

AFCAPS-TR-2012-0002

“Personality” and Mission Effectiveness

IMTA Conference
October 2006



Johnny J. Weissmuller
Kenneth L. Schwartz
AFPC/Strategic Research and Assessment
Branch



Kenneth L. Schwartz
Johnny J. Weissmuller

**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

Approved for Public Release. Distribution Unlimited

UNCLASSIFIED

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.

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 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				<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE		3. DATES COVERED (From - To)	
		Final		October 3-5 th , 2006	
4. TITLE AND SUBTITLE "Personality" and Mission Effectiveness				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Johnny J. Weissmuller and Kenneth L. Schwartz				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Personnel Center Force Management Liaison Office Randolph AFB TX 78150				8. PERFORMING ORGANIZATION REPORT NUMBER	
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-TR-2012-0002	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release. Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Paper covers the history of Air Force personality research, the current need for personality measures, and the move towards a change in psychometrics that will replace Item Response Theory in performance testing and personality assessment. Details the history and development of the Self Description Inventory (SDI) and SDI+; the need and development of personality measures to enhance person-job-match and increase retention; and the future move towards psychometrics being used to evaluate items based on the decision logic approach. This approach is based on the ability to distinguish between desired classification decisions and not the correlation with a total test vector representing a criterion based upon hypothesized areas of importance.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT U	18. NUMBER OF PAGES 41	19a. NAME OF RESPONSIBLE PERSON Kenneth L. Schwartz
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code) 210-565-3139

(Page Intentionally left blank)

“Personality” and Mission Effectiveness

Johnny J. Weissmuller

HQ Air Force Personnel Center,
Military Testing (AFPC/DPPPWT-1)
550 C St West, Suite 9
Randolph AFB, Texas 78150-4771 USA
Johnny.Weissmuller@Randolph.AF.mil

*Presentation at the Personality and the Military Symposium II
48th International Military Testing Association (IMTA) Conference,
Kingston Ontario, Canada, 4 October 2006*

Introduction

In support of our Air Staff initiatives, we in the Military Testing Section (Ken Schwartz, Chief, USAF Military Testing and I) have spent the last two months working to revitalize the USAF personnel research program for strategic force management issues. This has consumed a large portion of our attention up through 4:00 PM on Saturday, September 30th, the last day of our fiscal year. I am happy to report that we were successfully in launching four major “general personnel research” contracts totally nearly two million dollars from unexpected fallout funds.

In one of these contracts there is a task to evaluate the effectiveness and potential for using the new Self Description Inventory (SDI+) which has become part of the Air Force Officer Qualifying Test (AFOQT) since August 2005. We now have nearly 10,000 AFOQT answer sheets which represent not only current USAF Officers, but also “applicants” in the sense that they entered the commissioning source, but may not have progressed through final commissioning. This dataset will be an invaluable asset in our first, near-term plans for evaluating personality measures in any future operational Air Force personnel program.

Now that the stage has been set (as to why this symposium discussion can cover topics as recent as last week), let me outline the three topics to be covered today. These include 1) an overview of the last forty-nine years of Air Force personality research (1957-2006), 2) the currently recognized need for personality measures, and 3) the coming revolution in psychometrics which will replace Item Response Theory in performance testing and personality assessment. As we have 15 minutes for presentation in today’s symposium and I am known to speak quickly – upon completion, I yield back the balance of my time to our esteemed colleagues who follow.

Part 1: Forty-Nine Years of Air Force Personality Research

Given today's time constraints, I will focus in on only that particular research stream which resulted in the Self-Description Inventory (SDI+) currently on the Air Force Officer Qualifying Test (AFOQT).

Personality theories go back a long way and many papers in the first half of the twentieth century proposed "traits" to describe individuals (see Tupes and Christal review: <http://www.icodap.org/papers/SDI/TUPES/ASD-TR-61-97.pdf>). It was, however, a 1958 landmark paper by Air Force researchers, Tupes and Christal which collected several proposed models, established a measurement instrument, collected measures with varying populations under varying rating scenarios, and subjected all this data to the emerging power of the new electronic data processing machine – the computer (Tupes, E.C. & Christal, R.E. *Stability of personality trait rating factors obtained under diverse conditions*. Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Center, May 1958, Technical Note WADC-TN-58-61, ASTIA Document No. AD-151 041.) What emerged here and in follow-up studies was a recurrent factor structure with five main components. This is was the birth of what is now known as the "Big Five" model of personality.

While the Air Force did not do the original conceptual formulation of the personality traits being measured, they did the steps necessary to combine desperate proposed models and organize them into a common framework for operational application. Air Force personnel research often has done a review of the literature, integrated competing models, and developed a workable instrument to achieve operational results rather than produce a journal-quality article. While this point may seem a little sharp, it is, in fact, the point of part three of this discussion.

