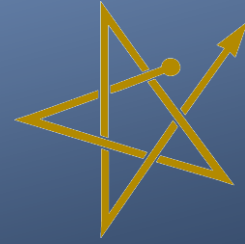# NPRST

Navy Personnel Research, Studies, and Technology
5720 Integrity Drive • Millington, Tennessee 38055-1000 • www.nprst.navy.mil

research at work

Evaluation of Tests of Processing Speed, Spatial Ability, and Working Memory for use in Military Occupational Classification

Janet D. Held
Navy Personnel Research, Studies, and Technology

Thomas R. Carretta
Air Force Research Laboratory

Evaluation of Tests of Processing Speed, Spatial Ability, and Working Memory for use in Military Occupational Classification

Janet D. Held
Navy Personnel Research, Studies, and Technology

Thomas R. Carretta
Air Force Research Laboratory

Reviewed by
Tanja Blackstone, Ph.D.

Approved and released by
D. M. Cashbaugh
Director

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

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)
2-12-2013	Technical Report	October 2012-May 2013

4. TITLE AND SUBTITLE

Evaluation of Tests of Perceptual Speed/Accuracy, Spatial Ability, and Working Memory for use in Military Occupational Classification

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROJECT ELEMENT NUMBER

6. AUTHORS

Janet D. Held and Thomas R. Carretta

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Navy Personnel Research, Studies, & Technology (NPRST/BUPERS-1)
Bureau of Naval Personnel
5720 Integrity Drive
Millington, TN 38055-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

NPRST-TR-14-1

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

10. SPONSOR/MONITOR'S ACRONYM(S)

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

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The Armed Services Vocational Aptitude Battery (ASVAB) is a joint-service battery used by the U.S. military services and the Coast Guard for enlistment selection and occupational classification. A subset of the ASVAB consisting of a composite of math and verbal tests is used for selection while various combinations of ASVAB tests, in composites formed to optimally predict training performance, are used for occupational qualification. The Navy has conducted extensive research to support the benefits of including a former ASVAB perceptual speed/accuracy test, Coding Speed (CS), as an adjunct ASVAB classification test as well as a working memory test, Mental Counters (MCt), shortly to be administered to both Navy and Air Force applicants. This technical report summarizes the supporting evidence for using CS and MCt in military classification and for use of the current ASVAB spatial ability test, Assembling Objects (AO), which is now only used by the Navy. The supporting evidence includes (1) incremental validity to the ASVAB in predicting training performance for a range of occupations, (2) lower adverse impact for females and some minority groups in qualifying for important occupations, and (3) increases in the proportion of recruit populations qualified for occupations in the aggregate.

15. SUBJECT TERMS

ASVAB, Coding Speed, Assembling Objects, Mental Counters, Improved Classification

16. SECURTIY CLASSIFICATION OF:			17. LIMITATATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCLASSIFIED			UNLIMITED	35	Wendy Douglas
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED			(901)874-2218

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Foreword

This technical report supports three cognitive tests for inclusion in the Armed Services Vocational Aptitude Battery (ASVAB) or as adjunct special occupational classification tests. The ASVAB is a joint-service cognitive test battery used by the U.S. military services and the Coast Guard for selection of their enlisted members and subsequent classification of those selected into military occupations. The content of the ASVAB is academic and technical knowledge based, with the exception of a spatial ability test, Assembling Objects (AO), which is only used by the Navy at the point for occupational classification. A former ASVAB test, Coding Speed (CS), also only used by the Navy, is a perceptual speed/accuracy test and is highly relevant in predicting performance for a number of occupations, including Air Traffic Controller. A third test, Mental Counters (MCt) is a Navy developed working memory test highly relevant for operations and multitasking types of occupations. This technical report provides empirical evidence on three dimensions for including the three tests in the military's occupational classification systems as they (1) increment the validity of the ASVAB in predicting training performance for a broad array of occupations, (2) reduce adverse impact defined as test score barriers for women and some minority groups, and (3) improve classification in terms of matching recruits to occupations and increasing the proportion of recruit populations occupation qualified.

This effort is supported by the Navy's Selection and Classification Office (N132G). The point of contact for this effort is Ms. Janet Held, Navy Personnel Research, Studies, and Technology, (901) 874-4650.

D. M. CASHBAUGH
Director

Summary

The Armed Services Vocational Aptitude Battery (ASVAB) is a joint-service cognitive test battery used by the U.S. military services and the Coast Guard for selection and classification of its enlisted members. The content of the ASVAB is mainly academic and technical knowledge based, considered measures of crystallized intelligence (gC), with the exception of a spatial ability test, Assembling Objects (AO). The AO test, which is only used by the Navy at this point for occupational classification measures non-verbal reasoning, an indicator of fluid intelligence (gF). A balance of both types of intelligence tests augments the ways in which recruits can be optimally assigned to occupations.

A former ASVAB test, Coding Speed (CS), is a perceptual speed/accuracy test also not dependent on academic or technical knowledge. The CS test is highly relevant in predicting performance for more than a few occupations that require attentional focus and multitasking, including Air Traffic Controller. The CS test has been revised by the ASVAB developer and executive agent, the Defense Manpower Data Center – Personnel Testing Division (DMDC-PTD), to be resistant to score impact from computer hardware changes, the primary reason the test was removed from the ASVAB. Similarly, a third test, Mental Counters (MCt), is a working memory test considered highly relevant for classification into operations and multitasking types of occupations. The MCt test, previously evaluated positively in the joint-service Enhanced Computer Administered Test (ECAT) battery project, is now being administered to Navy applicants on the computerized adaptive version of the ASVAB (CAT-ASVAB) in an evaluation phase.

This technical report provides evidence for including AO, CS, and MCt as either part of the ASVAB or as adjunct occupational classification tests recognizing that many occupations in the military have evolved over time to require personnel skills measured by these tests. The evidence is empirical and includes (1) increments to the validity of the ASVAB in predicting training grades for a broad array of occupations, (2) reduced adverse impact defined as test score barriers for women and some minority groups, and (3) improved classification in terms of matching recruits to occupations and increasing the proportion of recruit populations occupation qualified.

An additional benefit of the AO, CS, and MCt tests is that they are conducive to automated item generation, therefore greatly reducing or even eliminating item development costs that are traditionally associated with knowledge based tests. The Navy will be re-evaluating the three tests in ASVAB validation/standards studies in 2015 to reformulate optimal classification composites for many of its occupations.

Table of Contents

List of Figures

List of Tables

Background

Many U.S. military jobs have significantly increased technology use and multitasking requirements. On one hand, technology that automates job tasks reduces the need for deep learning of a particular system. For example, a ship's automated navigation system can reduce the need for Sailors to learn the underlying mathematics and navigation principles that were once required to guide ships. On the other hand, policy makers and the acquisition team might not have successfully addressed the Human Systems Integration (HSI) requirement that takes into account the capacity/limitations of individuals to successfully use the system, especially when other task requirements are added (Wickens, 2008). The perception of some not familiar with the underlying principles of HSI may be that automated systems allow individuals to take on many more job tasks without degradation of task performance, which may not be the case.

