REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
22-08-2014	Final	12 July 2012 – 22 July 2014

4. TITLE AND SUBTITLE

Development of a Cyber/Information Technology Knowledge Test for Military Enlisted Technical Training Qualification

5a. CONTRACT NUMBER
In-House

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)

D. Matthew Trippe[1], Karen O. Moriority[1], Teresa L. Russell[1],Thomas R. Carretta[2,] and Adam S. Beatty[1]

5d. PROJECT NUMBER
5329

5e. TASK NUMBER
09

5f. WORK UNIT NUMBER
53290902

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
See back of this form.

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711 Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Supervisory Control and Cognition Branch
Wright-Patterson AFB OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHCI

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

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

13. SUPPLEMENTARY NOTES
88 ABW Cleared 08/22/2014; 88ABW-2014-3945.

14. ABSTRACT

An Armed Services Vocational Aptitude Battery (ASVAB) Review Panel with expertise in personnel selection, job classification, psychometrics, and cognitive psychology developed recommendations for changes to the military enlistment test. One recommendation was to develop a test of cyber/information and communications technology literacy to supplement current ASVAB content. This article summarizes a multi-phased Cyber Test development process: (a) a review of information/computer technology literacy definitions and measures, (b) development and pilot testing of a cyber knowledge measure, (c) validation of test scores against final school grades (FSGs) for selected technical training courses, (d) development of an operational reporting metric and subgroup norms, and (e) examination of construct validity. Results indicate the Cyber Test has predictive validity against technical training school grades and incremental validity comparable to the ASVAB technical knowledge tests when used with the Armed Forces Qualification Test (AFQT) verbal/math composite as a baseline.

15. SUBJECT TERMS
Cyber, information and communications technology, technical knowledge, selection and classification, Armed Services Vocational Aptitude Battery

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	SAR	19	Antonio Ayala

7. PERFORMING ORGANIZATIONS NAMES AND ADDRESSES:

D. Matthew Trippe[1]
Karen O. Moriarty[1]
Teresa L. Russell[1]
Adam S. Beatty[1]
Human Resources Research Organization
Alexandria, VA

Thomas R. Carretta[2]
Air Force Research Laboratory
Wright-Patterson AFB, OH

Military Psychology
2014, Vol. 26, No. 3, 182–198

0899-5605/14/$12.00 http://dx.doi.org/10.1037/mil0000042

Development of a Cyber/Information Technology Knowledge Test for Military Enlisted Technical Training Qualification

D. Matthew Trippe, Karen O. Moriarty, and Teresa L. Russell
Human Resources Research Organization, Alexandria, Virginia

Thomas R. Carretta
Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio

Adam S. Beatty
Human Resources Research Organization, Alexandria, Virginia

An Armed Services Vocational Aptitude Battery (ASVAB) Review Panel, with expertise in personnel selection, job classification, psychometrics, and cognitive psychology developed recommendations for changes to the military enlistment test battery. One recommendation was to develop and evaluate a test of cyber/information and communications technology literacy to supplement current ASVAB content. This article summarizes a multiphased Cyber Test development process: (a) a review of information/computer technology literacy definitions and measures, (b) development and pilot testing of a cyber knowledge measure, (c) validation of test scores against final school grades (FSGs) for selected technical training courses, (d) development of an operational reporting metric and subgroup norms, and (e) examination of construct validity. Results indicate the Cyber Test has predictive validity versus technical training school grades and incremental validity comparable to the ASVAB technical knowledge tests when used with the ASVAB Armed Forces Qualification Test (AFQT) verbal/math composite as a baseline.

Keywords: cyber, information and communications technology, technical knowledge, selection and classification, Armed Services Vocational Aptitude Battery

The use of computer and information technology (IT) is pervasive in modern society. It affects all aspects of everyday life including commerce, communications, finance, government, military, transportation, utilities, and others. While the increased use of computer and IT have contributed to greater efficiency and cost savings, it also has led to increased vulnerability (e.g., information security, malicious intent, and theft). Over the last decade, computer and network security and vulnerability issues have increased dramatically in importance. A National Academy of Science (National Research Council, 2002) report emphasized the importance of cyber security in the wake of 9/11.

In the military, computer and IT are integral to the concept of net-centric operations. The objective of net-centric operations is to leverage an information advantage enabled in part by IT, into a competitive advantage through the networking of geographically dispersed forces. A strong, effective IT network improves information sharing which enhances the quality of information and shared situational awareness. Shared situational aware-

D. Matthew Trippe, Karen O. Moriarty, and Teresa L. Russell, Human Resources Research Organization, Alexandria, Virginia; Thomas R. Carretta, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio; Adam S. Beatty, Human Resources Research Organization, Alexandria, Virginia.

The opinions expressed are those of the authors and are not necessarily those of the United States Government, the Department of Defense, or the United States Air Force.

Correspondence concerning this article should be addressed to D. Matthew Trippe, 10503 Timberwood Circle, Suite 101, Louisville, KY 40223. E-mail: mtrippe@humrro.org

ness, in turn, enhances collaboration and synchronization of activities, speed of command, and overall mission effectiveness.

In 2006, the U.S. Air Force announced that cyberspace would constitute a new mission domain and in 2010 the Department of Defense (DoD) announced the establishment of the U.S. Cyber Command (McMichael, 2010) that was tasked to coordinate offensive and defensive cyber-related activities. Competition among industry, the government, and military for high quality cyber/IT personnel is great (Gould, 2013).

Selecting the Right People for Military Cyber Training

In 2005, the Defense Manpower Data Center (DMDC) initiated an Armed Services Vocational Aptitude Battery (ASVAB)[1] review process at the request of accession policy (Office of the Assistant Secretary of Defense). Factors driving the review included concerns that current ASVAB content was dated and the perceived potential of new measures to increase its predictive validity and classification efficiency. An expert review panel was convened to consider the current status of the ASVAB program and make recommendations for improvements. To this end, the ASVAB Review Panel (ARP) was briefed at three meetings in 2005 by military personnel, technical and policy experts from the Services and DMDC. The briefings included information regarding test development (item specifications, development, and evaluation), current ASVAB use (psychometric properties, validity, and classification efficiency), supplemental measures used by the Services (ability, temperament), and job analysis methods (and their relations to test content). The ARP presented its findings (Drasgow, Embretson, Kyllonen, & Schmidt, 2006) in March, 2006 that included 22 recommendations grouped into five broad areas: (a) content specifications, (b) test development and administration, (c) content changes, (d) development of a standardized validation and performance database, and (e) English language proficiency and its effect on test scores.

Proposed content changes included the development and evaluation of measures of noncognitive characteristics, nonverbal reasoning, and information/communications technology literacy (ICTL). The Air Force took the lead on the development of a cyber/ICTL measure as Air Force leadership had identified cyberspace operations as a critical and major growth area. The ARP speculated that an updated technical knowledge test along the lines of the ASVAB Electronic Information test might improve predictive validity and classification efficiency. This recommendation is consistent with a 2006 report by the National Academy of Engineering and the National Research Council regarding technological literacy (Garmire & Pearson, 2006).

A series of studies was conducted with the goal of development and psychometric evaluation of a cyber/ICTL test. These were: (a) literature review of ICTL (hereafter referred to as cyber knowledge) definitions and measures, (b) development and pilot testing of a cyber knowledge measure, (c) validation of cyber knowledge test scores against final school grades for selected technical training courses, (d) development of subgroup norms, and (e) examination of construct validity. The following sections summarize each of these studies.