The Air Force, for whatever reasons, did not pursue the "Big Five" for its own uses. The "Big Five" model was, however, picked up in the civilian community and is available from commercial sources. The climate and acceptance of personality and related non-cognitive measures has changed over the past fifty year (see part 2 below). In fact, today the Air Force uses a commercial "Big Five" instrument (among other similar tests) to establish medical/psychological baselines for pilots used for comparison after serious injury to assist in "return to duty" determinations.

The second author on that landmark paper was Dr. Raymond E. Christal. Dr. Christal eventually became the Chief of the Occupation and Manpower Research Division of the Air Force Human Resources Laboratory. During this time I became his chief analyst in the Computation Sciences Division and he became my mentor. In 1980 we discussed his early work in personality and I insisted that I could not imagine a scenario under which the Air Force would ever be convinced to deploy such measures operationally.

In 1992 we conducted a Critical Incident analysis of Air Force missions from the first Gulf War (*Critical Incidents Technique to Identify Attributes Associated with Aircrew Combat Performance*, Final Report F33615-91-D-0010-DO5, 15 Dec 1993). The study collected 100-140 “incidents” for each of seven aircraft/weapon system types. Seven panels of pilots (n=14-20) were assembled and asked to sort index cards with their aircraft’s incidents into 2 to 10 piles to represent different kinds of problems. The intent of the researchers was to identify classes of problems that might suggest the kinds of training that could be provided in the future to avert these problems.

I used the ALSCAL program to perform a multi-dimensional scaling (MDS) solution to integrate all the pilots’ decisions into the optimal set of stacks. The mathematically determined result indicated that six stacks best represented the intent of the combined pool of pilots and aircraft. I was not surprised by the largest and most overwhelming solid pile which read: “encountered a mechanical problem – solved by the book.” Subsequent analysts and authors were bothered by this stack because there was no *one* outstanding element was evident by which to label this dimension other than as “effective, professional pilot and crew.”

The content of other much smaller five smaller stacks took me quite by surprise. It should be noted that the pilot panel members were asked to label each pile after they were done. These hundred of labels included many colorful phrases, but among them were: “loss of situational awareness”, “cowboy”, ‘I’ve got a secret”, “ineffective planning,” etc. Although the mapping was not one-to-one, I knew I was looking at a solution which was best described by the Big Five model. I went back to Dr. Christal and apologized for my 1980 comment saying, “the scenario I had failed to envision was the Air Force delivery of ordnance in combat.”

What followed was eleven years of research (1993-2004) to re-tool the original “trait rating by others” approach into a self-report format. The combination of behavioral phrases (n=99) and bi-polar traits (like “silent-talkative”, n=64) were all replaced by self-descriptive phrases (n=220*). The items are now like “**I enjoy reading poetry.**” And the rating scale is a five-point, bi-polar, Likert scale from “Strongly Disagree” to “Strongly Agree.” The revised instruments were tested on both enlisted personnel and officers from the USAF and the Royal Air Force. As noted in the opening remarks, the new Self-Description Inventory (SDI+) went operational on 1 August 2005 and currently has 10,000 answer sheets awaiting analysis.

*As final note on this topic, the instrument is referred to as the SDI+. The plus sign is to indicate that two additional factors (covered by 80 items) have been added to the “Big Five” for operational use within the military. These two factors include “Service Orientation” and “Team Orientation.” These factors measure the degree to which a person is likely to make the service a career (i.e. for long-term return on investment analyses) and a person’s ability to accept differing roles under changing conditions (i.e. support evolving mission statements with fewer and fewer personnel.) So, in the SDI+, the mnemonic for remembering the factors now becomes “OCEAN ST” for Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism, Service, and Team scales.

Part 2: Recognized Need for Personality Measures

An article in a recent Air Force Times (Terry D. Stevens, 25 Sept 2006, p 54) highlighted DoD's recognized need to improve their track record in selecting recruits who can both complete training and serve out their full military commitment. Two quotes from this article by Terry D. Stevens tell the essential story:

1. "According to Defense Department figures, only 70 percent of high school graduates make it through their first enlistment."
2. "There is an inexpensive process available for enhancing the quality and reducing first-term separations of enlistees, but the Defense Department doesn't use it. Millions of job applicants must complete a personality or behavioral assessment before being hired..."

While the referenced article is an editorial for a given perspective, there is current evidence that the use of personality assessments may make sense for the military. The traditional use of these instruments is to filter-out or deny entry based on the results. We, however, are undertaking research with a different focus.