Weiner (1989) coined the term "clumsy automation" to denote the role awkward systems can play in contributing to human errors in technically complex jobs such as pilot or air traffic controller. Awkward human-system interfaces contribute to error by increasing rather than by decreasing the cognitive workload of human operators during periods when they are preoccupied with other tasks that require attention. Wickens (2008) provides a good discussion on "Workload" stressors and the adequacy of humans to deal with them given in terms of the nature of the tasks and individuals' available mental processes. Under the multiple resources model (Wickens, 1984), a strategy to reduce workload could involve changing the resource requirements for a task. For example, if two tasks each require the visual processing of information, it might be possible to use auditory rather than visual presentation for one of the tasks to reduce visual workload (auditory and visual perception use different mental resources).

Although the Workload/Resource area of study is complex, an accumulation of research points to working memory, attention, and perceptual speed as critical abilities necessary for efficient and accurate complex multitasking (Ackerman & Beier, 2007; König, Bühner, & Mürling, 2005; Oberlander, Oswald, Hambrick, & Jones, 2007). Such cognitive constructs are not currently measured as part of the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB is used by all of the U.S. military services to qualify military applicants for enlistment and to classify those enlisted to military occupations. With the exception of the ASVAB's spatial ability test [Assembling Objects (AO)], the battery content is academic and knowledge based, and therefore considered to largely measure crystallized intelligence (gC). As such, gC test performance may depend somewhat upon access to quality education, which may be linked to socioeconomic status.

In contrast to gC, spatial ability, perceptual (or processing) speed, and working memory are measures of fluid intelligence (gF) that are less culturally bound – that is, linked less to education and knowledge acquisition. This report describes past and ongoing Navy and DoD research that considers the inclusion of measures of gF to supplement the ASVAB thereby enhancing occupational classification.

As such, the report provides empirical evidence that supports the use of three tests, two of which are operational and administered at the Military Entrance Processing Stations (MEPS) on the computer platform that delivers the adaptive version of the ASVAB (CAT-ASVAB) and the other, a working memory test, in an evaluation phase. The two operational tests are presently a part of the Navy's classification system, and the working memory test is being considered. The three tests are (1) Assembling Objects (AO), an operational ASVAB spatial ability test, (2) Coding Speed (CS), a former ASVAB perceptual speed/accuracy test only used by the Navy in occupational classification, and (3) Mental Counters (MCt), the working memory test recently (June, 2013) deployed at the MEPS on the CAT-ASVAB system and administered only to Navy applicants for initial evaluation. Table 1 provides brief descriptions of the ASVAB and CS tests that are currently in operational use for Navy occupational classification.

Table 1
Description of the ASVAB and Coding Speed (CS) Tests

Test Name and Abbreviation	Test Description
General Science (GS)	Knowledge of physical and biological sciences
Arithmetic Reasoning (AR)	Ability to solve arithmetic word problems
Word Knowledge (WK)[a]	Ability to select the correct meaning of words presented in context and correct synonyms
Paragraph Comprehension (PC)[a]	Ability to obtain information from written passages
Mathematics Knowledge (MK)	Knowledge of high school mathematics principles
Electronics Information (EI)	Knowledge of electricity and electronics
Auto and Shop Information (AS)	Knowledge of automobile and shop technologies, tools, and practices
Mechanical Comprehension (MC)	Knowledge of mechanical and physical principles
Assembling Objects (AO)[b]	Ability to determine correct spatial forms from separate parts and connection points
Coding Speed (CS)[b]	Ability to quickly identify correct word/number pairings from a key with many options

[a]WK and PC are combined to form the Verbal (VE) composite that is a component of the AFQT and several Navy ASVAB classification composites. [b]Not all recruits enter the Navy with AO and CS test scores. CS is only given by computer at the MEPS at the end of the CAT-ASVAB. AO, also given on the CAT-ASVAB, is not given to high school students taking the paper and pencil version of the ASVAB under the Career Exploration Program, but is given in paper and pencil ASVAB forms in the Enlisted Testing Program.

WK and PC are combined to form the Verbal (VE) composite, with WK weighted approximately 2/3 and PC 1/3. The ASVAB individual tests, including CS, are scored on a standard score scale that was derived to have a mean of 50 and standard deviation of 10 developed for the ASVAB normative youth population in the 1997 Profile of American Youth study (PAY97; Segall, 2004).

Although only the Navy uses CS or AO in occupational classification and is projected for an early use of MCt, all of the military services may view these tests more favorably if the current stellar military recruiting environment deteriorates. The military has sustained a long running positive recruiting environment due mainly to (1) the nation's economic downturn resulting in a high unemployment rate and limited good job opportunities, (2) a sense of patriotism following the 9/11 attacks, and (3) the military's offer of a low cost quality education both while enlisted and post service.

If the recruiting environment deteriorates the AO, CS and MCt tests will provide expanded options for qualifying youth interested in joining the military and assigning them to occupations to which they might not otherwise be qualified. Further, as U.S. demographics change and with an expanded focus on the role of women in the military, these tests will become more important as tools to enable the "right" occupational assignments for all military enlisted members. The following sections describe the AO, CS and MCt tests and their benefits.

The Coding Speed (CS) Test

From 1976 to 2002, the ASVAB contained two speeded tests, Coding Speed (CS) and Numerical Operations (NO), that were useful in classifying military recruits to predominately clerical types of occupations. The two tests were eliminated from the battery because the scores were sensitive to changes in test format and item response input modes. For example, in the paper and pencil (P&P) format, NO and CS scores were impacted when the paper answer sheet with round bubbles was replaced with an answer sheet that had narrow rectangles. The rectangles took less time to fill in and given that the tests were timed; examinees had a greater number of correct responses than those taking the test with the bubble answer sheet. Score impact was also observed (mainly for NO) in a computer hardware study that varied computer features (e.g., CPU, screen size, and response input mode) (Segall, 1997).

The rationale for eliminating both NO and CS from the ASVAB was not that the tests were classification inefficient but rather that it would be impractical to conduct the studies required to equate CS scores now that CS is only administered by computer (not in P&P format). Obviously the life cycle of computer hardware ends at some point and new computers would be purchased. However, before NO and CS were eliminated, the Navy provided enough empirical evidence to support continued use of CS as a Navy special classification test (NO was not supported and CS was less susceptible to hardware changes). As a result of the evidence, and because the Navy, the Defense Manpower Data Center – Personnel Testing Division (DMDC-PTD), and others expended resources to make CS more robust to inadvertent hardware changes, CS was retained for the Navy (Abrahams, Walton-Paxton, Alf, Barton, Cole, & Kieckhaefer, 1996; Segall, 1997). One major improvement to CS was a new scoring method, the rate score, which is less sensitive to speed of responding (Segall, 1997). In 2004, DMDC-PTD scaled CS scores for four CS forms to the PAY97 ASVAB normative population score scale (Segall, 2004).

Score sensitivity for CS is again an issue as the CAT-ASVAB undergoes another major hardware change. The hardware change this time is the replacement of the special CAT-ASVAB keyboard configured with the A, B, C, D, E answer keys on one row with a mouse. The mouse, currently used by examinees to input their responses for all CAT-ASVAB tests, will be the only response input mode for the CS test. The CAT-ASVAB special keyboards are being replaced with standard commercial issue keyboards (with various keys used for testing purposes other than answering test questions).