Concept Definition and Initial Test Development

Literature Review: Definitions and Measures

The DoD sponsored a literature review on cyber/IT measurement (Russell & Sellman, 2007, 2008b) where the specific objectives were to develop a working definition based on prior research and to identify and review existing tests. To arrive at a working definition of cyber/IT, taxonomies of information and computer literacy concepts developed by the National Research Council and others were reviewed and

[1] ASVAB tests include Arithmetic Reasoning (AR), Assembling Objects (AO), Auto and Shop Information (AS), Electronics Information (EI), General Science (GS), Math Knowledge (MK), Mechanical Comprehension (MC), Paragraph Comprehension (PC), and Word Knowledge (WK). The verbal tests (PC and WK) are combined into a verbal (VE) composite. VE and the math tests (AR and MK) are combined into the Armed Forces Qualification Test (AFQT) composite, which is used by all U.S. military Services for enlistment qualification. Each Service develops its own composites to qualify applicants for technical training.

compared. The resulting working definition contained seven common elements across the taxonomies: (a) using computers (basic), (b) communicating, (c) gathering information, (d) using information technology (IT) tools and resources, (e) using networks, (f) programming, and (g) taking the broad view. The literature search also revealed several cyber/IT measures. Russell and Sellman (2007) compared existing cyber/IT measures against the working definition of cyber/IT literacy and several technical criteria and concluded that none of them covered all aspects of the working definition. Nonetheless, several of the measures demonstrated useful testing approaches and unique item types.

The initial cyber/IT literacy working definition was based entirely on literature and existing definitions, primarily civilian in nature. Russell and Sellman (2007) recommended that the cyber/IT literacy requirements of military jobs be integrated with the working definition to focus content development.

Development and Pilot Testing of a Cyber Knowledge Test

Identification of knowledge, skills, abilities, and other characteristics for measurement. Once a working definition of cyber/IT literacy was developed, the next step was to create a taxonomy of knowledge, skills, abilities, and other characteristics (KSAOs) required for successful performance in cyber/IT occupations. The taxonomy was used to create Cyber Test (CT) content specifications. Activities included (a) a review and integration of existing taxonomies, (b) interviews with military cyber/IT subject matter experts (SMEs), and (c) an online survey of additional military IT SMEs to evaluate and modify the initial taxonomy.

Review and integration of existing taxonomies. Several sources were reviewed to identify a set of KSAOs for measurement. These included the National Workforce Center for Emerging Technologies Web site, which contains industry-derived skills standards for IT, knowledge base categories from an IT publication focused on IT managers (Computerworld.com), and occupational information (education and training plans) for cyber/IT-related career fields for the Air Force, Army, and Navy. The resulting taxonomy consisted of 79 specific knowledge statements organized into four broad areas: (a) networking and telecommunications, (b) computer operations, (c) security and compliance, and (d) software programming and Web design.

There were two main concerns with the original knowledge taxonomy. The first was that it was civilian-centric. The second was that it was not known if the KSAOs were entry-level or more appropriate for advanced positions. To address these issues, military cyber/IT SMEs were recruited to review and modify the taxonomy to make it more appropriate for qualifying military applicants for entry-level technical training.

Interviews with military SMEs. Seventy-two cyber/IT SMEs from the Air Force (31), Army (3), and Navy (38) were interviewed by phone or face-to-face in small groups about the initial taxonomy. It was explained that the objective was to develop an entry-level technical knowledge test that could be administered as a part of the ASVAB. SMEs were asked which knowledge statements were entry-level. They also were asked to add new knowledge statements they thought were important and to make wording changes as needed. Finally, SMEs were shown some examples of different types of test items and asked for ideas about potential test item types. The revised taxonomy, summarized in Table 1, consisted of 39 knowledge statements.

It became apparent during the SME interviews that they viewed basic abilities, particularly reasoning, as important for success in IT or cyber-related training. With this in mind, we reviewed two well-known individual differences taxonomies (Carroll, 1993; Fleishman, Costanza, & Marshall-Mies, 1999) and defined abilities thought to be important for IT and cyber-related occupations. Drafts of the abilities list were discussed with SMEs over the course of the interviews to determine occupational relevance. The final list of 12 abilities appears in Table 2.

Online SME survey. An online survey was administered to gather data from SMEs on the cyber/IT knowledge and abilities identified in the previously described steps. Thirteen Air Force and 37 Navy SMEs completed the survey, which had four parts. Part 1 collected participant background data. In Part 2, SMEs made judgments about each of the 39 statements in

Table 1
IT Career Clusters and Core Skills

Broad area	IT cluster	Example of specific knowledge statement
Networking and Communications	• Network communications and maintenance	• Knowledge of network protocols and standards
	• Telecommunications	• Knowledge of telecommunications topologies
Computer Operations	• PC configuration and maintenance	• Knowledge of file structure
	• Using IT tools/software	• Knowledge of features and general uses of word processing software
Security and Compliance	• System security	• Knowledge of security methodologies for routing devices
	• Offensive methods	• Knowledge of encryption and decryption methods
Software Programming and Web Design	• Software programming	• Knowledge of basic language constructs
	• Database development and administration	• Knowledge of database querying methods
	• Web development	• Knowledge of web-based data environments
	• Data formats	• Understanding the differences between data formats
	• Numbering systems	• Understanding the different numbering systems such as hex and binary

the final knowledge taxonomy. They were asked to indicate whether the knowledge was basic or advanced, rate its importance, and the likelihood that the knowledge will change in the future. These three judgments were designed to help identify important, stable, basic knowledge areas that were good candidates for measurement on the Cyber Test. In Part 3, SMEs were asked to imagine that they were creating a test and to indicate how many items should be distributed across the four broad knowledge areas. Finally, in Part 4 SMEs rated the importance of the 12 abilities (see Table 2). The purpose of this part was to document the importance of

Table 2
Definitions of Abilities for Cyber/IT Occupations

Ability	Definition
Verbal reasoning	Ability to solve verbal/word problems by reasoning logically
Nonverbal reasoning	Ability to solve nonverbal problems (graphical, puzzles, and diagrammatic) by reasoning logically
Mathematical reasoning	Ability to reason mathematically and choose the right mathematical methods or formulas to solve a problem
Problem sensitivity	Ability to tell when something is wrong or is likely to go wrong. It does not involve solving the problem, only recognizing there is a problem.
Originality	Ability to come up with unusual or cleaver ideas about a given topic or situation or to develop creative ways to solve a problem.
Information ordering	Ability to arrange things or actions in a certain order or pattern according to a specific rule or set of rules (e.g., patterns of numbers, letters, words, pictures, mathematical operations)
Written communication	Ability to read and understand information and ideas presented in writing
Oral comprehension	Ability to listen to and understand information and ideas presented through spoken words and sentences
Perceptual speed	Ability to quickly and accurately compare similarities and differences among sets of letters, numbers, objects, pictures, or patterns
Advanced written comprehension	Ability to read and understand technical and/or government documents
Written expression	Ability to communicate information and ideas in writing so others will understand
Near vision	Ability to see details at close range (within a few feet of the observer)

abilities that might be measured by the CT or by other ASVAB tests.

SMEs considered all four broad cyber/IT knowledge areas to be important. They indicated that most of the items should focus on Networking and Telecommunications (29.4%), Computer Operations (28.3%), and Security and Compliance (27.0%), with less emphasis on Software Programming and Web Design (15.2%). SMEs considered most of the Computer Operations knowledge statements (74.0%) to be entry-level, with smaller percentages attributed to Networking and Telecommunications (48.0%), Security and Compliance (29.6%), and Software Programming and Web Design (30.6%).

SMEs rated nearly all of the 12 abilities as very important. The communications related abilities (Written Comprehension, Advanced Written Comprehension, Written Expression, and Oral Comprehension) held four of the top five ratings of importance. These results suggested that IT and cyber-related jobs are very cognitively demanding and that it may be useful to expand the coverage of communications skills in predictors of cyber related occupations.