In an era where the selection ratio is very low (perhaps one acceptance to every 1.6 applicants by some estimates), a "screen-out" policy doesn't make a lot of sense. Rather, then, if one uses the personality "constellation pattern" to improve the person-job match of the selected individuals, perhaps one can make better use of the people brought into the service. In fact, with changing missions and Air Force internal restructuring, current service members are viewed as a resource which competes with incoming recruits for filling emerging roles. We, in the Military Testing Section, are acutely aware of this fact whenever we're called upon to "re-norm" existing DoD enlistment and classification test scores for people already in the Air Force master personnel system. This re-norming is essential to proper decision making when you have the option of using experienced service members or recently trained troops in newly defined mission areas.

There is, however, concern on the privacy issue and the ease with which "personality measures" may be misused. How this information could be integrated into the assignment system is not at all clear. Some have suggested that the SDI+ be made available over the Internet and that service members could self-administer the inventory at will and take the confidential results to a military career counselor for off-the-record advice. There are many issues to be worked out before we go down any of these paths. We are just now embarking on the research needed to determine the possible benefits from a range of utilization options so that an informed policy decision can be made in the future.

Part 3: The Coming Revolution in Psychometrics

Had this paper been finalized before last week, it would have ended above. I spent this last week at the Performance Testing Council Summit Meeting in Washington DC. As a guest at this conference, I attended the meetings of the Psychometric Committee. My purpose in attending was to find the conceptual framework required to “score” and possibly integrate future performance test measures into the Air Force’s current promotion system. What I found looks like an effective approach to handling personality “constellations” for improving future person-job-match systems.

Background on the Air Force Promotion System

The Military Testing Section is the policy shop for oversight on the development of Air Force promotion tests used in the Weighted Airman Promotion System (WAPS.) Over 200,000 promotion test answer sheets are scanned each year.

The WAPS methodology was developed in the late 1960’s and was revalidated as recently as the summer of 2004. In the WAPS for promotion to pay grades E-5 through E-7 there are two basic types of test – one for job knowledge (the Specialty Knowledge Test – SKT) and one for general Air Force and military knowledge (the Promotion Fitness Exam – PFE). These are both knowledge power tests. For promotion to the two highest enlisted ranks (E-8 and E-9 – the Senior NCO Promotion Program - SNCOPP), those two tests are replaced by the Supervisory Exam (USAFSE) and a Promotion Board Score which alone accounts for half of the total points available.

The Air Force promotion system, WAPS, adds together weighted values for the test scores along with points for Performance Reports, Decorations, Time-in-Service, Time-in-Grade and Board Score for (E-8 & E-9 cycles only) to produce a WAPS Total Score. The WAPS Total Score is used to rank-order eligible airmen for promotion.

For each pay grade (E-5 through E-9), the needs of the Air Force determine the percent of all eligibles to be promoted during the annual cycle. This percentage is applied, essentially across the board to each of the 150+ job families in the Air Force at that pay grade. Service members compete only against others in their own job family (Air Force Specialty -AFS) for promotion. For this reason, the total WAPS score needed for promotion (the “cut score”) is different for each AFS. Note that promotion eligibility requirements for each grade include Time-in-Service and Time-in-Grade minimums so that “good test takers” cannot skyrocket to the top.

To further complicate this system from a psychometric standpoint, there are over 400 different promotion tests administered each year. To ensure the test area weights reflect the current needs of the Air Force, the test outline area weights are driven by a panel of 3-5 Subject Matter Expert (SME) item-writers using a recent survey of Testing Importance

recommendations from 50-75 senior field supervisors and a census task-level occupational survey that is no more than three-years old. The tests are changed annually. These test development projects were formally known as “major” and “minor” revisions to denote the percent of the test being replaced. The Air Force Occupational Measurement Squadron (AFOMS) recently fielded a new Automated Test Development System (ATDS) which will facilitate major changes in these Air Force promotion tests every year.

The most salient point here is that the order-of-merit for promotion is the result of multiple factors. While there is a “cut score” on the WAPS total score for promotion it is NOT applied directly to the test scores themselves (i.e., there is no minimum “passing score.”) In fact, this is a compensatory model in which the “high drivers” can vary from year to year (depending on the demographics of the eligible pool) and offers a self-correcting feature which values knowledge, experience and demonstrated performance.

Performance Testing

On recommendation from our Navy counterparts, Ken Schwartz (Chief, USAF Military Testing Section) and I attended the Performance Testing Council Summit. The Navy is further along in moving their high-stakes promotion (“advancement”) tests to computerized format. Their initial evaluations have used Navy ratings (job families) which could benefit most from multi-media presentations (MMP). These MMP items begin to cross the line from the typical psychometric assumptions which underlie the Item Response Theory for traditional multiple choice questions (MCQ) and the more complex analysis requirements of scenario-linked data that are common in performance tests.