DMDC-PTD in consideration of the Navy's use of the CS test has conducted a CS equating study involving the mouse and the special CAT-ASVAB keyboard; however, from the data analyses it appears that a special equating is not necessary – that is, the mouse response mode has not impacted CS scores (Pommerich, 2013). In the lead up to the equating study, because of long standing concerns about speeded tests and the anticipation that examinees using the mouse would have higher scores, DMDC-MPT developed a new version of CS, called Processing Speed (PS). The PS test was developed with the intended purpose of reducing or eliminating response input mode impact on scores by eliminating the time it takes from the examinee's realization of the correct response (or the chosen response) to the time it takes to input that response (Segall, 2010), which in a of itself could be an individual difference variable.

DMDC-PTD's new version of CS, called Processing Speed (PS), displays each item separately rather than as a block of seven items, but with the key containing 10 paired number/word answer choices positioned the same as CS, at the upper portion of the screen (Segall, 2010). Unlike CS with a generous screen display, the PS item display time is controlled to systematically shorten as the test progresses making PS a purer measure of processing speed than CS. The PS total test score is simply the total number of items correct – measuring the threshold of the examinee's ability to perform the simple but involved task. Because of the lenient screen viewing time allowed for CS, the new PS test with increasing time constraints may not completely measure the same constructs as CS (to be determined). Segal (2012) established that at least some part of the CS test measures intrinsic motivation that relates to salary (possibly a proxy for promotion) over time.

The Assembling Objects (AO) Test

Both the CS and AO tests were developed by the Army but at different times. CS has historical military roots that evolved from use in classifying Army recruits to clerical occupational whereas AO was developed later during the Army's Project A in search of new cognitive and non-cognitive military performance predictors (Buscigilo, Palmer, King, & Walker, 1994; Campbell & Zook, 1992; Russell & Peterson, 2001). One of the first steps in Project A was the identification of abilities and characteristics important for Army occupations that were not measured by the ASVAB. Spatial ability was identified as a key area. Several spatial constructs were identified (such as spatial relations, spatial orientation, and spatial visualization); 10 spatial tests were developed, six of which survived field testing and were included in validation studies (Russell, Peterson, Rosse, Hatten, McHenry, & Houston, 2001).

Factor analyses of the Army's Project A spatial tests indicated the presence of a general spatial factor and that reasoning and assembly type items were the best measures of this factor. Additional analyses revealed that there were small or no gender differences for the spatial tests, Assembling Objects (AO) and Figural Reasoning (FR) (Peterson, Russell, Hallam, Hough, Owens-Kurtz, Gialluca, & Kerwin, 1990). Further, in a study of the effects of practice and coaching on test performance, only small to moderate mean score improvements were observed for AO and FR (Buscigilo & Palmer, 1996). Both AO and FR were included in the DoD's ECAT (Alderton, Wolfe, & Larson, 1997) project. Analyses of the ECAT data showed that AO could increment the validity of the ASVAB for predicting job performance and improve classification of personnel into some military occupations (Sager, Peterson, Oppler, Rosse, & Walker, 1997; Wolfe, Alderton, Larson, Bloxom, & Wise, 1997). The AO test subsequently became an ASVAB test in 2002 when NO and CS were eliminated.

It should be noted that on a theoretical basis, the Navy, in their use of AO and CS, has not combined the two tests in the same classification composite. One reason is that the tests are considered measurements of separate constructs linked to different occupations. AO measures the ability to visually construct spatial forms from the forms' parts and also to identify connection points of form parts. On the face of it, these types of test items map well to tasks performed in mechanical occupations (Held, Fedak, & Johns, 2004). CS, on the other hand, requires quick and accurate thinking, which applies to many operations types of occupations in addition to clerical (e.g., Navy SEALs). The AO and CS tests, in more comprehensive analyses, will be evaluated in combination in the future across a wide variety of Navy occupations.

The second reason for not combining AO and CS in the same classification composite is logistical in that not all Navy applicants are administered both tests. For example, Navy applicants testing on the paper-and-pencil version of the ASVAB receive AO but do not receive CS. Further, those who take the ASVAB in the high school testing program (Career Exploration Program, currently administering ASVAB in paper-and-pencil) do not receive either AO or CS.

The third reason for not combining AO and CS in the same composite is that initial validity analyses with data available for both tests were not supportive. For example, in an ECAT study, Held and Wolfe (1997) added the "best" ASVAB test to the operational ASVAB classification composite that applied to an occupation and compared the incremental validity with that provided by the best ECAT test. The ECAT incremental validity results showed AO did not add to the ASVAB operational composite for the six occupations that used CS in their ASVAB composite (Held & Wolfe, 1997, p. 81).

The Navy wants to retain the ability to measure the underlying constructs that both CS and AO measure. For CS this may be a constellation of cognitive abilities, persevering on mundane but important tasks, and alertness and vigilance on short fused time/critical tasks. For AO this would also be cognitive abilities, but in the domain of spatial ability. Empirically, the Navy supported three reasons for retaining both CS and AO in its classification system in that these tests provide (1) non-trivial incremental validity in predicting training performance grades for many Navy occupations (ratings) when combined with specific ASVAB tests, (2) reduced gender and some minority group

score differences that effectively lowers barriers to occupations, and (3) increases in the proportion of annual Navy recruit populations qualified for Navy occupations in the aggregate. The following sections provide a summary of the evidence on each supporting reason for CS and AO followed by a section on the theoretical and initial empirical evidence that supports use of the ECAT working memory test, Mental Counters (MCt).

Coding Speed and Assembling Objects: Incremental Validity

Coding Speed

Table 2 lists ASVAB predictive validity coefficients for composites with and without the CS test, presented to the Manpower Accession Policy Working Group (MAPWG) and the Defense Advisory Committee-Military Personnel Testing (DAC-MPT) during the 1990s DoD wide assessment of the ASVAB speeded tests, CS and NO.

Table 2
Range Corrected Validities of ASVAB Composites
with and without Coding Speed

Rating	N	VE+AR+MK	AR+2MK+GS	VE+MK	VE+MK+CS	VE+MK+CS − VE+MK
Signalman	1,548	.56	.54	.56	**.59**	.03
Radioman	2,263	**.62**	.61	.61	**.62**	.01
Operations Specialist	1,676	**.74**	.73	.72	**.74**	.02
Dental Technician	516	.63	.61	.62	**.64**	.02
Personnelman	942	.62	**.63**	.60	**.63**	.03
Ship's Serviceman	801	**.49**	.48	.48	**.49**	.01
Storekeeper	2,201	.65	.64	.63	**.66**	.03
Aviation Maint/Admin	873	.72	.70	.71	**.74**	.03
Aviation Storekeeper	801	.63	.62	.63	**.65**	.02
Mess Management	2,589	**.65**	.63	.64	**.65**	.01

Notes. (1) Validities were corrected for multivariate range restriction; largest validity coefficients are in bold. (2) The VE+MK composite of ASVAB tests was judged a suitable replacement for the Services' administrative composites that contained CS (and in some cases that also contained NO). (3) Validity coefficients were developed on final school grades that pertained to each Navy rating's initial technical training course. (4) Validity coefficients were corrected for range restriction using the PAY80 normative population correlation matrix as the unrestricted population from which, theoretically, future recruits would be selected using the ASVAB composites and cutscores (the ASVAB standards). (5) Some of the ratings listed have since been consolidated, eliminated, or renamed as occupations have changed since the 1990s.