Development of an initial experimental item pool. Once the KSAOs to be measured were defined, attention turned to identifying item types and measurement methods. Although several item types were considered, including information/knowledge, logic-based reasoning, situational judgment, nonverbal reasoning, scenario/stimulus-based, and biographical data, there were three important constraints. First, the items needed to have a format that would be consistent with other ASVAB items and capable of being administered on the CAT-ASVAB platform.[2] Second, the new test needed to be relatively short and efficient, ultimately about 20 min in length for the operational form. The first two constraints virtually dictated a selected response test. The third constraint was that the new test needed to provide incremental validity beyond that provided by the ASVAB. There is a wealth of evidence that the ASVAB is a good measure of cognitive aptitude for a number of constructs such as mathematical and verbal aptitude. This meant that the new test needed to focus on KSAOs not already tapped by the ASVAB.

Based on discussions with the cyber/IT SMEs, it was decided to focus the experimental item pool on information or knowledge, logic based reasoning, and biographical data items. Information tests were among the most successful and most highly valid printed classification tests created by the Army Air Forces (AAF) Aviation Psychology Program during World War II. Guilford and Lacey (1947) saw information tests as maximal performance interest measures. That is, information tests are thought to be indirect measures of interest, motivation, aptitude, and skill in a particular area. Moreover, they are not intended to certify an individual at a particular skill level or identify who does not need training. Rather, they are designed to assess knowledge and skill at a very general level whereas also providing an objective measure of interest and motivation in a technical content area. Knowledge or information tests continue to serve military selection and classification well today. The ASVAB General Science, Electronics Information, and Auto and Shop Information tests are all measures of technical knowledge or information in their respective content domains. For these reasons, technical knowledge items in the cyber/IT knowledge domain were expected to be good candidates for inclusion on the cyber/IT aptitude test. After decades of use, they have proven successful for use in military selection and classification (Oppler, Russell, Rosse, Keil, Meiman, & Welsh, 1997). We concluded that information or knowledge items were likely to be very useful predictors of performance in training for cyber-related jobs.

Logic-based reasoning (LBR) items assess inductive or deductive reasoning skills by presenting examinees with a premise or set of premises and asking them to choose the one valid conclusion among a series of conclusions (Colberg, Nester, & Trattner, 1985). Although LBR did not appear among the critical cyber/IT KSAOs, they were included because military cyber/IT SMEs indicated they believed reasoning ability to be an important determinant of job performance. We thought LBR items might be a useful way to assess reasoning skills needed for cyber-related jobs.

Deductive LBR items are essentially formal syllogisms placed in the scaffolding of a tradi-

[2] The CAT-ASVAB is a computerized adaptive testing platform for administering the ASVAB at the Military Entrance Processing Stations (MEPS).

tional verbal reasoning test item. Inductive LBR items are similar in structure, but rely on probabilistic rather than necessary premises and conclusions. The LBR items were expected to show a small amount of incremental validity when used in combination with the ASVAB as both assess general mental ability (*g*) (Stauffer, Ree, & Carretta, 1996).

Biodata items (Stokes, Mumford, & Owens, 1994) are based on the notion that the best indicator of future performance is past performance (Wernimont & Cambpell, 1968). Such items assess biographical information relevant to job performance. Past research has indicated that well-constructed biodata measures can exhibit good levels of criterion-related validity (e.g., Carlson, Scullen, Schmidt, Rothstein, & Erwin, 1999; Rothstein, Schmidt, Erwin, Owens, & Sparks, 1990; Schmidt & Hunter, 1998) and small subgroup differences (e.g., Reilly & Chao, 1982). The main drawback with biodata items is that they could be subject to response distortion when applicants are seeking highly valued occupations. Further, biographical data measures have been shown to demonstrate little incremental validity for predicting training and job performance when used in combination with measures of general mental ability (Schmidt & Hunter, 1998). Even so, we chose to develop biodata items because they are efficient to administer, inexpensive to develop, and offer a very different methodology.

Finally, SMEs had emphasized the importance of reasoning ability. Many talked about the ability to solve puzzles like Sudoku as occupationally relevant. To evaluate nonverbal reasoning ability, we administered a Figural Reasoning (FR) assessment. FR was previously used in the Army's Project A (Russell, Peterson, Rosse, Hatten, McHenry, & Houston, 2001) and the Enhanced Computer Adaptive Test (ECAT) project (Alderton, Wolfe, & Larson, 1997). Administering a nonverbal reasoning test would allow us to estimate how well such a measure would work for cyber/IT jobs. As with the LBR items, the FR test as a measure of nonverbal reasoning was expected to demonstrate a small amount of incremental validity when used in combination with the ASVAB, as both measure *g* (Stauffer et al., 1996).

Although a traditional multiple choice format was used for most of the information or knowledge and LBR items (75%), some were developed using nontraditional formats (e.g., multiple response, matching). The main advantages of nontraditional items are that they add face validity and variety for examinees, and are expected to result in less guessing. However, it was recognized that these item formats would be difficult to integrate into the CAT-ASVAB system.[3]

The number of items developed by knowledge area was based on discussions with SMEs about which content areas best reflected entry-level training requirements. The initial item pool had 219 items: 162 knowledge/information, 43 logic, and 14 biodata items[4] (Russell & Sellman, 2008a). Following the technical and sensitivity reviews, several items were edited and those thought to be too difficult were replaced with easier items. The final item pool consisted of 206 items: 148 knowledge or information, 44 logic, and 14 biodata items. Figural Reasoning was not included in the pilot test stage because it had been through rigorous development and review in the Army's Project A (Russell et al., 2001) and the ECAT project (Alderton et al., 1997).

Pilot test procedures, data processing, and sample demographics. Four forms of the CT were developed to minimize the effects of fatigue and item order on psychometric results. Each version included all of the items, but the items were presented in different orders.

The pilot test sample consisted of 684 examinees from two groups: 586 U.S. Air Force Basic Recruits at Lackland AFB, TX, and 98 U.S. Navy trainees attending a Cryptologic Technician Networks (CTN) course at Pensacola, FL. The USAF sample contained a higher proportion of women than did the Navy sample (29.7% vs. 18.4%). The two groups were similar in race or ethnic representation with about 79% White and 89% non-Hispanic in each

[3] Nontraditional formats such as multiple response or matching may violate assumptions of the item response theory (IRT) model used in CAT-ASVAB, such as local independence. Polytomous scored items also present a challenge for integration with CAT-ASVAB that uses only dichotomously scored items using the three parameter logistic model (3PL; Lord & Novick, 1968). IRT models appropriate for polytomously scored items (e.g., Muraki, 1997) are available, and mixing of models is not problematic within the IRT framework. Nevertheless, the current CAT-ASVAB infrastructure is configured to work with the 3PL model only, and revising it to include other models would require substantial changes to the current system.

[4] The 14 biodata items were multiple response format, representing 79 discrete items.

group. Although the U.S. Army showed some interest in the cyber/IT test, other research priorities precluded their involvement in the pilot study.

Biodata results. Examination of responses to biodata items indicated that both the USAF Basic Recruits and Navy CTN trainees used computers and information technology in their daily lives. Common activities included instant messaging, playing Internet games, participating in virtual environments, and downloading or listening to podcasts. They also were knowledgeable about computer operations (e.g., set up wired or wireless home network, set up/install/upgrade operating system on a home PC, and scan for/remove viruses). The Navy CTN sample was more experienced than the USAF Basic recruits on technical computer network tasks (e.g., set up wired and nonwired networks) and on computer programming languages. This was not surprising, as the USAF Basics represented a cross-section of technical training specialties, while the Navy CTN trainees were already assigned to an IT-related training course.