The Performance Testing Council and its member organizations have been addressing these issues for the past several years. This organization was underwritten, in large part, by the Information Technology (IT) community which moved into performance testing in order to certify large number of people into critically understaffed IT job areas. While the founding members may have been those who train, certify, and employ IT people, the current Performance Testing Council (PTC) is open to a wider range of members and issues related to all types of performance tests.

Although my intent in attending was to identify a conceptual framework for integrating performance testing into a nearly 40-year old promotion program, I believe I have found the tools appropriate to the analysis and proper use in person-job-match employing personality assessments.

Classification Testing Using Decision Theory

The PTC Psychometric Committee had an invited speaker, Dr. Larry Rudner of the Graduate Management Admission Council (GMAC®). Dr. Rudner presented *Classification Testing Using Decision Theory*. Although I believe that most of us psychometricians in the audience were overwhelmed by the speed, mathematical complexity and technical issues (my Java applet didn't work) from this presentation, the concepts presented were very clear. This approach is very much akin to a discriminate analysis approach. At the very core of this approach is the insight that the purpose of any test is to distinguish between those who will be successful (masters) and those who will not (non-masters.) This method does generalize to multiple classifications, but for illustratively purposes, this master/non-master dichotomy communicates best. At the heart of this approach is the ability to create a benchmark sample of your classifications, such as a group of recognized "masters" and "non-masters." Following this ultimate, high-level, global determination (in essence, a pre-test validation study) one administers the proposed test to determine how each and every items fares in distinguishing between the classifications at hand. These performance characteristics of the items determine their information loading and power in achieving the true goal of the test – distinguishing membership between classification choices.

This method uses an information-based metric to characterize an items ability to distinguish between the responses of "masters" versus the responses of "non-masters." In addition to this item-level approach, Dr. Rudner showed how this metric can be employed to produce an adaptive testing system as an ALTERNATIVE to the adaptive testing based on Item Response Theory (IRT). The concept of adaptive testing is to resolve down to an end result by taking a Bayesian approach in which subsequent items are selected to provide maximum information (resolving power) given all responses up to this point. Dr. Rudner not only demonstrated how this would work with information-based metrics but also showed how one might incorporate the cost of "wrong decisions" to enhance the outcome of the decision logic.

What Dr. Rudner made clear was the set of assumptions implicit in IRT that one may not wish to endorse. IRT assumes that a higher score on the test is better than a lower score which drives the discrimination parameter. In this model, "discrimination" means the ITEM's ability to distinguish higher scoring individuals from lower scoring individuals and that which doesn't seem that far off base.

The approach that "more is better" in terms of score assumes that all items are "good" items and relate to the uni-polar (zero-anchored) uni-dimensional ability (theta) measured by the test. (Note that personality measures do NOT meet this assumption.) IRT further assumes that these items are presented and scored properly and a "correct" response increases the implication of mastery. Adaptive tests built on the IRT model assume that all items measure the same uni-dimensional ability as reflected by "total score."

Testing in a Multi-Dimensional World

There was a fundamental assumption which was not explicitly mentioned, but which, given our Air Force test construction practices, became very clear to me. The Air Force occupational analysis program drives test development starting with job tasks distributed in census surveys. These surveys are put through a hierarchical clustering system which elucidates the number of different “job types” actually required to accomplish the work domain covered by the target job title. In addition to complaints about translating tasks (i.e. performance requirements) into knowledge test items, Subject Matter Experts also labor over creating the test blueprint (i.e., formulating a test outline and assigning appropriate weights (number of items per topic) to each area and sub-area.)

When outlines are properly weighted by Subject Matter Experts it is expected that some areas (typically “safety” for example) will show lower item-total correlations and (consequently) reduced test reliabilities – even though it was agreed that the test validity was, in fact, improved. The world of performance is not uni-dimensional. Any model which assumes uni-dimensionality should limit its attention to aptitude tests where the use of real-world performance tests is not appropriate.

Correlation with total test score is the result of adding all these assumed correct/incorrect scores together. The test outline or evidence based design establishes the number of test items in each area. Larger areas contribute more weight to the total score. In other words, how the items were selected for inclusion drives the balance (i.e. weight) of areas covered by the test and drives the primary dimension with which “agreement” will be measured.