Table 2 shows range corrected validities using the multivariate formulas (Lawley, 1943) for a set of ASVAB composites that included the Navy's operational VE+MK+CS composite (Verbal, Mathematics Knowledge, and Coding Speed all unit weighted).[1] This "CS" composite had, on average, .02 higher validity than the VE+MK composite proposed as the replacement composite (at the time) for all military services that used an ASVAB composite with NO/CS for classification into administrative types of occupations. A .02 increment in predictive validity may seem small but in large scale testing programs such as the ASVAB, this level of increment can translate to a substantial reduction in training attrition and cost savings (Schmidt, Dunn, & Hunter, 1995). However, the .02 validity increment is not the total picture and a test's inclusion for military occupational classification must demonstrate other psychometric qualities such as validity standing alone, test stability (retest or parallel forms reliability), reduced adverse impact, and improved recruit classification outcomes without lowering ASVAB cutscores.

Although the NO and CS tests were most known for use in military classification for clerical types of occupations, the rating names in Table 2 obviously apply to a mix of clerical and non-clerical occupations (i.e., Signalman, Radioman, Operations Specialist, and Dental Technician). More recently, two other non-clerical ratings, Air Traffic Controller and SEALs, have adopted a composite using CS (VE+MK+MC+CS), which has been confirmed twice for each rating as having the highest criterion related (predictive) validity when compared to other rationally and empirically derived ASVAB composites (Held, 2006, 2011). In addition, an independent source confirmed the VE+MK+MC+CS composite for the Navy SEALs.[2]

Finally, in an Army study of many cognitive and non-cognitive measures, CS was shown in one optimally derived test battery equation to increase mean predicted performance (MPP) across the study's largest set of occupations (and across a number of performance outcomes) improving what the Army terms, Classification Efficiency (CE) (Scholarios, Johnson, & Zeidner, 1994). The CS test also improved CE as differential assignment capability (Scholarios et al., 1994), which simply means assigning individuals to jobs for which they exhibit the relevant aptitudes, skills and abilities, recognizing there are individual differences in these personnel attributes. In one experimental battery, Scholarios et al. found that CS was selected first based upon the differential assignment index. Differential assignment capability is discussed in a later section as improved classification, which means increasing the proportion of recruit populations occupation qualified without having to lower cutscores.

---

[1] Held and Foley (1994) demonstrate the multivariate range correction formulas in an ASVAB restriction in range simulation study where selection stringency was modeled applying a range of ASVAB composite cutscores.
[2] "Follow on Research Findings" submitted by the Gallup Consulting, Inc. in 2011 to Director, Naval Special Warfare Recruiting Directorate, NAVSPECWARCEN, San Diego, CA.

Assembling Objects

Table 3 lists the ASVAB predictive validity coefficients for composites with and without the AO test, which were presented to the MAPWG during the Navy's early 2000 assessment of AO (Held, Fedak, Crookenden, & Blanco, 2002).

Table 3
Range Corrected Validities of ASVAB Composites
with and without Assembling Objects

Rating	Rating Description	Best ASVAB Validity	Best ASVAB+AO Validity	Validity Dif.
Aviation Boatswain's Mate (Equipment) (N = 244)	Services hydraulics and arresting gear maintenance	VE+AR+MK+MC (.626)	AR+GS+AS+AO (.639)	.013
Builder (N = 339)	Performs wood and concrete construction	AR+MC+AS (.628)	AR+MC+AO (.643)	.015
Construction Mechanic (N =260)	Services gasoline and diesel engines	AR+MC+AS (.573)	AR+GS+AS+AO (.586)	.013
Parachute Rigger (N = 293)	Rigs parachutes, maintains survival equipment	AR+MK+EI+GS (.656)	AR+MK+EI+AO (.678)	.020
Quartermaster (N = 250)	Steers ships; logs compass readings, tides, bearings, etc.	VE+AR+MK+MC (.750)	VE+AR+MK+AO (.762)	.012
Signalman (N = 149)	Operates assorted visual and communications devices	AR+MK+EI+GS (.542)	AR+MK+EI+AO (.577)	.035

Note. (1) Validities were corrected for multivariate range restriction; differences between composites were computed as best ASVAB plus AO composite minus the best ASVAB composite. (2) The authors recognize that, as with any statistic, there are confidence intervals in validity differences between composites not reported here.

Table 3 shows for AO composites, as previously reported for CS, about a .02 validity increment on average across a subset of Navy ratings that are quite different than the CS ratings (indicating classification effectiveness). Although Table 3 provides only brief descriptions of the Navy's occupations (ratings), the major duties listed show the obvious relevance of spatial ability. The .02 incremental validity AO provides the ASVAB appears robust as it has been observed for several military occupations and performance criteria in studies conducted by the Army (Anderson, Hoffman, Tate, Jenkins, Parish, Stachowski, & Dressel, 2011; Russell, Le, & Putka, 2007), Marine Corps (Carey, 1994), and Navy (Held et al., 2002; Held et al., 2004).

Coding Speed and Assembling Objects: Reduced Adverse Impact

Prediction bias for a minority race/ethnic group is detectable when the test used for selection (or occupational classification) does not predict performance scores with the same sensitivity (not the same regression weights) as for the major group. Wise, Welsh, Grafton, Foley, Earles, Sawin, and Divgi (1992) found that across many military technical occupations, the ASVAB technical composites were equally sensitive (unbiased) for the studied major (Caucasian) and minority (African American) groups within the relevant occupational composite cutscore range. However, score barriers (adverse impact) were noted to some extent for the minority group and so the recommendation was for the military services to consider adding valid tests to the ASVAB (or to their classification systems) that reduce or eliminate such barriers (to occupational assignments). In response, the Navy adopted the ASVAB spatial ability test, Assembling Objects (AO) (See Table 1) for which both gender and minority group differences were considered less of an issue compared to the technical tests.

Table 4 provides a gender and minority group breakout of mean score differences as effect sizes for the ASVAB tests that included both AO and for CS before the test was eliminated from the battery.

Table 4
Effect Size Analysis for Gender and Subgroups

ASVAB	Male Effect Sizes Caucasian Group N = 22,230				Female Effect Sizes Caucasian Group N = 4,454			
	Af.Am. N = 6,117	Hisp. N = 4,049	Asian N = 1,777	Nat. Am. N = 1,523	Af.Am. N = 1,911	Hisp. N = 1005	Asian N = 383	Nat. Am. N = 410
GS	*.93	*.68	*.78	.03	*.87	*.68	*.53	.16
AR	*.70	.31	.09	.03	*.62	.29	-.07	.10
VE	*.65	*.59	*.73	-.01	*.66	*.57	.45	.12
MK	.19	.04	-.42	.05	.11	.02	-.41	.06
MC	*.93	.43	.43	-.01	*.83	.42	.34	-.03
AS	*1.13	*.73	*1.04	-.11	*1.09	*.84	*.94	.01
EI	*.76	*.52	.46	-.01	*.68	*.61	.39	.14
AO	*.58	.18	-.04	-.05	*.58	.22	.02	-.03
CS	.21	.10	-.08	.06	.17	.18	-.10	.07

Note. *denotes an effect size greater than |.5| (half a standard deviation), .5 being considered moderate. Effect size was calculated for the Navy's 1999 fiscal year recruit population as the major group mean (Caucasian) minus the minor group mean, the difference divided by the pooled groups' standard deviation.

Table 4, (from Held et al., 2002) shows mean differences in major and minor groups broken out by gender and transformed to effect size calculated as the difference between the major (Caucasian) and the specific minority group mean divided by the pooled group standard deviation (SD). Cohen (1988) characterizes standardized mean score differences of .2 as small, .5 moderate, and .8 large.

