04	-0.10
SelfControl	-0.40	-0.16	-0.34		-0.53	-0.22	-0.48		-0.32	-0.18	-0.14
Selflessness	-0.40	-0.31	-0.40		-0.53	-0.44	-0.56		-0.24	-0.04	-0.08
Sociability	-0.28	0.06	-0.27		-0.35	0.07	-0.35		-0.20	0.07	-0.06
Tolerance	-0.35	-0.32	-0.35		-0.44	-0.41	-0.50		-0.20	-0.03	-0.02
Adjusted R2	0.29	0.23	0.31						0.07	0.01	0.01

Note: Observed correlations less than -0.09 are significant at $p = .05$ (one-tailed) and less than -0.13 are significant at $p = .01$ (one-tailed).

Table 8. Scale Correlations with Communication

	Communication									
	Honest				Honest Disattenuated				Faking	
Scale	SS	MFC	UFC		SS	MFC	UFC		SS	MFC	UFC
Achievement	0.43	0.20	0.25		0.52	0.28	0.37		0.31	0.00	0.13
Adjustment	0.30	0.09	0.31		0.38	0.12	0.41		0.26	0.19	0.21
Attention Seeking	0.07	0.24	0.03		0.09	0.33	0.04		0.20	0.18	0.12
Cooperation	0.32	0.08	0.26		0.42	0.11	0.37		0.30	0.16	0.09
Dominance	0.43	0.36	0.37		0.52	0.47	0.49		0.35	0.31	0.27
Even-Tempered	0.34	0.02	0.30		0.44	0.02	0.43		0.25	-0.06	0.15
Intellectual Efficiency	0.48	0.25	0.29		0.63	0.33	0.41		0.34	0.18	0.19
Non-Delinquency	0.28	-0.05	0.15		0.35	-0.07	0.22		0.11	-0.05	0.02
Optimism	0.37	0.15	0.29		0.47	0.20	0.42		0.31	0.17	0.15
Order	0.32	0.19	0.26		0.42	0.25	0.36		0.15	0.10	0.14
Physical Condition	0.15	-0.13	0.14		0.18	-0.16	0.20		0.30	0.07	0.20
SelfControl	0.51	0.02	0.24		0.68	0.02	0.34		0.29	0.03	0.10
Selflessness	0.35	0.14	0.18		0.47	0.20	0.26		0.21	0.07	0.06
Sociability	0.27	0.05	0.24		0.34	0.06	0.32		0.36	0.20	0.20
Tolerance	0.42	0.22	0.33		0.54	0.29	0.48		0.18	0.06	0.06
Adjusted R2	0.41	0.23	0.22						0.17	0.15	0.06

Note: Observed correlations greater than 0.09 are significant at $p = .05$ (one-tailed) and greater than 0.13 are significant at $p = .01$ (one-tailed).

Table 9. Scale Correlations with Decision-Making and Managing Resources

	Decision-Making and Managing Resources									
	Honest				Honest Disattenuated				Faking	
Scale	SS	MFC	UFC		SS	MFC	UFC		SS	MFC	UFC
Achievement	0.57	0.36	0.43		0.68	0.50	0.60		0.30	0.07	0.23
Adjustment	0.31	0.09	0.30		0.39	0.12	0.39		0.20	0.09	0.18
Attention Seeking	0.09	0.23	0.01		0.11	0.31	0.01		0.01	0.10	-0.05
Cooperation	0.29	0.05	0.20		0.36	0.08	0.28		0.33	0.09	0.19
Dominance	0.48	0.37	0.43		0.57	0.47	0.55		0.24	0.20	0.22
Even-Tempered	0.29	-0.02	0.28		0.37	-0.03	0.39		0.22	0.01	0.18
Intellectual Efficiency	0.43	0.24	0.28		0.55	0.30	0.39		0.27	0.13	0.21
Non-Delinquency	0.35	-0.07	0.25		0.43	-0.09	0.35		0.19	0.05	0.13
Optimism	0.39	0.15	0.31		0.48	0.20	0.43		0.29	0.09	0.17
Order	0.38	0.20	0.40		0.48	0.26	0.54		0.21	0.18	0.19
Physical Condition	0.28	0.01	0.27		0.34	0.01	0.37		0.24	0.03	0.19
SelfControl	0.53	0.02	0.28		0.69	0.03	0.39		0.37	0.10	0.21
Selflessness	0.40	0.12	0.23		0.51	0.17	0.32		0.22	0.08	0.11
Sociability	0.31	-0.02	0.23		0.38	-0.03	0.29		0.18	-0.04	0.07
Tolerance	0.41	0.24	0.32		0.51	0.30	0.45		0.18	0.08	0.20
Adjusted R2	0.45	0.25	0.31						0.13	0.04	0.04

Note: observed correlations greater than 0.09 are significant at $p = .05$ (one-tailed) and greater than 0.13 are significant at $p = .01$ (one-tailed).