If this balance does not reflect the dimensionality and weighting of the real work, this synthetic criterion vector may take on a uni-dimensional character even if the true criterion space requires multiple dimensions. The way in which reliability is computed will lead to lower and lower reliability estimates as the actual validity (ability to predict success in jobs with multiple domains) actually goes up and up. Conversely, ignoring dimensions necessary for “master performance” can lead to higher reliability but less validity. In other words, using “reliability” as a primary consideration in creating a test may become a self-fulfilling prophecy much to the detriment of actual validity.

The coming revolution in psychometrics will be driven by the observation that the Decision Logic approach evaluates each item against its ability to distinguish between desired classification decisions and not the correlation with a total test vector representing a synthetic criterion resulting from test designer’s pre-conceived notions of area importance.

Does this mean that the Decision Logic approach is a panacea for test outline development? No, definitely not. It simply means that an item’s evaluation really is independent of other items under consideration and whether or not they are constructed / scored properly and represented in the necessary balance by the test designers.

The Air Force's traditional problem of re-working "safety" items because they do not correlate with the test total score will be corrected. If knowledge of safety issues distinguishes masters from non-masters, safety items will receive a reasonable "information" weight and appear more important than in the past under classical analysis approaches (i.e. increased validity will be recognized.)

Can the Decision Logic create inappropriate uni-dimensional tests like Item Response Theory? Yes, definitely. While each and every test item is an independent measure, there is no inherent magic to generate the SET and RANGE of items required to explain the work space. The use of Subject Matter Experts is still required to establish outline weights to be covered. If important areas are not covered there will be no indication of missing variance to account for by adding additional items. One cannot measure dimensions for which no test items are written. It's basic epistemology that every "sense" requires a sensor tuned to the proper input. If items are written for even the smallest "real" dimension in the master's skill set, however, the Decision Logic approach will acknowledge its contribution in defining "master."

One serious caveat for the Decision Logic approach is that some test items may be essential for certifying "competent" behavior. This is a second class of items which need to be treated separately. These items, whether or NOT they distinguish between masters and non-masters (i.e., offer no discrimination/information), still need to be asked and correct response documented.

Where will the Decision Logic Approach Offer an Advantage?

The Decision Logic approach seems to hold promise in the two application areas for which I am involved: the complex Air Force Promotion Test Development process and in utilizing Personality Assessment Inventories in the real world.

Promotion Tests under WAPS

Recall from Part 1 above that airmen are promoted based on a total score which includes several other scores in addition to percent correct on the promotion tests. Using the Decision Logic approach, I can classify those who were promoted (masters) and those who were not (non-masters). Now, the value of item test item can be evaluated for its informational value in making that broad "whole person" distinction. Because promotion rates vary from year to year, this process may be accomplished at several standard levels, say 10%, 20% and 30% promotion rates. A sensitivity analysis might be performed to assess impact of promotion rate on item information loading and final performance decisions. For the first time, some areas, such as safety, may no longer be identified as targets for re-write because of low item-total correlation. With the Decision Logic approach their true value may be more accurately reflected.

Personality Assessment Items

Personality Assessment items are bi-polar rather than zero-anchored measures. In these measures (such as Neuroticism) – “more” is not always “better.” Even using reverse scaling strategies is not a viable answer to negative or bi-polar scales. For example, some jobs like Public Affairs may be more acceptable to people with a high extroversion score while the Intelligence Analyst job may be more suited to people who are more methodical and more introverted (a low or negative extroversion score). In other words, the meaning of the “best” personality profile may vary by job title.

Real Air Force jobs, even before future classification mergers, often show that a single job title may, in fact, represent several different recognizable job types. Although people are assigned to positions or open billets, the work required at the job location may be any of the job types identified for that job family. These job types require differing skill constellations and are, at present, considered interchangeable for assignment purposes. This finding came from the Job Structuring Technology (JST) contract (1992-1996). In the JST contract, task modules (called performance dimensions) within job families were rated on the traditional skill set categories used in the Dictionary of Occupational Titles (DOT) – namely Sydney Fine’s “People – Data – Things” dimensions. It was shown that various performance dimensions covered by one job title do have very different “People – Data – Things” requirements profiles. The point is that a single job title may also have different and distinct “ideal” personality constellations.

The need for some type of external criteria to determine effective classifications became obvious. The Decision Logic approach has the advantage of using the differential resolving power of an item without relying on *a priori* assumptions about the value of a “correct” or “incorrect” response. In fact, this approach can use items which do not have a scoring “key” at all – like bi-polar self-report personality trait affirmations. If the goal is to improve retention, then one suspects that rather than a classification category based on “master” versus “non-master,” that the classification dichotomy should be “satisfied with the Air Force and current job” and “not satisfied with the Air Force or current job.”