Table 4 shows the same effect sizes patterns for males and females across the race/ethnic groups (Caucasian being the common comparison group) suggesting cultural, experience or interest differences. Only effect sizes equal to or greater than |.5| are discussed because a .5 effect size is considered meaningful. African Americans had the largest number of effect size differences across tests and gender followed by Hispanics, and Asians. No meaningful effect size differences were found for Native Americans for either males or females. Auto/Shop (AS) had the largest effect size, favoring Caucasians (male and female), for the three major and minor group comparisons. No effect size difference at the |.5| SD criterion was observed for Mathematics Knowledge (MK) although the difference approached .5 for both genders favoring Asians (-.42 and -.41, respectively). AO, when compared to AS, had trivial effect sizes across three of the groups reaching the |.5| effect size criterion only for African Americans (for both genders). Finally, CS displayed trivial effect sizes for all group comparisons for both genders.

Figure 1 graphically shows the effect sizes for the ASVAB and CS tests for gender. [Figure 1 applies to the Table 4 data collapsed across all groups (Males, N = 35,831; Females, N = 8,246)].
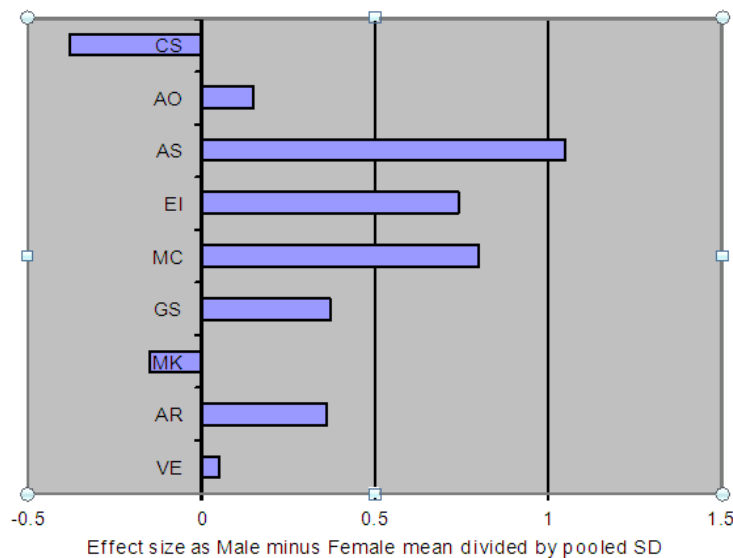

Figure 1
FY99 Navy ASVAB effect sizes for gender.

Figure 1 shows that CS was the only test to favor females to the extent of nearly reaching the study's |.5| effect size criterion. This outcome is consistent with previous research showing that females outperform males on clerical speed/accuracy tests (Majeres, 1988). The Mathematics Knowledge (MK) test also favored females but to a lesser extent than CS, which may be somewhat related to inconstant appeal for males and females in choosing the military as a career option. Effect sizes favoring males exceeded the +.5 SD criterion for the three technical knowledge tests, AS, EI, and MC. The effect size was smaller for the GS test, which has a large verbal content and therefore is not considered a pure technical test. AO showed lower effect sizes than these four tests and demonstrates, along with validity evidence, that it has more than one benefit in military occupational classification.

We note that the ASVAB AS and other technical tests have high utility in military classification because they measure not only knowledge of the subject matter relevant to the training and jobs, but potentially experience and interest that results in motivated engagement in technical endeavors (which involves/enhances the learning process). The goal is not to exclude technical knowledge tests from military occupational classification because of some group differences but to include other highly valid tests that can be used as alternative qualifiers. For example, the Navy currently uses the AO test as an alternative to AS in ASVAB composites that apply to some mechanical and engineering ratings so that recruits can qualify if they meet the operational cutscore on either composite (VE+AR+MK+AS "or" VE+AR+MK+AO). In all cases the cutscores are set on alternative composites to qualify about the same level of aptitude/ability reflected in the composites.

Because AO and CS are not administered to all recruits (see Note "b" in Table 1), by necessity an ASVAB-only composite is provided as an alternative. To be operationalized, the AO and CS composites (combined optimally with other ASVAB tests) must display a validity coefficient at least as large as that of the non-AO and non-CS composites; however, the Navy has had a more stringent requirement of incremental validity.

Coding Speed and Assembling Objects: Improved Classification

During the time that CS was being evaluated for elimination from the ASVAB, in addition to being concerned about lowering predictive validity and increasing adverse impact, the Navy was also concerned about losing differential assignment capability (Johnson, & Zeidner, 1991). As mentioned earlier, Scholarios et al. (1994) showed that CS provided differential assignment capability as well as increased mean predicted performance (Brogden, 1951). The Navy also was concerned that the evaluation at the time for negative impact in eliminating the CS test only addressed how many recruits qualified for an occupation on an ASVAB composite that contained CS could qualify for another on a composite without CS. The question not answered at the time was what if the number of recruits now not qualified for a rating because of elimination of CS was much larger than the number of rating slots (i.e., due to fill of a rating's annual goal). The implication was that, potentially, many recruits would not be assigned a rating.

The Navy's evaluation of differential assignment capability took the form of simulating recruit occupation assignments (ratings or, alternatively, jobs) using and not using CS and AO in ASVAB classification composites. Under varying scenarios using operational ASVAB composites as appropriate for ratings and composites that contained CS and AO where the validity warranted, the objective of the simulation exercise was to see how many recruits would not be classified/assigned across ratings given each rating had a fixed year's recruiting goal (school seats to be filled).

In all cases, the cutscores were set on the "alternative" composites to reflect about the same level of aptitude/ability in the ASVAB normative population. The algorithms did not, however, consider limitations such as physical or security clearance requirements, thus limiting the results to ASVAB impact only. Two classification algorithms in two different applications described in this section were used in the Navy's recruit classification/assignment simulation studies.

The Navy's Operational RIDE Algorithm

The first algorithm was developed by Navy Personnel Research, Studies, and Technology (NPRST) (Folchi, 2007; Folchi & Watson, 1997) and operationalized by EDS Federal Engineering and Logistics under contract (EDS Federal, 2001). The algorithm is now incorporated in the Navy's operational rating classification system called the Rating Identification Engine (RIDE) (Crookenden & Blanco, 2002). The algorithm's purpose is to generate a ranking of ratings to which a person would be optimally classified considering input personnel data and two utility functions. One utility function rewards applicants with high ASVAB classification composite scores (relevant to a particular rating's ASVAB standard) whereas the other function discourages classifying largely overqualified applicants to ratings whose jobs are considered not optimally challenging.

Data for recruits are entered into the classification system in a sequential manner (that involves a random selection component) in order to mimic the operational assignment process (in contrast to assigning all recruits in a batch). Four composite sets were formulated for the simulation that applied to the ratings. Composite Set 1 (baseline) contained only the ASVAB composites without CS or AO. Composite Set 2 contained some composites with AO where predictive validity warranted the composite's use. Likewise, Composite Set 3 contained some composites with CS. Composite set 4 contained all of the CS and AO composites (for their respective ratings). Differential assignment capability was defined as (1) the increase in the percentage of the recruit population "assigned" to ratings and (2) the lowest standard deviation of fill rate (indicating even fill).