Table 10. Scale Correlations with Leading Others

	Leading Others									
	Honest				Honest Disattenuated				Faking	
Scale	SS	MFC	UFC		SS	MFC	UFC		SS	MFC	UFC
Achievement	0.44	0.20	0.31		0.55	0.29	0.46		0.23	-0.04	0.14
Adjustment	0.35	0.16	0.34		0.46	0.24	0.47		0.14	0.10	0.16
Attention Seeking	0.16	0.28	0.16		0.20	0.39	0.23		0.05	0.18	0.11
Cooperation	0.53	0.23	0.31		0.70	0.34	0.45		0.26	0.04	0.09
Dominance	0.43	0.31	0.42		0.54	0.42	0.57		0.21	0.25	0.15
Even-Tempered	0.33	0.04	0.26		0.43	0.07	0.39		0.10	-0.02	0.08
Intellectual Efficiency	0.28	-0.03	0.10		0.38	-0.04	0.15		0.15	-0.01	0.06
Non-Delinquency	0.34	-0.03	0.23		0.44	-0.04	0.34		0.04	-0.15	0.05
Optimism	0.45	0.26	0.37		0.58	0.37	0.55		0.27	0.10	0.10
Order	0.23	0.06	0.20		0.30	0.09	0.29		0.07	0.09	0.05
Physical Condition	0.27	-0.06	0.20		0.34	-0.08	0.28		0.26	0.08	0.12
SelfControl	0.30	-0.04	0.16		0.41	-0.06	0.23		0.15	-0.02	0.05
Selflessness	0.47	0.23	0.25		0.64	0.33	0.36		0.14	0.05	0.02
Sociability	0.43	0.23	0.37		0.55	0.31	0.50		0.21	0.10	0.19
Tolerance	0.34	0.19	0.21		0.45	0.25	0.31		0.02	0.04	0.03
Adjusted R2	0.38	0.24	0.27						0.10	0.07	0.01

Note: Observed correlations greater than 0.09 are significant at $p = .05$ (one-tailed) and greater than 0.13 are significant at $p = .01$ (one-tailed).

Table 11. Scale Correlations with Displaying Professionalism

	Professionalism									
	Honest				Honest Disattenuated				Faking	
Scale	SS	MFC	UFC		SS	MFC	UFC		SS	MFC	UFC
Achievement	0.49	0.22	0.30		0.60	0.32	0.44		0.22	0.03	0.09
Adjustment	0.33	0.12	0.28		0.42	0.18	0.37		0.15	0.04	0.18
Attention Seeking	0.12	0.23	0.07		0.15	0.32	0.10		0.00	0.13	0.02
Cooperation	0.46	0.14	0.22		0.60	0.20	0.32		0.32	0.08	0.16
Dominance	0.38	0.25	0.41		0.47	0.34	0.55		0.16	0.15	0.09
Even-Tempered	0.43	0.08	0.35		0.56	0.12	0.51		0.19	0.00	0.14
Intellectual Efficiency	0.32	0.08	0.21		0.42	0.11	0.31		0.17	-0.04	0.04
Non-Delinquency	0.39	0.02	0.28		0.51	0.02	0.41		0.17	-0.07	0.12
Optimism	0.40	0.19	0.30		0.51	0.27	0.44		0.27	0.06	0.13
Order	0.23	0.04	0.23		0.30	0.06	0.32		0.17	0.08	0.02
Physical Condition	0.19	-0.06	0.15		0.23	-0.08	0.21		0.21	-0.03	0.08
SelfControl	0.45	0.03	0.22		0.60	0.04	0.32		0.21	0.09	0.15
Selflessness	0.41	0.19	0.21		0.55	0.27	0.31		0.22	0.11	0.11
Sociability	0.32	0.07	0.24		0.40	0.09	0.32		0.17	-0.02	0.09
Tolerance	0.38	0.22	0.30		0.49	0.29	0.43		0.16	0.19	0.13
Adjusted R2	0.37	0.13	0.26						0.06	0.05	0.02

Note: Observed correlations greater than 0.09 are significant at $p = .05$ (one-tailed) and greater than 0.13 are significant at $p = .01$ (one-tailed).