Standard USAF Job Inventories include some basic job satisfaction questions which could serve as the basis for these classification determinations. If the SDI+ could be administered concurrently with standard occupational analysis surveys, the information loading contribution of each item could be assessed in its ability to draw this distinction. In other words, using the Decision Logic approach, SDI+ item-level constellation could be constructed which predict “success” in not only job titles, but also in each of the job types determined by the occupational analysis.

Why Item Response Theory (IRT) Is No Longer the Best Alternative

The Big Five Factor model of personality has already established there are five relatively independent dimensions. Any “total score” approach will degrade reliability estimates which include items from the other four dimensions. The SDI+ adds two additional dimensions to complete the “OCEAN ST” model. In other words, the IRT model cannot operationally enhance either the Air Force test development process (with both multi-dimensional test blueprints and non-test items driving promotion) nor in the projected use of personality constellations based on the Big Five.

IRT Remediation Suggestions

To be fair, to remediate the shortcomings of the IRT approach, there are always suggestions to treat each dimension separately and develop “sampling” and “bridging” algorithms to support a robust adaptive testing system. Suggestions typically build on a General Linear Model which will determine appropriate weights to explain the variance mandated by external criterion vector. If one is required to obtain an external criterion vector, then the model is evolving towards the Decision Theory approach in which the criterion, not the test instrument comes first.

For the IRT approach, the devil is in the operational details. The number of observations required to “calibrate” items for use in a 3-parameter model are onerous. When the “one dimension” assumption is accepted, a given number of observations is implied (the Air Force has used 2000 observations per item when it worked on standardized tests for use at the DoD level.) If one accepts that multiple dimensions are at work, then there is no economy of scale and that same number of INDEPENDENT observations is required for each item in a dimension. The Air Force also recognizes items which have non-trivial loadings on more than one dimension. While the Decision Logic approach would have no problem with such items, IRT would be forced into two separate calibrations which, in all probability, would look quite different. For Air Force Promotion Tests, larger proportions of the test are being replaced each year to keep up with field-determined changes in the mission. Practical reasons preclude the ability to calibrate items before each annual test. Moreover, for some Air Force job specialties, the total population at a given pay grade in a target job family may be only 50 to 100 service members and “calibration” would mean exposure to the intended test subjects themselves.

Operationalizing Personality-Based Person-Job-Match

The backbone of IRT is built on the test instrument. Each item is evaluated as to how contributes to the central focus of that backbone (i.e. reliability in a one-dimensional “success” model). Item calibration is build on the assumption that items are properly keyed (i.e., scored) and that correct responses indicate success and that all distracters equally indicate failure. This model cannot be remediated for use in measuring the value of various personality constellations in differing job settings due to the bi-polar nature of the measurement and the fact that there are multiple “right” answers. In this case, the criterion must come first. By collecting behavior on the SDI+ instrument at the same time that detailed job information and job satisfaction is being collected (i.e., service members are being classified into job types), the Decision Logic approach will provide a sound methodology for isolating which (perhaps multiple) constellations of self-reported personality traits are associated with the kind of attitude that will improve retention in the armed forces.

Headquarters Air Force Personnel Center

Integrity - Service - Excellence



Personality and Mission Effectiveness

**Mr. Johnny J. Weissmuller
Chief, Military Testing Research
HQ Air Force Personnel Center**

48th International Military Testing Association (IMTA) Conference, Kingston, Ontario, Canada

Symposium – Role of Personality in the Military Organization- 14:00 4 Oct 2006 – Limestone “C”



Outline of The Paper

- ***Introduction***
- **Part 1: Forty-Nine Years of Air Force Personality Research**
- **Part 2: Recognized Need for Personality Measures**
- **Part 3: The Coming Revolution in Psychometrics**



The Highlights – 1 of 2

- **The USAF has just completed a 5-Year Research Plan to Support the AF Strategic Personnel Plan**
- **As of 4:00 PM EDT on 30 Sept 2006, nearly \$2 Million of Fall-out money has been committed to**
 - 1. Job Performance Measurement,**
 - 2. Force Management Studies,**
 - 3. AFSC Structuring,**
 - 4. Corporate Memory Documentation, and**
 - 5. Revitalization of the Air Force Human Resources Research Databank**



The Highlights – 2 of 2

- **The USAF is just beginning the Strategic, Return-on-Investment Analysis for Personality Measures**
- **We have developed a Comprehensive Occupational Data Acquisition and Analysis Plan (CODA²P)**



Part 1: Forty-Nine Years of Air Force Personality Research

- **1957-2006**
- ***Tupes, E.C. & Christal, R.E. Stability of personality trait rating factors obtained under diverse conditions. Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Center, May 1958, Technical Note WADC-TN-58-61, ASTIA Document No. AD-151 041.***
- **The Air Force Approach – Review the Professional Literature and forge an operational system to effectively measure and deploy “best thinking.”**