Over 80 Navy ratings and over 150 rating/advanced program combinations were involved in the study, all referred to as "jobs". Males and females were simulated in separate analyses because some jobs were not open to females. Finally, four scenarios were applied for the four sets of composites that varied the ratio of "job slots" to recruits to mimic different recruiting environments (e.g., either too many or not enough recruits for slots). Table 5 shows the simulation study results.

Table 5
Simulated Rating Classification Results

	Composite Set without AO or CS	Composite Set with AO	Composite Set with CS	Composite Set with AO and CS
Scenario #1: 1.7% less female jobs than females (8,134 jobs; 8,275 females)				
Unassigned Recruits	469	413	389	288 - 303 (range with 4 runs)
Job Fill Standard Dev.	16.1%	15.1 %	14.6 %	13.7 % – 14.8 %
Scenario #2: 2.5% more female jobs than females (8,484 jobs; 8,275 females)				
Unassigned Recruits	501	440	279	279 – 300 (range with 4 runs)
Job Fill Standard Dev.	20.2 %	18.7 %	16.7 %	16.4 % – 17.4 %
Scenario #3: 6.2% more male jobs than males (38,402 jobs; 36,154 males)				
Unassigned Recruits	938	661	785	492 – 555 (range with 4 runs)
Job Fill Standard Dev.	13.8 %	13.4 %	13.0 %	12.6 % – 14.4 %
Scenario #4: 13.4% more male jobs than males (40,995 jobs; 36,154 males)				
Unassigned Recruits	387	71	213	0 (range with 4 runs)
Job Fill Standard Dev.	15.9 %	15.6 %	15.6 %	18.2 % – 19.3 %

Table 5 shows that, in each of the four classification simulation scenarios, providing an ASVAB composite set that included some composites with the AO or CS tests, and especially with both, resulted in fewer recruits "unassigned" to jobs. Also, the standard deviation of job fill (indicating evenness of distribution) tended to decrease with the addition of the AO and CS tests. The obvious exception is for scenario #4 (13.4% more male jobs than males to assign) where everyone was assigned to a job using the AO and CS composite set, but with less of an even fill across ratings (standard deviations ranged from 18.2% to 19.4% for 4 runs, as compared to the high of 15% for the other runs).

The Smallest ASVAB Delta Algorithm

The other Navy sequential assignment simulation application, developed by the Lewin Group, Inc. (Hogan & Simonson, 2004) is used when conducting Navy ASVAB validation/standards studies. The algorithm assigns a recruit (drawn randomly from a recruit population) to a rating for which the difference between the recruit's ASVAB composite score compared to the rating's operational composite's cutscore (compared across all ratings) is lowest. Despite this minimum delta criterion for a rating assignment (which involves a random tie breaker routine when ties occur), all ratings end up with a fairly wide range (distribution) of ASVAB scores to the right of the cutscore because there are relatively few low ASVAB scorers.

Consistent with the operational RIDE algorithm simulation study results, the study that used the Lewin Group, Inc. (Hogan & Simonson, 2004) algorithm also showed more recruits in the aggregate assigned to the ratings when the CS and AO composites were used compared to when they were not. [3] Not only was differential assignment improved (defined as more recruits being assigned to ratings at the same relative cutscores, thus with potentially at least the same expected performance levels), but at lower recruiting costs. The lower recruiting costs were due to a lower proportion of high AFQT category recruits (AFQT at least 50 but still with a high school degree diploma) required to fill the ratings. The observed lower proportion of high AFQT category recruits needed to fill the ratings was due to the lower correlation of CS and AO with AFQT, with many of the test used predominately in rating qualification. Historically higher AFQT youth with a high school diploma are more expensive to recruit.

The operational RIDE and the Lewin Group, Inc. algorithms are very different approaches to assessing the increased potential for qualifying large recruit populations to military occupations with ASVAB and adjunct classification tests. Both support use of CS and AO for military occupational classification.

Working Memory and
the Navy's Mental Counters (MCt) Test

Working memory (WM) at some level is obviously necessary in order to mentally register and manipulate written, visual, and auditory information in real time. WM is complicated considering all of its co-systems, subsystems, and neurological linkages and, even after years of research, is not fully understood (Burgess, Gray, Conway, & Braver, 2011; Logan, 2004). However, recent research has led to more clarification on the relationship of WM to intelligence, which is usually measured by fluid intelligence (gF) tests. The research comes from different disciplines, that is, neuroscience, cognitive science, developmental, and individual differences/experimental psychology. Much of the research involves the Raven Advanced Progressive Matrices (RAPM) (Raven, Raven, & Court, 1998), an instrument thought to be highly reflective of gF.

Studying the uniqueness and overlap in gF tests and the ASVAB gC academic/ knowledge tests is complicated and the interpretation of construct overlap is still a focus of concern. For example, Stauffer, Ree, and Carretta (1996) examined the relations between a battery of cognitive components tests (including measures of working memory) and the ASVAB and found a strong correlation (.98) between the general factors obtained from their hierarchical structures. The almost perfect correlation implies the same rank ordering of individuals if factor scores on the two general factors were used for selecting individuals for jobs. We provide a discussion of working memory in this section in order to more fully grasp the construct's theoretical underpinnings and relationship to gF.

---

[3] The Lewin Group, Inc. application is used for many purposes including comparing diversity across ratings to the operational classification state, as well as, to assess the fill potential for women in technical ratings (females historically score lower on the ASVAB technical tests).

Working Memory is generally thought to have an executive control center responsible for providing attentional focus (Baddeley, 1986). Without having to construct a formal experiment, we can readily observe that external distractions impact individuals' performance on WM tasks, particularly attending to one task while having to deal with another. WM has been identified by Air Force subject matter experts (SMEs) as a key ability related to successful job performance for Air Traffic Controllers (ATC) (Carretta & Siem, 1999) and unmanned aerial vehicle operators (Paullin, Ingerick, Trippe, & Wasko, 2011). Consistent with this finding, a military project that examined the linkage between knowledge, skills, and abilities to an occupation's major job duties (that included ATC) showed a requirement for attentional focus and speed/accuracy for several of the limited number of occupations that were included in the study (Waters, Russell, Shaw, Allen, Sellman, & Geimer, 2009).

Recent neurologically based research explored the commonality of the ability to focus attention, referred to as "interference-control ability brain activity" (I-CA) as part of both WM and gF (Burgess et al., 2011). WM, measured by several types of span tasks and gF, measured by the RAPM and the Cattell Culture Fair Test (Cattell, 1973) produced I-CA brain activity that was depicted in a path model with both directly influencing gF and indirectly influencing gF through WM; however, the direct path became statistically non-significant after accounting for the indirect (through WM) path (p. 684). While not conclusive, the study supported WM as the mediator of gF performance rather than the other way around.

Also recently, Wiley, Jarosz, Cushen, and Colflesh (2011) offered new detailed insights into the component level relationship between measures of WM capacity and the RAPM. The study measure of WM was an operation span test that, similar to the Navy's working memory test, MCt, involves storing information while at the same time processing new information (the OSpan task, Conway, Krane, Bunting, Hambrick, Wilhelm, & Engle, 2005). The strongest relationship between the OSpan scores and the RAPM scores was for RAPM items that required successful application of new rules in combinations. A weaker relation was observed for items that involved increased difficulty of the rules or the increased number of repeated rule combinations. The authors stated that "...the quality of the executive function, and not capacity per se, may be responsible for the relationship between WM capacity and the RAPM (p. 261)". The quality of the executive function is taken to include the ability to control one's focus of attention.