6.0 CONCLUSIONS

Three personality inventories were developed. Form A is a traditional Likert scale where respondents are instructed to rate each statement on a 4-point scale. Form B and Form C are forced-choice scales where respondents are instructed to choose the statement that is "more like me." In Form B, statements measuring the same construct but having different extremities are paired. In Form C, statements measuring different constructs but having similar extremity and desirability are paired. Scoring algorithms for each form were provided.

An initial study investigating the psychometric properties of the three scales was conducted. The reliabilities of the Likert format scales were good, ranging from .70 to .85, with a mean of .77. The reliabilities were lower for the forced-choice measures, which are less likely to capitalize on single subject response consistency error, with a mean of .56 for the unidimensional forced-choice scales and .62 for the multidimensional forced-choice scales.

Given the lower than desired reliabilities for the forced-choice scales, two options might be considered. First, as with the Army's TAPAS, computer adaptive measurement might be considered. For example, in a simulation study, Stark, Chernyshenko, Drasgow, and White (2012) found that a 10-facet assessment with 5 items per facet had a reliability of .73 as a static test but a reliability of .84 when administered adaptively. Another, less burdensome approach is to use empirical Bayes estimation, which the trait estimate for each trait "borrows strength" from the other trait estimates. The "borrowed strength," usually called ancillary information in the statistical literature, can substantially improve reliability. This method, described in detail by Wainer et al. (2001), was recently used on data described by Zhang et al. (in press). IRT marginal reliability increased by about .05 on average, with the largest gains coming from the least reliable scales. The IRT marginal reliability of the Selflessness scale, for example, increased from .64 to .75. Moreover, the test-retest reliability for the empirical Bases Selflessness scores was .72 versus .67 for the original scores.

The convergent validity cross-method correlations were found to be very good to excellent for data collected in the Honest condition. For the observed scores, they ranged from a mean correlation of .52 for the Likert format scales with multidimensional forced-choice scales to a mean correlation of .66 for the Likert format scales with the unidimensional forced-choice scales. After correcting for measurement error, the mean disattenuated correlations ranged from .75 to .98. For data collected in the Faking condition, cross-method correlations were much lower, ranging from a mean of .33 to .46 for the observed correlations and from .57 to .91 for the disattenuated correlations.

Finally, correlations of the AF TAPAS scales with five criterion variables were examined. They were found to be relatively high for the AF TAPAS scales that shared a response format (Likert) with the criterion scales, but still substantial for the forced-choice scales.

7.0 REFERENCES

Allender, C. J. M. (2005). Predicting leadership and performance in uniformed organizations using the five factor model of personality. Unpublished doctoral dissertation, University of Plymouth.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis, *Personnel Psychology*, *44, 1*, 1-26.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment, *Applied Psychological Measurement*, *6, 4,* 431-444.

Brown, A., Inceoglu, I., & Lin, Y. (2017). Preventing rater biases in 360-degree feedback by forcing choice, *Organizational Research Methods*, *20, 1,* 121-148.

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires, *Psychological Methods, 18, 1,* 36-52.

Campbell, J. P., & Knapp, D. J. (2001). *Exploring the Limits in Personnel Selection and Classification,* Mahwah, NJ: Erlbaum.

Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology, 104, 11*, 1347-1368.

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, 1*, 88-106.

Chiaburu, D. S., Oh, I. S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality Traits and organizational citizenship behaviors: A Meta-analysis, *Journal of Applied Psychology, 96, 6,* 1140-1166.

Christal, R. E., Barucky, J. M., Driskill, W. E., & Collis, J. M. (1997). *The Air Force Self-Description Inventory (AFSDI): A summary of continuing research,* F33615-91-D-0010, Armstrong Laboratory, Brooks AFB, San Antonio, TX.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment, *Human Performance*, *18, 3,* 267-307.

Costa, P. T., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R), In D. H. Saklofske (Ed.), *The SAGE Handbook of Personality Theory and Assessment, Vol. 2: Personality Measurement and Testing* (pp. 179-198), Thousand Oaks, CA: Sage.

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 dupport Army personnel selection and classification decisions,* Technical Report 1311, Drasgow Consulting Group, Urbana, IL.