Part 1: Forty-Nine Years of Air Force Personality Research

- ***Critical Incidents Technique to Identify Attributes Associated with Aircrew Combat Performance, Final Report F33615-91-D-0010-DO5, 15 Dec 1993***
- **1993-1995 Dr. Raymond E. Christal begins development of a Self-Description Inventory (SDI) to Capture the “Big Five” for Operational Air Force use**



Part 1: Forty-Nine Years of Air Force Personality Research

- **1995-2004 After Dr. Christal passed away, the SDI development continued through various research iterations**
- **The final Air Force form is known as the SDI+ or “OCEAN ST”**
- **Since August 2005, part of the Air Force Officer Qualifying Test (AFOQT) – 10,000 operationally completed forms on file**



“OCEAN ST” – The Self-Description Inventory – Plus 2

- **“O” – Openness to cultural differences**
- **“C” – Conscientiousness**
- **“E” – Extroversion**
- **“A” – Agreeableness**
- **“N” – Neuroticism (1-Emotional Stability)**



“OCEAN ST” – The Self-Description Inventory – Plus 2

- **“OCEAN” – Plus –**
- **“S” – Service Orientation (Service over Self)**
- **“T” – Team Orientation (Role Flexibility)**



Part 2: Currently Recognized “Need” for Non-Cognitive Measures

- **“Editorial” Article by Terry D. Stevens in the Air Force Times, 25 Sept 2006, Page 54**
- **Title: “*Personality test could save billions of recruiting dollars.*”**
- **Quote: “*According to Defense Department figures, only 70 percent of high school graduates make it through their first enlistment.*”**



Part 2: Currently Recognized “Need” for Non-Cognitive Measures

- “Editorial” Article by Terry D. Stevens in the Air Force Times, 25 Sept 2006, Page 54
- Quote: “*There is an inexpensive process available for enhancing the quality and reducing first-term separations of enlistees, but the Defense Department does use it.*”
- Quote: “*The Defense Department should require prospective recruits to complete a behavioral / objective assessment before final enlistment, in addition to the Armed Services Vocational Aptitude Battery.*”



Part 2: Currently Recognized “Need” for Non-Cognitive Measures

The Air Force Times article was citing DoD-sponsored work on Recruiters already completed by HumRRO

DoD is just completing a review of the effectiveness of the Armed Services Vocational Aptitude Battery and the potential benefit from “Non-Cognitive” Measures has been documented.



Part 3: The Coming Revolution in Psychometrics – 1 of 2

- **Background on the Air Force Promotion System**
- **Performance Testing**
- **Classification Testing Using Decision Theory**
- **Testing in a Multi-Dimensional World**



Part 3: The Coming Revolution in Psychometrics – 2 of 2

- **Where will the Decision Logic Approach Offer an Advantage?**
 - **Promotion Tests under WAPS**
 - **Personality Assessment Items**
- **Operationalizing Personality-Based Person-Job-Match**



Comprehensive Occupational Data Acquisition and Analysis Plan (CODA²P)

For Officers

- **Now Operational in the Air Force Officer Qualifying Test (AFOQT)**
- **Since August 2005, nearly 10,000 AFOQT (Form “S”) have been collected**
- **Represents the Officer Commissioning Pipeline**



Comprehensive Occupational Data Acquisition and Analysis Plan (CODA²P)

For Enlisted Personnel 1

- **Experimentally include SDI+ in Standard USAF Job Inventories (surveyed every Three (3) Years)**
- **Ensure that Job Satisfaction and Reenlistment Intent Items are used to establish Criterion “Levels”**



Comprehensive Occupational Data Acquisition and Analysis Plan (CODA²P)

For Enlisted Personnel 2

- **After Routine Job Clustering based on Task Performance...**
- **Identify “SDI+ Constellations” at the Job Type by Criterion Level**
- **Use Hierarchical Clustering of Constellations to Determine Effectiveness of “Big Five” in Explaining Job Satisfaction and / or Reenlistment Intent**