Jarosz and Wiley (2012) shed further light on the quality of the executive control in WM capacity that most strongly relates to the RAPM. In a carefully designed counterbalanced experiment, individuals were administered two versions of the RAPM. One version included the wrong answer to the specific item that was most frequently selected by a normative sample, thereby considered the most distractive, that is, the most difficult to reject. The other version replaced this difficult to reject distracter with an easy to reject wrong answer. The complex RSpan (Reading) as well as OSpan (Conway et al., 2005) were used as the WM capacity measures. The result of interest was that the WM capacity measures were more strongly correlated with RAPM scores for the RAPM version that contained the high level distracters.

Jarosz and Wiley (2012) contend that their results are consistent with other research that shows that "...WMC aids in resisting interference from visually distracting stimuli... (p. 433)". Finally, from the developmental research area, we see support for both speed and working memory involved in performance on the RAPM. Fry and Hale (1996) tested a group of children, adults, and adolescents on multiple processing speed and working memory tests as well as the RAPM and showed a "...developmental cascade in which increases in processing speed result in improvements in working memory that, in turn, contribute to improvements in fluid intelligence (p. 241)." Although there were age related differences in all measures, after controlling them, speed and working memory were statistically significant in predicting level of cognitive ability.

We clearly see from the working memory literature the relevance of the construct for many military occupations, including air traffic controller, foreign language interpreters in their endeavors to learn a foreign language, pilots tracking the positions of enemy planes and otherwise multi-tasking, and unmanned aerial vehicle (UAV) operators. Military psychologists understood the utility of working memory and other gF tests in occupational classification as evidenced by a large scale military test development research project that spanned the late 1980s and early 1990s. These efforts involved the military's development of such gF measures to complement the largely gC ASVAB. One of the gF tests was the working memory test that has been accepted for implementation on the military's CAT-ASVAB platform. The test, Mental Counters (MCt), is described and pictured in Alderton et al. (1997). The MCt test displays three counters on the computer screen (three short horizontal lines displayed on one row), each initialized to the same numerical value. The examinee's task for each item is to determine the final numerical value for each of the three counters after a series of empty boxes are randomly displayed one at a time either above or below a targeted counter. A counter's value is incremented by 1 if the box is displayed above the counter and decremented by 1 if below. Item difficulty is manipulated by number of boxes displayed and box exposure time.

A predictive validity study involving the ASVAB and the ECAT tests found MCt to add .05 and .10 validity above the ASVAB classification composite (VE+AR) that the Air Force uses for qualifying ATC candidates (Held & Wolfe, 1997, p. 81). The outcome variable was performance (scores) from the Air Force ATC training modules that mimic real tower operations (.05 and .10 incremental validity for basic and advanced approach tower control operations, respectively).

Also in the ASVAB/ECAT predictive validity study, MCt was shown to add incremental validity to the Navy's CS composite, VE+MK+CS, which is used for classifying recruits to several "Operations" type of occupations, including the Navy's Operations Specialist (OS) rating. The OS rating has major job duty tasks that include quick and accurate plotting of a ship's position and providing updated target identification data to the ship's Command Information Center (CIC). In an ECAT study, MCt provided a .05 increment to the .71 validity of VE+MK+CS (Held & Wolfe, 1997). The validity estimates and incremental validities for the OS rating were considered relatively stable unrestricted population estimates (range corrected) due to (1) a large (n = 815) sample and (2) a not too stringent operational cutscore.

The Navy SEALs, also an "Operations" rating, recently requested a working memory test for their screening system after establishing a job requirement for quick and accurate strategic decision making from synthesis of multiple/sequential/dispersed inputs. As mentioned earlier, CS is already being used a part an alternative ASVAB qualification system and the Navy expects to evaluate MCt along with CS and AO in the near future.

There are many other operations types of occupations in the military including the Navy's Cryptologic Technician Interpretive (CTI) rating. CTIs attend difficult and fast paced foreign language courses at the Defense Language Institute, Foreign Language Center (DLIFLC). CTIs on the job are required to efficiently decode lengthy intelligence transmissions in both written and listening modes. Most recently, a major DoD rework of the Defense Language Aptitude Battery (DLAB) used to screen military candidates for DLIFLC found that a working memory test contributed uniquely to the prediction of foreign language proficiency over and above measures of verbal intelligence, artificial language rules application measures (a deductive reasoning task), and inductive pattern application (Bunting et al., 2011). Bunting concluded from his research in working memory (Bunting, 2006; Conway et. al, 2005) that the Navy's MCt is a suitable substitute for the study measure of working memory.

Inclusion of a WM test in a new DLAB (DLAB2) is consistent with recent principled approaches to establishing working memory as a foreign language aptitude (Wen, 2012; Wen & Skehan, 2011).[4] Other second language learning research involving hierarchical models of aptitude complexes have shown both speed and working memory to be a part of different aptitude complexes (Robinson, 2007).

Conclusions

This paper provided both theoretical and empirical support for three tests used, or to be used shortly (MCt), in the Navy's enlisted occupational classification system. As with AO and CS, the MCt test now being administered to Navy applicants testing at the MEPS is expected to (1) increment the validity of the ASVAB for predicting training performance for important military occupations, thereby reducing academically related failure and setback rates, (2) reduce adverse impact for females and several minority groups in qualifying for some occupations due to these groups having less exposure, for whatever reason, to content measured by the ASVAB technical tests, and (3) increase the proportion of recruits each year filling Navy rating goals (on an ability/aptitude basis).

---

[4]The International Language Learning Roundtable on Memory and Second Language Acquisition was held in 2012 with multidiscipline participation in the cognitive and psycholinguistic areas to further the understanding of the theoretical and methodological issues regarding the role of memory in second language learning. The Proceedings are located at http://lc.ust.hk/~center/conf2012/doc/2012%20Roundtable%20Booklet.pdf.

The AO, CS, and MCt tests have been shown to measure different domains when factor analyzed in their respective batteries (Alderton et al., 1997, p. 30).[5] That is, the tests are at least somewhat non-overlapping and therefore offer unique contributions to differential assignment capability. CS belongs to a "Clerical Speed" factor within the ASVAB (Ree & Carretta, 1994) from which it was eliminated. AO belongs to an ECAT "Spatial Ability" factor within the ECAT from which it originated. MCt (also an ECAT test) belongs to an ECAT "Working Memory" factor (Alderton et al. 1997, p. 30). These three gF related factors complement the more academic/knowledge factors, Verbal, Technical, and Mathematics, of the ASVAB.

Logically the ECAT and other gF tests could be thought of as indicators of what individuals can reason out given unfamiliar (non-academic) content of the problem (Cattell, 1971, p.99). In this regard gF tests could be more important than crystallized intelligence (gC) in predicting job performance, as noted about the ECAT (Wolfe et al., 1997). The augmentation of the ASVAB in some form to include both types of measures is consistent with Cattell's assertion that both fluid and crystallized intelligence are important aspects of mental ability.