Knowledge and logic results. A major objective of this project was to use the pilot test data to evaluate test items and assemble alternate test forms containing a subset of the items. We began by screening all 192 cognitive items (148 knowledge and 44 logic) using Classical Test Theory (CTT) based item statistics. Items were flagged based on proportional *p* values and item-total correlations. Items with proportional *p* values greater than .80 were flagged as "easy" and those with values less than .20 as "hard." Items with item-total correlations less than .20 were flagged as "weak." Although this information was used in the decision process, items were not necessarily removed because they were too easy or hard or had a low item-total correlation. Ninety-eight items (72 multiple choice and 26 nontraditional) survived this initial screening process.

Test items then were evaluated based on their psychometric characteristics and content. Three pre-equated knowledge test forms and three pre-equated logic test forms were assembled to be parallel with respect to item discriminability, difficulty, and content. Internal consistency reliabilities ranged from .62 to .79 across the forms and samples. Values of this magnitude were not unexpected, given the range of content and the relatively small number of items.

Sex and racial group mean score differences in performance favored males and Whites. Male-female mean score differences were generally small (d = .27 to .40) by Cohen's (1988) guidelines. Although White–Black mean score differences were large (d = .93 to .98), they were consistent with those observed in other aptitude measures (Gottfredson, 2002; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmidt & Hunter, 1998) and for the ASVAB tests (Russell, Reynolds, & Campbell, 1994).

Correlations between the CT knowledge and logic forms and ASVAB scores were examined to explore relations between the tests. Analyses also included the Armed Forces Qualification Test (AFQT), a composite of the four ASVAB verbal and math tests (Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, and Mathematics Knowledge). The AFQT is used by all U.S. military services for enlistment qualification and is an indicator of *g*. Correlations were corrected for multivariate range restriction (Lawley, 1943) because of prior selection on the ASVAB. The 1997 national profile of American youth (PAY97; Segall, 2004) served as the reference population for this correction. After correction, the CT knowledge and logic test forms had moderate correlations with the ASVAB tests. Corrected correlations ranged from .55 to .77 between the AFQT and CT knowledge forms and from .53 to .81 between the AFQT and CT logic forms. Among the ASVAB technical knowledge tests, CT scores correlated most strongly with General Science (.56 to .71 for knowledge forms, .44 to .64 for logic forms).

Correlations between average CT knowledge and logic scores and biodata items revealed several moderate relationships. The strongest relationships occurred between those who claimed to have experience working with computer hardware (e.g., ordered computer parts, read manufacturer specifications, and built or repaired computers) and cyber knowledge. Biodata items generally had weak relationships with ASVAB scores, with the exception of the Electronics Information (EI) test. The EI test had moderate relationships with many of the same items to which the CT was related.

Technical Training School Validation

Predictive Validity Versus Final School Grades

Russell and Sellman (2010) examined the predictive validity of the CT knowledge, logic, and biodata measures against technical training grades. Six Air Force technical training courses and two Navy "A" courses were included in the study. All Air Force occupations were cyber/IT-related and drawn from intelligence and communications-computer functional communities. Nearly all of the Air Force occupations have since been reclassified with new occupational titles and specialty codes, but represent substantial coverage of what are now considered cyber warrior occupations (Scott, Conley, Mesic, O'Connell, & Medlin, 2010). See Table 3 for a list of courses.

The predictor battery consisted of the CT, a biodata measure, and Figural Reasoning. The tests were administered to students at the beginning of technical training. Final school grades (FSGs) were collected at the end of training to serve as criteria for validating the measures. In total, 1,127 students had both predictor data and FSGs.

Table 3 summarizes the observed validities for the predictors. Validity coefficients are summarized across occupations at the bottom of the table with sample size weighted means. For comparison purposes, Table 3 includes the ASVAB EI test. EI was a part of the selection composite for several of the cyber-related jobs. The AFQT had the highest weighted mean validity, (.41) followed by the CT (.37), FR (.25), EI (.22), and biodata (.19). The CT predicted FSGs significantly for all but one of the occupations (Network Intelligence Analyst – 1N4 × 1). Results suggested that the CT measure was a better predictor than EI that is currently part of composites used to qualify military applicants for many of the cyber/IT occupations.

Table 4 summarizes the validities of the predictors after multivariate correction for range restriction (Lawley, 1943) to the military enlisted applicant sample. All validities increased in magnitude after correction. The AFQT (.73) and CT (.64) had the highest mean validities for the eight courses.

CT Incremental Validity Versus Final School Grades

The AFQT was used as a baseline (observed $r = .41$) by which to evaluate the incremental validity of the other measures for predicting FSGs. The CT showed a small amount of incremental validity when used in combination with the AFQT and compared favorably with the other measures (EI, FR, and biodata). For the observed correlations, the weighted mean incremental validities for the eight courses were: CT (.051), EI (.031), FR (.012), and biodata (.008). After correction for multivariate range restriction on the ASVAB, the weighted mean incremental validities for the eight courses were: CT (.022), EI (.016), FR (.006), and biodata (.006).

Table 3
Observed Validity Estimates by Course

Service/course	*N*	Correlation with final school grade				
		AFQT	EI	Biodata	CT	FR
Air Force						
1N4 × 1–Network Intelligence Analysis	79	.25*	.26*	.16	.15*	.04
2E1 × 1–Satellite Wideband Telemetry	138	.37**	.24**	.13	.34**	.27**
2E1 × 3–Ground Radio Communication	170	.54**	.12	.10	.43**	.31**
2E2 × 1–Communications, Network, Switch, and Crypto Systems	161	.33*	.34**	.03	.43**	.21**
3C0 × 1–Communications-Computer Systems Operations	188	.44**	.29**	.30**	.46**	.20**
3C2 × 1–Communications-Computer Systems Controller	147	.47**	.18*	.27**	.35**	.23**
Navy						
Information Systems Technician (IT)	183	.37**	.21**	.15*	.31**	.17
Crypotologic Technician–Networks (CTN)	61	.35**	.07	.10	.34**	.22
Weighted mean	1,126	.41	.22	.19	.37	.25

* $p < .05$. ** $p < .01$.

Table 4
Validity Estimates Corrected for Multivariate Range Restriction by Course

Service/course	*N*	Correlation with final school grade				
		AFQT	EI	Biodata	CT	FR
Air Force						
1N4 × 1–Network Intelligence Analysis	79	.61	.48	.36	.46	.14
2E1 × 1–Satellite Wideband Telemetry	138	.72	.58	.36	.66	.44
2E1 × 3–Ground Radio Communication	170	.82	.59	.27	.77	.53
2E2 × 1–Communications, Network, Switch, and Crypto Systems	161	.73	.68	.31	.74	.40
3C0 × 1–Communications-Computer Systems Operations	188	.73	.53	.48	.69	.49
3C2 × 1–Communications-Computer Systems Controller	147	.65	.38	.35	.48	.42
Navy						
Information Systems Technician (IT)	183	.76	.52	.50	.61	.47
Crypotologic Technician–Networks (CTN)	61	.69	.28	.14	.53	.54
Weighted mean	1,127	.73	.53	.37	.64	.45

A couple of issues should be kept in mind when evaluating the incremental validities. First, the incremental validity analyses do not reflect the way the ASVAB is used operationally. Incremental validity analyses address how much additional prediction the new test would provide if the AFQT were used optimally (i.e., as a top-down selection tool, not as a dichotomized score). Because the AFQT is not used optimally, the incremental validity estimates are conservative and may underestimate the actual selection efficiency of the CT and other measures.