Grijalva, E., & Newman, D. A. (2015). Narcissism and counterproductive work behavior (CWB): Meta-analysis and consideration of collectivist culture, Big Five personality, and narcissism's facet structure, *Applied Psychology*, *64, 1*, 93-126.

Houston, J. S., Borman, W. C., Farmer, W. L., & Bearden, R. M. (2005). *Development of the Enlisted Computer Adaptive Personality Scales (ENCAPS), Renamed Navy Computer Adaptive Personality Scales (NCAPS),* Technical Report No. 502, Navy Personnel Research, Studies, and Technology (NPRST). Navy Personnel Command, Millington, TN.

Huang, J. L., Ryan, A. M., Zabel, K. L., & Palmer, A. (2014). Personality and adaptive performance at work: A meta-analytic investigation, *Journal of Applied Psychology*, *99, 1,* 162-179.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review, *Journal of Applied Psychology*, *87, 4,* 765-780.

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. (2007). Self-efficacy and work-related performance: The integral role of individual differences, *Journal of Applied Psychology*, *92, 1,* 107-127.

Lentz, E., Horgen, K. E., Borman, W. C., Dullaghan, T. R., Smith, T., Schwartz, K. L., & Weissmuller, J. J. (2009). *Air Force officership survey Volume II: Performance requirement linkages and predictor recommendations* (Technical Report AFCAPS-FR-2010-0010). Randolph AFB, TX: HQ AFPC/DSYX, Strategic Research and Assessment.

Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires?, *Personality and Individual Differences*, *32, 2,* 247-256.

Pavlov, G., Maydeu-Olivares, A., & Fairchild, A. J. (2018). Effects of applicant faking on forced-choice and Likert dcores, *Organizational Research Methods*, Advanced on-line publication.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2005). A general item response theory model for unfolding unidimensional polytomous responses, *Applied Psychological Measurement*, *24, 1,* 3-32.

Schmidt, F. L., & Hunter, J. E., (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings, *Psychological Bulletin*, *124, 2,* 262-274.

Stark, S., Chernysheko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model, *Applied Psychological Measurement*, *29, 3,* 184-203.

Stark, S., Chernyshenko, O. S., Drasgow, F., & White, L. A. (2012). Adaptive testing with multidimensional pairwise preference items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods, 15,* 463-487.

Tett, R. P., Steele, J. R., & Beauregard, R. S. (2003). Broad and narrow measures on boths Sides of the personality–job performance relationship, *Journal of Organizational Behavior, 24,* 335-356.

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis, *Journal of Applied Psychology*, *97, 3*, 499-530.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakeability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59,* 197–210.

Wainer, H., Vevea, J. L., Camacho, F., Reeve, B. R. III, Rosa, K., Nelson, L., Swygert, K. A., & Thissen, D. (2001). Augmented scores – "Borrowing strength" to compute scores based on small numbers of items. In D. Thissen & H. Wainer (Eds.), *Test Scoring* (pp. 343-387), Mahwah, NJ: Erlbaum.

White, L. A., Nord, R. D., Mael, F. A., & Young, M. C. (1993). The Assessment of Background and Life Experiences (ABLE), In T. H. Trent & J. H. Laurence (Eds.), *Adaptability Screening for the Armed Forces* (pp. 101-161), Office of the Assistant Secretary of Defense, Washington, DC.

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.

Zhang, B., Sun, T., Drasgow, F., Chernyshenko, O. S., Nye, C.D., Stark, S., & White, L. A. (2019). Though forced, still valid: Psychometric equivalence of forced choice and single statement measures. *Organizational Research Methods*, in press.

Zimmerman, D. W., & Williams, R. H. (1997). Properties of the Spearman correction for attenuation for normal and realistic non-normal distributions. *Applied Psychological Measurement, 21,* 253-270.

8.0 LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

ABLE	Assessment of Background & Life Experiences
AIM	Assessment of Individual Motivation
EAP	Expected A Posteriori
GGUM	Generalized Graded Unfolding Model
IRT	Item Response Theory
MDPP	Multidimensional Pairwise Preference
NCAPS	Navy Computer Adaptive Personality Scales
NEO-PI-R	Neuroticism-Extraversion-Openness Personality Inventory-Revised
TAPAS	Tailored Adaptive Personality Assessment System
TDSI	Trait Self-Description Inventory
UPP	Unidimensional Pairwise Preference
USAF	United States Air Force