Headquarters Air Force Personnel Center



Integrity - Service - Excellence



Johnny's OCEAN Profile

- There are free on-line resources to obtain your own “Big Five” Description
- Collaborative Efforts to Identify “Normative” samples for counseling purposes
- Below is the link to one such site at Penn State:
- <http://www.personal.psu.edu/%7Ej5j/IPIP/ipipneo300.htm>
- *The full IPIP contains 1,699 items assembled by Dr. Lewis R. Goldberg. The URL for Dr. Goldberg's IPIP is <http://ipip.ori.org/> . The IPIP is in the public domain and its items can be freely downloaded from that site.*
- *The IPIP-NEO is not equivalent to the commercial inventory on which it is based, the NEO PI-R™, authored by Paul T. Costa, Jr. and Robert R. McCrae. The genuine NEO PI-R™ (240 items) is considered by many psychologists to be the best inventory for measuring traits within the Five Factor Model (FFM) of personality. The NEO PI-R™ is copyrighted by Psychological Assessment Resources (PAR) in Florida, and can only be ordered by professionals and used by permission. You can contact PAR at: 1-800-331-TEST, or <http://www.parinc.com>.*



Johnny's OCEAN Profile

- <http://ipip.ori.org/>
- International Personality Item Pool:
A Scientific Collaboratory* for the Development of Advanced Measures of Personality and Other Individual Differences
- ~ Mission Statement ~
- This IPIP Website is intended to provide rapid access to measures of individual differences, all in the public domain, to be developed conjointly among scientists worldwide. Later, the site may include raw data available for reanalysis; in addition, it should serve as a forum for the dissemination of psychometric ideas and research findings.
- (Finholt, T. A., & Olson, G. M. From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, January 1997; vol. 8, no. 1; pp. 28-36.)
- Contact [the webmaster](#) with comments about this website.



Johnny's OCEAN Profile - O

■ ***Openness to Experience***

■ Domain/Facet.....	Score	0-----10-----20-----30-----40-----50-----60-----70-----80-----90-----
■ OPENNESS TO EXPERIENCE.....	48	
■ ..Imagination.....	94	*****
■ ..Artistic Interests.....	3	***
■ ..Emotionality.....	22	*****
■ ..Adventurousness.....	66	*****
■ ..Intellect.....	91	*****
■ ..Liberalism.....	21	*****

- ***Your score on Openness to Experience is average, indicating you enjoy tradition but are willing to try new things. Your thinking is neither simple nor complex. To others you appear to be a well-educated person but not an intellectual.***



Johnny's OCEAN Profile - C

■ **Conscientiousness**

■	Domain/Facet.....	Score	0-----10-----20-----30-----40-----50-----60-----70-----80-----90-----
■	CONSCIENTIOUSNESS.....	90	*****
■	..Self-Efficacy.....	93	*****
■	..Orderliness.....	72	*****
■	..Dutifulness.....	69	*****
■	..Achievement-Str.....	96	*****
■	..Self-Discipline.....	73	*****
■	..Cautiousness.....	83	*****

- ***Your score on Conscientiousness is high. This means you set clear goals and pursue them with determination. People regard you as reliable and hard-working.***



Johnny's OCEAN Profile - E

■ Extraversion

■ Domain/Facet.....	Score	0-----10-----20-----30-----40-----50-----60-----70-----80-----90-----
■ EXTRAVERSION.....	55	*****
■ ..Friendliness.....	49	*****
■ ..Gregariousness....	27	*****
■ ..Assertiveness.....	86	*****
■ ..Activity Level...	99	*****
■ ..Excitement-Seeking.....	1 *	
■ ..Cheerfulness....	81	*****

■ ***Your score on Extraversion is average, indicating you are neither a subdued loner nor a jovial chatterbox. You enjoy time with others but also time alone.***



Johnny's OCEAN Profile - A

■ **Agreeableness**

■ ***Disagreeable people can make excellent scientists, critics, or soldiers.***

■	Domain/Facet.....	Score	0-----10-----20-----30-----40-----50-----60-----70-----80-----90-----
■	AGREEABLENESS.....	69	*****
■	..Trust.....	74	*****
■	..Morality.....	48	*****
■	..Altruism.....	46	*****
■	..Cooperation.....	84	*****
■	..Modesty.....	47	*****
■	..Sympathy.....	69	*****

■ ***Your high level of Agreeableness indicates a strong interest in others' needs and well-being. You are pleasant, sympathetic, and cooperative.***



Johnny's OCEAN Profile - N

■ **Neuroticism**

■ Domain/Facet.....	Score	0-----10-----20-----30-----40-----50-----60-----70-----80-----90-----
■ NEUROTICISM.....	10	*****
■ ..Anxiety.....	34	*****
■ ..Anger.....	5	****
■ ..Depression.....	23	*****
■ ..Self-Consciousness....	24	*****
■ ..Immoderation.....	7	*****
■ ..Vulnerability.....	16	*****

- ***Your score on Neuroticism is low, indicating that you are exceptionally calm, composed and unflappable. You do not react with intense emotions, even to situations that most people would describe as stressful.***