Regardless, incremental validity is an index the Services have used to evaluate new predictors for many years. It should be noted that the estimates reported here are similar to those for the ASVAB technical knowledge tests (General Science, Mechanical Comprehension, EI, and Auto and Shop Information) that are independent of the AFQT. For example, Oppler et al. (1997) reported incremental validities from a Joint-Service study that included 13 technical training courses. Validities for each ASVAB test were computed using only the training courses that included that test in their composites. Average incremental validity estimates beyond the AFQT, after correction for multivariate range restriction, ranged from .012 for EI to .034 for Auto and Shop Information.

Finally, it is important to note that military research suggests that even small validity increments (e.g., .02) can have utility in large selection programs (Held, Fedak, Crookenden, & Blanco, 2002; Schmidt, Dunn, & Hunter, 1995).

Additional Navy Training School Validation

Near the conclusion of the original training school validation study, the Navy significantly altered training in the Crypotologic Technician–Networks (CTN) course. Therefore, the Navy wanted to know whether the CT was a significant predictor of performance in the new course format. An additional sample of 118 CTN trainees completed the CT predictor battery during their first week of training in the revised course format. Two criterion variables were available for the validation analyses—grade point average (GPA) and graduation status (pass/fail). GPA was the average score computed from 19 course modules. Only individuals who ultimately passed the course had a reported final GPA. Nevertheless, course module scores were available for nongraduates up to the point of failure. That is, students continued in the course until they scored below 70% on a module. At that point the academic review board determined if the student should be dropped from the class. Students had to maintain a course average of 75% or higher and pass all module tests by scoring 70% or better. Because validating the CT against GPAs of only the successful candidates would restrict the variance in criterion scores, we imputed GPAs for all students using the average of the course module scores available.

Table 5 contains multiple correlation values that resulted from the regression of GPA on existing ASVAB composite predictors and the CT. It should be noted that Navy personnel can

Table 5
Validity and Incremental Validity of CT in Predicting GPA

	Observed				Corrected			
	Reported GPA		Imputed GPA		Reported GPA		Imputed GPA	
	$n = 76$		$n = 118$		$n = 76$		$n = 118$	
Predictor(s)	R	ΔR	R	ΔR	ρ	Δρ	ρ	Δρ
Cyber Test (CT)	.39**	—	.46**	—	.66	—	.69	—
AFQT	.41**		.40**		.80		.79	
AFQT + CT	.49**	.08**	.52**	.12**	.81	.01	.82	.02
Composite 1[a]	.45**		.44**		.82		.81	
Comp1 + CT	.52**	.07**	.55**	.11**	.84	.02	.84	.03
Composite 2[b]	.44**		.45**		.81		.81	
Comp2 + CT	.50**	.06**	.54**	.09**	.82	.01	.82	.02

Note. ρ indicates coefficients that were corrected for multivariate range restriction (Lawley, 1943).
[a] AR + 2*MK + GS. [b] VE + AR + MK + MC.
** $p < .01$.

qualify for CTN training on either of two composites (Composite 1 = AR + 2*MK + GS; Composite 2 = VE + AR + MK + MC). As a result, we examined the incremental validity of the CT against the AFQT and each of the Navy CTN composites. All incremental gains in the observed multiple correlation values were statistically significant at the .01 level. Values corrected for multivariate range restriction were more modest than the observed values. This may be due in part to the fact that the ASVAB variances were adjusted directly to the population values (that tends to result in a larger adjustment) whereas the CT variances were indirectly adjusted.

The second training criterion was graduation status (pass/fail). Results of the logistic regression analysis in which graduation status was regressed on the CT score alone and the CT score in combination with the ASVAB composite predictors are found in Table 6. The table includes Nagelkerke's (1991) adjusted coefficient of determination as well as the χ^2 value for each model. The χ^2 values for all models were statistically significant at the .01 level. The increment in χ^2 associated with the CT added to a model including the AFQT or Composite 1 were both statistically significant at the .05 level. The increment in chi-square associated with the model adding the CT to Composite 2 was not statistically significant ($\chi^2_{(crit)} = 3.84$, $p < .05$). Results indicated that the CT had significant value as a predictor of performance in the CTN course and provided incremental prediction over two of the three ASVAB composites.

Testing of Military Applicants at Military Entrance Processing Stations. Subsequent studies involved data collection on military applicants tested at the Military Entrance Processing Stations (MEPS). The objectives of these studies were to (a) estimate psychometric properties of the CT items in an applicant sample, (b) finalize two operational forms, (c) develop norms in military applicant samples, (d) further examine the relations between the CT and ASVAB tests, and (e) initiate longitudinal predictive validation studies.

To prepare for MEPS testing, new CT items were generated and four forms of the test were developed. In addition to entirely new items being written, some previous nontraditional format items were converted to the multiple choice format. Biodata items were eliminated from all

Table 6
Logistic Regression Results

Predictor	$R^2_{(Nag)}$	$\Delta R^2_{(Nag)}$	χ^2	$\Delta\chi^2$
Cyber Test (CT)	.11	—	9.84	—
AFQT	.10	—	8.95	—
AFQT + CT	.15	.05	13.78	4.82
Composite 1[a]	.15	—	13.76	—
Comp1 + CT	.20	.05	18.55	4.79
Composite 2[b]	.16	—	14.49	—
Comp2 + CT	.19	.03	17.77	3.29

Note. $n = 76$ graduates and $n = 42$ nongraduates.
[a] AR + 2*MK + GS. [b] VE + AR + MK + MC.

CT forms over concerns of potential response distortion as well as generally low predictive validity. Each CT form included 26 anchor items and 14 unique items and used the same content specifications. This was done to produce tests of similar length to the current ASVAB technical knowledge tests and to collect item-level data on a large set of items. The test plan was to administer the four CT forms to a combined sample of 50,000 Air Force, Army, and Navy applicants. The large sample sizes were needed to enable subgroup analyses on the four CT forms.

One of the objectives of the MEPS administration was to refine the available item pool based on a large-scale applicant administration. The majority of items that comprised the forms administered at the MEPS had been pilot-tested on relatively smaller samples of Air Force or Navy recruits who had already passed several selection hurdles. We expected to remove some of the items from the pool for psychometric reasons (e.g., inappropriate difficulty level, low item-total score correlations, and large subgroup differences) or because of flaws in experimental items that would only be revealed after pilot testing. Four items were removed from the pool because post hoc SME review of the item content in the context of the psychometric information revealed item flaws such as misleading language or more than one response option that could be considered correct. Twelve items were removed because they did not perform well in the applicant population. That is, some items had low or negative item-total correlations, extremely high or low *p* values, or poorly calibrated item response theory (IRT) parameters (i.e., extreme or out of bounds values) despite the absence of any apparent flaw in the item content. Of those items removed from the item pool, the majority were removed for being too difficult in the applicant sample.

The final CT item pool was calibrated and analyzed using an IRT measurement model known as the Three Parameter Logistic Model (3PL) (Lord, 1980; Lord & Novick, 1968). In essence, IRT assumes that test item responses by examinees are the result of underlying levels of ability possessed by those individuals. IRT provides a seamless approach to a variety of test analysis, development, and reporting activities and is facilitated by fitting, or calibrating, statistical models to examinee responses. Application of these statistical models results in the simultaneous scaling of item difficulty and examinee (population) ability. Calibration was executed via the software program MULTILOG (Thissen, 2003).

Another goal of the MEPS administration was to construct two operational forms from the items that comprised the experimental forms and develop a reporting metric. The target length of the two operational forms was 30 items with no overlap. We began the form assembly effort with the 65-item pool retained from the 82 unique items administered to the applicant sample. The resulting forms needed to be balanced with respect to (a) item content, (b) item subcontent, (c) difficulty, (d) discrimination, (e) reliability, and (f) keyed responses. We also needed to consider item "enemies" (i.e., items that assess identical or highly similar content) when making form assignments. To determine the optimal assignment of items to forms to balance the competing test specifications, we utilized Automated Test Assembly (ATA; van der Linden, 2005). The final operational forms contained 29 items each, one short of the original goal of 30 items per form. The 29-item solution resulted in the best balance of content, difficulty, discrimination, and reliability across the two forms. The inclusion of additional items upset the balance at a cost that we felt was greater than any benefit achieved in reliability or information.

Subgroup Norms

Standardized mean difference comparisons were computed across five subgroups: males (n = 39,951), females (n = 11,859), non-Hispanic Blacks (n = 7,524), non-Hispanic Whites (n = 25,607), and Hispanic Whites (n = 5,251). These groups were chosen to be consistent with designations used by the ASVAB testing program (Defense Manpower Data Center, 2011). Results for the CT, several ASVAB tests, and the AFQT are found in Table 7. The CT had smaller standardized mean differences than the ASVAB technical knowledge tests in the male–female comparison. Male–female differences were larger in the CT than in Assem-

Table 7
Standardized Subgroup Mean Differences of the Cyber Test and ASVAB Technical Tests in Applicant Sample

Test	$d_{(\text{male–female})}$ [a]	$d_{(\text{White–Black})}$ [b]	$d_{(\text{White–Hispanic})}$ [c]
Cyber Test (CT)	0.44	0.55	0.36
Armed Forces Qualification Test (AFQT)	0.30	0.81	0.48
Assembling Objects (AO)	0.19	0.59	0.14
Auto and Shop (AS)	1.05	1.14	0.62
General Science (GS)	0.56	0.99	0.61
Electronics Information (EI)	0.83	1.00	0.60
Mechanical Comprehension (MC)	0.82	1.09	0.55

[a] Male (n = 39,951) vs. female (n = 11,859). [b] Non-Hispanic White (n = 25,607) vs. Non-Hispanic Black (n = 7,524). [c] Non-Hispanic White (n = 7,524) vs. Hispanic White (n = 5,251).

bling Objects[5] (AO) or the AFQT. Differences between non-Hispanic Whites and non-Hispanic Blacks were smaller for the CT than any of the other technical knowledge tests. Similarly, differences in the non-Hispanic White versus Hispanic White comparison were smaller in the CT than in any other test with the exception of AO.

Construct Validity

To evaluate the construct validity of the CT, we tested a series of Confirmatory Factor Analysis (CFA) models depicted in Figures 1 through 3 using the technical training school validation sample. We tested highly similar models in the MEPS applicant sample and obtained comparable results, but present the training school modeling results here because of the availability of multiple nonverbal reasoning variables in the training school sample. Model 1 is based on prior factor analytic work on the ASVAB (Kass, Mitchell, Grafton, & Wing, 1983) and serves as a benchmark or baseline with which to compare subsequent models including the CT. Observed variables in Model 1 included the nine ASVAB tests and the FR test administered with the CT. The four hypothesized latent variables in Model 1 were factors representing Quantitative (QUANT), Verbal (VERBAL), Technical Knowledge (TECH), and Non-Verbal Reasoning (NVR). Model 2 added the CT as an observed variable hypothesized to load on the technical factor. The CT is conceptually similar to the other technical knowledge tests (General Science, Auto-Shop, EI, and Mechanical Comprehension) in that it represents an information test designed to assess knowledge and aptitude in a technical domain. Model 3 was a revision to Model 2, in which the CT was hypothesized to load on both the Technical and Verbal factors. Kass et al. (1983) found the General Science test to load on both technical and verbal factors. The CT is similar to GS in that its reading requirements are relatively more demanding than for the quantitative or other technical knowledge tests.

Table 8 summarizes the fit indices for the models. The χ^2 value associated with each model was statistically significant, indicating poor model fit. However, the χ^2 test is not generally relied on as an index of overall model fit in models tested on samples larger than 200. CFI and TLI/NNFI values above .95, RMR values below .05, and RMSEA values below .08 are generally indicative of good model fit (Kenny, 2009). CFI, TLI/NNFI, and RMR values all suggested that Models 1–3 exhibited good fit. The RMSEA index suggested poor model fit. The higher than desirable RMSEA value was likely due in part to that index's sensitivity to the ratio of parameters to degrees of freedom. Given the complexity of Models 1–3, it is reasonable to conclude that their fit to the data are within the acceptable range. Models 2 and 3 are nested and thus, their relative fit can be directly compared via the change in χ^2 value. Model 3 fits

[5] The Assembling Objects test is a nonverbal reasoning test that requires examinees to determine how an object will appear when its parts are put together.

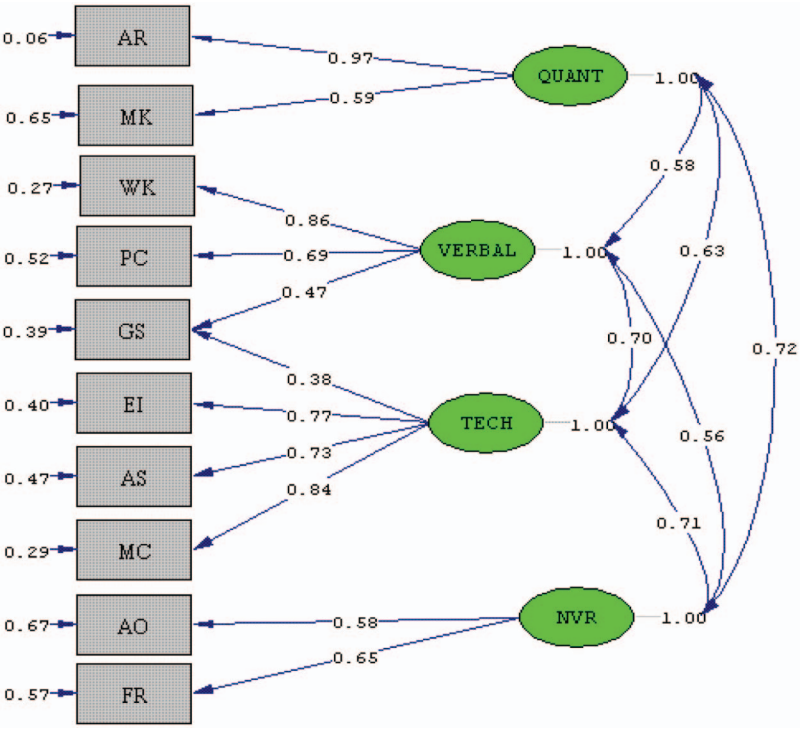

Figure 1. Confirmatory factor analysis Model 1. The tests were Arithmetic Reasoning (AR), Math Knowledge (MK), Word Knowledge (WK), Paragraph Comprehension (PC), General Science (GS), Electronics Information (EI), Auto and Shop Information (AS), Mechanical Comprehension (MC), Assembling Objects (AO), and Figural Reasoning (FR). The factors were Quantitative (QUANT), Verbal (VERBAL), Technical Knowledge (TECH), and Non-Verbal Reasoning (NVR). See the online article for the color version of this figure.

the data significantly better than Model 2 ($\chi^2 = 78.88$, $df = 1$, $p < .01$), suggesting that both the technical knowledge and verbal factors contributed significantly to the CT.

Discussion

Cyberspace is both an established and emerging national security front (Smart, 2011). As this fact becomes increasingly apparent as critical to national defense, we will undoubtedly observe a concomitant demand to select, classify, and train cyber warriors. Indeed, shortages of cyber security personnel are being reported in the military and federal agencies (Beidel & Magnuson, 2011). Although there is no single solution to address gaps in cyber knowledge and available cyber personnel within the Services, one way to address shortages and confront emerging threats is to begin identifying applicants most likely to succeed in cyber-related training. Expertise takes years to develop. The development of methods to assess suitability for cyber/IT career fields is only a first step.

Development and analysis of the CT is ongoing. The large-scale MEPS administration of the CT will serve as the foundation for the evaluation of new item pools and longitudinal validation studies. Nevertheless, the cumulative research to date has been sufficient to convince policymakers to begin preliminary operational use of the CT.

The current status of the CT is as a special test to be administered in static form on the CAT-ASVAB platform. In 2011, the Services formed a CT working group to address implementation issues. These include: (a) test maintenance (e.g., review of item specifications, development of expanded item pool, and

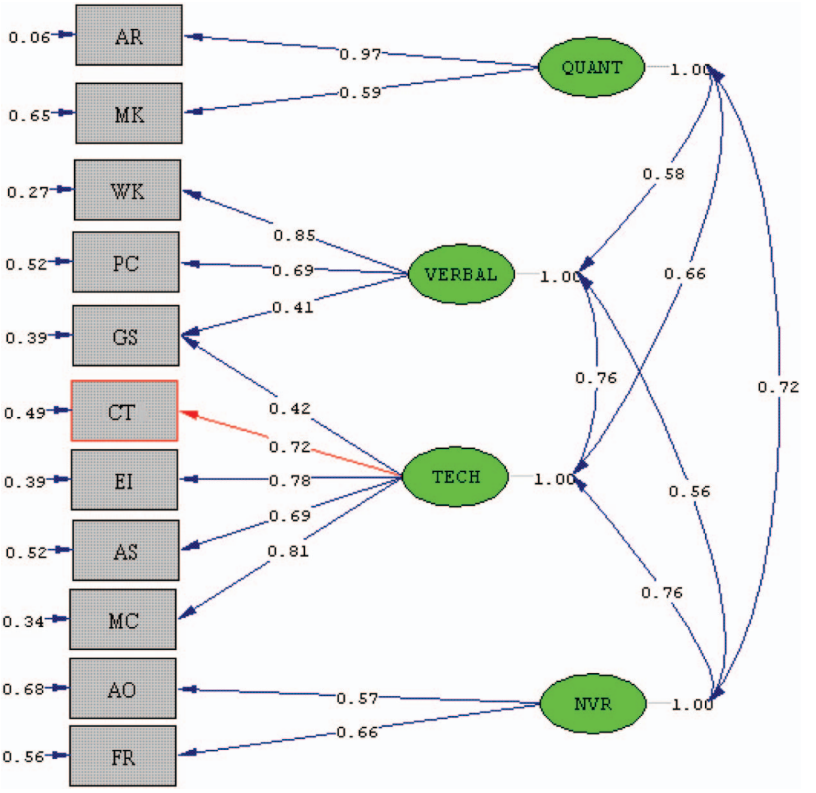

Figure 2. Confirmatory factor analysis Model 2. The tests were Arithmetic Reasoning (AR), Math Knowledge (MK), Word Knowledge (WK), Paragraph Comprehension (PC), General Science (GS), Cyber Test (CT), Electronics Information (EI), Auto and Shop Information (AS), Mechanical Comprehension (MC), Assembling Objects (AO), and Figural Reasoning (FR). The factors were Quantitative (QUANT), Verbal (VERBAL), Technical Knowledge (TECH), and Non-Verbal Reasoning (NVR). See the online article for the color version of this figure.

evaluation of item obsolescence), (b) identification of resources (funding, cyber/IT SMEs), (c) determination of frequency of planned updates, and (d) the development of Service ASVAB/CT composites.

The Services are currently ready to use the CT measure as a special test administered to a limited number of applicants who may expand the pool of available qualified applicants. In June 2014, the Air Force began operational use of the CT. Their model expands the qualified applicant pool for those who are five or fewer percentile points below existing cut scores for qualifying into cyber occupations. Those who score high enough on the CT (standard score $\geq$60) to compensate for missing the existing cut scores are added to the pool of qualified applicants. Additional work in the areas of composite formation and standard setting are underway. More specifically, we are examining the predictive validity of predictor composites that combine and weight the CT measure with other ASVAB tests, and measures of personality (Carretta & Manley, 2014) to achieve specific goals (e.g., maximize predictive validity, minimize adverse impact). We are also continuing to explore standard setting in the context of compensatory predictive models, like the one described above, such that cut scores optimize policy objectives (e.g., success in training, diversity).

The next phase of CT development will be to migrate the static test forms to an operational item pool suitable for computer adaptive testing (CAT). The existing item pool is relatively small in comparison to that of a CAT-ASVAB

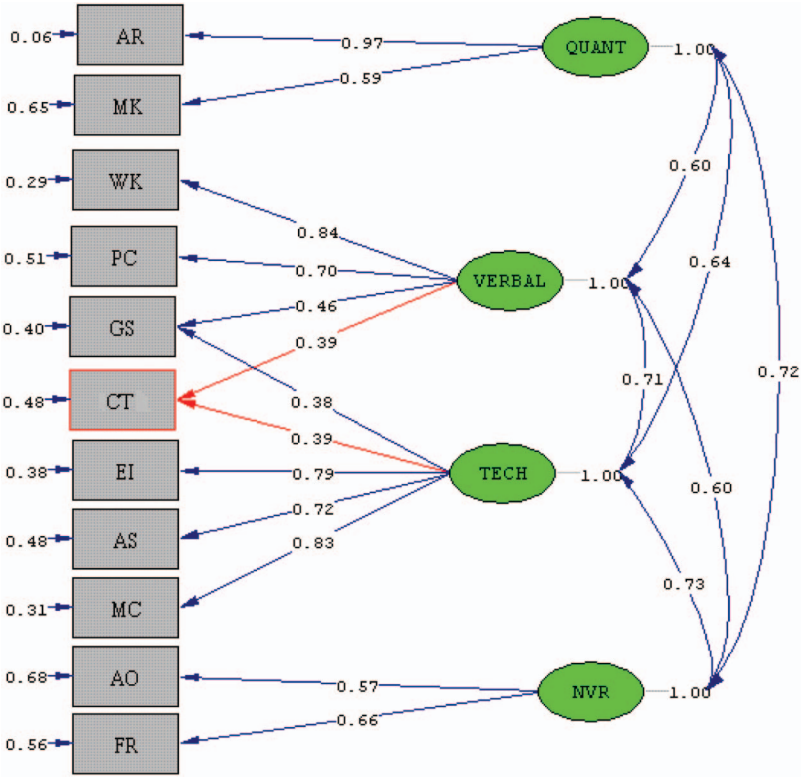

Figure 3. Confirmatory factor analysis Model 3. The tests were Arithmetic Reasoning (AR), Math Knowledge (MK), Word Knowledge (WK), Paragraph Comprehension (PC), General Science (GS), Cyber Test (CT), Electronics Information (EI), Auto and Shop Information (AS), Mechanical Comprehension (MC), Assembling Objects (AO), and Figural Reasoning (FR). The factors were Quantitative (QUANT), Verbal (VERBAL), Technical Knowledge (TECH), and Non-Verbal Reasoning (NVR). See the online article for the color version of this figure.

test and generally more subject to content obsolescence. Moreover, test "information"[6] is concentrated at the higher end of the ability distribution such that the test is relatively precise around the existing cut score, but relatively imprecise toward the middle and lower end of the ability distribution. Development efforts will focus on establishing a larger, contemporary item pool containing items that provide information along the entire continuum of ability. This kind of item pool is necessary to support CAT administration and to maintain proper item exposure controls for a test that is likely to be used increasingly for selection and classification.

Table 8
Fit Indices for CFA Models 1 Through 3

Model	*df*	χ^2	CFI	TL/NNFI	RMSEA	RMR
1	28	382.56	0.9635	0.9414	0.1046	0.04941
2	37	558.06	0.9582	0.9378	0.1118	0.05391
3	36	482.18	0.9642	0.9453	0.1031	0.04962

Note. Models 2 and 3 are nested. Sample size with complete data for all observed variables was 1,193.

[6] Test information is an index of measurement precision.

References

Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT battery. *Military Psychology, 9,* 5–37. doi:10.1207/s15327876mp0901_1

Beidel, E., & Magnuson, S. (2011, August). Government, military face severe shortage of cybersecurity experts. *National Defense*, 32–34.

Carlson, K. D., Scullen, S. E., Schmidt, F. L., Rothstein, H., & Erwin, F. (1999). Generalizable biographical data validity can be achieved without multi-organizational development and keying. *Personnel Psychology, 52,* 731–755. doi:10.1111/j.1744-6570.1999.tb00179.x

Carretta, T. R., & Manley, G. G. (2014). *Cyber/IT test development: Project overview*. Presentation to the Department of Defense Human Factors Engineering Technical Advisory Group (DoD HFE TAG), Aberdeen, MD.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511571312

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

Colberg, M., Nester, M. A., & Trattner, M. H. (1985). Convergence of the inductive and deductive models in the measurement of reasoning abilities. *Journal of Applied Psychology, 70,* 681–694. doi:10.1037/0021-9010.70.4.681

Defense Manpower Data Center. (2011). *ASVAB fairness information*. Retrieved from Official site of the ASVAB website: http://officialasvab.com/fairness_res.htm

Drasgow, F., Embretson, S. E., Kyllonen, P. C., & Schmitt, N. (2006). *Technical review of the Armed Services Vocational Aptitude Battery (ASVAB), FR No. 06–25*. Alexandria, VA: Human Resources Research Organization.

Fleishman, E. A., Costanza, D. P., & Marshall-Mies, J. (1999). Abilities. In N. G. Peterson, M. D. Mumford, W. C. Borman, P. R. Jeanneret, & E. A. Fleishman (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 49–69). Washington, DC: American Psychological Association.

Garmire, E., & Pearson, G. (Eds.). (2006). *Tech tally: Approaches to assessing technology literacy*. Washington, DC: National Academy Press.

Gottfredson, L. S. (2002). Where and why *g* matters: Not a mystery. *Human Performance, 15,* 25–46.

Gould, J. (2013, April 15). Cyber soldiers: Army set to load up 2 new enlisted MOSs. *Army Times, 1,* 8–19.

Held, J. D., Fedak, G. E., Crookenden, M. P., & Blanco, T. A. (2002). Test evaluation for augmenting the Armed Services Vocational Aptitude Battery. *Proceedings of the 44th Annual Conference of the International Military Testing Association* (pp. 291–297). Ottawa, Canada.

Kass, R. A., Mitchell, K. J., Grafton, F. C., & Wing, H. (1983). Factorial validity of the Armed Services Vocational Aptitude Battery (ASVAB), forms 8, 9, and 10: 1981 Army applicant sample. *Educational and Psychological Measurement, 43,* 1077–1087. doi:10.1177/001316448304300417

Kenny, D. A. (2009). *Measuring model fit*. Retrieved from David Kenny's website: http://davidakenny.net/cm/fit.htm

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62,* 28–30.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley.

McMichael, W. H. (2010). DoD cyber command officially online. *Air Force Times*. Retrieved from http://www.airforcetimes.com/news/2010/05/military_cyber_command_052110

Muraki, E. (1997). A generalized partial credit model. In W. van der Linden & R. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 153–164). New York, NY: Springer. doi:10.1007/978-1-4757-2691-6_9

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78,* 691–692. doi:10.1093/biomet/78.3.691

National Research Council. (2002). Cybersecurity today and tomorrow: Pay now or pay later. *National Academies Press*. Retrieved from www.nap.edu

Guilford, J. P., & Lacey, J. I. (Eds.). (1947). *Army Air Forces Aviation Psychology Program Research Reports: Printed classification tests, Report No. 5*. Washington, DC: U.S. Government Printing Office.

Oppler, S. H., Russell, T. L., Rosse, R. L. Keil, C. T., Meiman, E. P., & Welsh, J. R. (1997). *Item evaluation for the Armed Services Vocational Aptitude Battery (ASVAB) science and technical test specifications, TR No. 97–024*. Seaside, CA: Defense Manpower Data Center.

Reilly, R. A., & Chao, G. F. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35,* 1–62. doi:10.1111/j.1744-6570.1982.tb02184.x

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology, 75,* 175–184. doi:10.1037/0021-9010.75.2.175

Russell, T. L., Peterson, N. G., Rosse, R. L., Hatten, J. L. T., McHenry, J. J., & Houston, J. S. (2001). The measurement of cognitive, perceptual and psychomotor abilities. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel*

selection and classification (71–110). Mahwah, NJ: Erlbaum.

Russell, T. L., Reynolds, D. J., & Campbell, J. P. (1994). *Building a Joint-Service Classification Research Roadmap: Individual differences measurement, TP No. 1994–0009*. Brooks AFB, TX: Armstrong Laboratory.

Russell, T. L., & Sellman, W. S. (2007). *Review of information and communication technology literacy tests (FR No. 07–31)*. Alexandria, VA: Human Resources Research Organization.

Russell, T. L., & Sellman, W. S. (2008a). *Development and pilot testing of an information and communication technology literacy test for military enlistees: Vol. 1 Tech. Rep. No., FR No. 08–128*. Alexandria, VA: Human Resources Research Organization.

Russell, T. L., & Sellman, W. S. (2008b). Review of information and communication technology literacy tests. In T. L. Russell (Chair), *Measuring information and communication technology literacy*. Symposium conducted at the 23rd annual conference of the Society for Industrial and Organizational Psychology, San Francisco.

Russell, T. L., & Sellman, W. S. (2010). *Information and communications technology literacy test training school validation: Phase II final report, FR No. 09–89*. Alexandria, VA: Human Resources Research Organization.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High stakes testing in employment, credentialing and higher education: Prospects in a post affirmative action world. *American Psychologist, 56,* 302–318. doi:10.1037/0003-066X.56.4.302

Schmidt, F. L., Dunn, W. L., & Hunter, J. E. (1995). *Potential utility increases from adding new tests to the Armed Services Vocational Aptitude Battery (ASVAB), TN No. 95–5*. San Diego, CA: Navy Personnel Research and Development Center.

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,* 262–274. doi:10.1037/0033-2909.124.2.262

Scott, L. M., Conley, R. E., Mesic, R., O'Connell, E., & Medlin, D. D. (2010). *Human capital management for the USAF cyber force*. Santa Monica, CA: RAND.

Segall, D. O. (2004). *Development and evaluation of the 1997 ASVAB score scale, TR No. 2004–002*. Seaside, CA: Defense Manpower Data Center.

Smart, S. J. (2011). Joint targeting in cyberspace. *Air & space Power Journal, 25,* 65–75.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than *g*: An extension of Kyllonen's analyses. *The Journal of General Psychology, 123,* 193–205. doi:10.1080/00221309.1996.9921272

Stokes, G. S., Mumford, M. D., & Owens, W. A. (Eds.). (1994). *Biodata handbook*. Palo Alto, CA: CPP Books.

Thissen, D. (2003). *MULTILOG user's guide, version 7.03* [computer program]. Chicago, IL: Scientific Software International.

van der Linden, W. J. (2005). *Linear models for optimal test assembly*. New York, NY: Springer. doi:10.1007/0-387-29054-0

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples and criteria. *Journal of Applied Psychology, 52,* 372–376. doi:10.1037/h0026244