Discussions are currently underway within the MAPWG and the DAC-MPT about establishing a "Philosophy" of the ASVAB that will guide what kinds of tests to include in the battery. The timing of these discussions is most likely optimal as any potential downturn in the military recruiting environment (as the economy improves) will spur efforts to qualify youth for military enlistment based on other than purely gC measures. A potential additional benefit of including the tests discussed in this paper as part of a comprehensive military occupational classification battery is that they can be developed as automated item generation tests.

As an example of automated item generation, the MCt working memory test could be developed to display the boxes above and below the three counter lines by simply including a software routine that, on the fly, randomly assigns the boxes' positions. (While currently the MCt boxes appear to be randomly displayed, the item displays are the same for every examinee.) Automated item generated tests of this sort, which could include perceptual speed and spatial ability tests, do not require the traditional resource intensive item development efforts that apply to knowledge based tests, but also, they are less subject to test compromise.[6] In addition, efforts are now underway to assess what ASVAB tests can be consolidated where constructs if not the content are similar (e.g., the MK and AR tests). Consolidating ASVAB content where appropriate would leave more testing time for military relevant new tests that demonstrate utility improvements for military selection and classification.

---

[5] The ECAT Psychomotor Skill tests were not recommended for future evaluation due to peripheral equipment requirements that were intractable to maintain and also because of large mean differences in scores that would impact females.
[6] Practice effects do occur for these types of tests, but the CAT-ASVAB delivery system provides a generous time allotment during the tests instruction phase to practice with sample items.

References

Abrahams, N., Walton-Paxton, E., Alf, E. F., Barton, M. A., Cole, D. R., & Kieckhaefer, W. G. (1996, February). *Hardware independent measures of the ASVAB clerical proficiencies test.* (1996). (RGI, Inc. Contract N66001-95-D-960, Order No. 7J06). Report for Navy Personnel Research and Development Center (NPRDC), San Diego, CA.

Ackerman, P. L., & Beier, M. E. (2007). Further explorations of perceptual speed abilities in the context of assessment methods, cognitive abilities, and individual differences during skill acquisition, *Journal of Experimental Psychology, 13,* 249-272.

Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT Battery, *Military Psychology, 9,* 5-37.

Anderson, L. M., Hoffman, R., Tate, B., Jenkins, J., Parish, C., Stachowski, A., & Dressel, J. D. (2011). *Assessment of Assembling Objects (AO) for improving predictive performance of the Armed Forces Qualifications Test*, Technical Report 1282. Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Baddeley, A. (1986). *Working memory.* Oxford, UK: Oxford University Press.

Brogden, H. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement, 11,* 173-196.

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 183 - 196.

Bunting, M. F. with extensive co-author/contributor acknowledgement. (2011, March). *Reinventing DLAB: Potential new predictors of success at DLIFLC* (Technical Rep. TT0 82114E.3.3.1). Center for Advanced Study of Language (CASL), University of Maryland.

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General, 140,* 674-692.

Buscigilo, H. H., & Palmer, D. R. (1996). *An empirical assessment of coaching and practice effects on three Army tests of spatial aptitude,* ARI Research Note 96-70. Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Buscigilo, H. H., Palmer, D. R., King, I. H., & Walker, C. B. (1994). *Creation of new items and forms for the Project A Assembling Objects test,* Technical Report 1004. Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., & Zook, L. M. (Eds.) (1992). *Building and retaining the career force: New procedures for screening and assigning Army enlisted personnel: Annual report, 1990 fiscal year*, ARI Technical Report 952. Alexandria, VA: U. S. Research Institute for the Behavioral and Social Sciences.

Carey, N. B. (1994). Computer predictors of mechanical job performance: Marine Corps findings. *Military Psychology, 6*, 1-30.

Carretta, T. R., & Siem, F. M. (1999). Determinants of enlisted air traffic controller success. *Aviation, Space, & Environment Medicine 70*, 910-918.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. NY: Houghton Mifflin.

Cattell, R. B. (1973). *Measuring intelligence with the culture fair tests*. Champaign, IL: The Institute for Personality and Ability Testing.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (revised edition)*. Hillsdale, NJ: Erlbaum.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Whilhem, O., & Engle, R. W. (2005). Working memory and span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12,* 769-786.

Crookenden, M. & Blanco, T. (2002). *SCORE test, comparing composite tests*. (Delivery Order EDS 02-002). EDS draft technical report prepared for Navy Personnel Research, Studies, and Technology (NPRST), Millington, TN.

EDS Federal. (2001). *Alternative algorithms for job selection*. (Contract Number: N66001-00-F-CR00/GS-35F-0015K Delivery Order: FISC NOR (0014)); Sabre (GSA 01-002). EDS technical report prepared for Navy Personnel Research, Studies, and Technology (NPRST), Millington, TN.

Folchi, J. (2007). *RIDE vs. CLASP comparison and evaluation: Models and parameters*. (NPRST-TR-07 of March 2007). Millington, TN: Navy Personnel Research, Studies and Technology.

Folchi, J. & Watson, S. (1997). *Specifications for RIDE utility functions*. Unpublished technical report prepared for Navy Personnel Research, Studies, and Technology (NPRST), Millington, TN.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7,* 237 - 241.

Held, J. (2006). *Armed Services Vocational Aptitude Battery (ASVAB) standards: Air Traffic Control rating* (NPRST Letter Report Ser 3900, PERS-1/00047 31 May 2006): Millington, TN: Navy Personnel Research, Studies, and Technology.

Held, J. (2011). *Armed Services Vocational Aptitude Battery (ASVAB) standards: Special Warfare Operations (SO) SEAL from program to rating* (NPRST Letter Report Ser 3900, BUPERS-1/00092 11 Aug 2011). Millington, TN: Navy Personnel Research, Studies, and Technology.

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 International Military Testing Association, pp. 291-297. Ottawa, Canada.

Held, J., Fedak, G., & Johns, C. (2004). *Armed Services Vocational Aptitude Battery (ASVAB) standards: Aviation mechanics*, NPRST Letter Report Ser 3900 PERS-13/000011 of 24 Feb 04). Millington, TN: Navy Personnel Research, Studies, and Technology.

Held, J. D. & Foley, P. P. (1994). Explanations for accuracy of the general multivariate formulas for correcting for range restriction. *Applied Psychological Measurement, 18,* 355-367.

Held, J. D. & Wolfe, J. H. (1997). Validities of unit-weighted composites of the Armed Services Vocational Aptitude Battery (ASVAB) and Enhanced Computer Administered Tests (ECAT). *Military Psychology, 9,* 77-84.

Hogan, P. & Simonson, B. (2004). *Selection and classification cost effectiveness model* (Contract Number: MOBIS N00639-02-F-A-097). The Lewin Group, Inc. software and user manual prepared for Navy Personnel Research, Studies, and Technology, Millington, TN.

Johnson, D. & Ziedner, J. (1991). *The economic benefits of predicting job performance: Vol 2. Classification efficiency.* NY: Praeger.

Jarosz, A. F. & Wiley, J. (2012). Why does working memory capacity predict RAPM performance? A possible role of distraction. *Intelligence, 40,* 427 - 438.

König, C. J., Bühner, M., & Mürling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance, 18*, 243-266.

Lawley, D. (1943). A note on Karl Pearson's selection formula. *Royal Society of Edinburgh, Proceedings, Section A, 62,* 28-30.

Logan, G. D., (2004). Working memory, task switching, and executive control in the task span procedure. *Journal of Experimental Psychology: General, 133,* 218 - 236.

Majeres, R. L. (1988). Serial comparison processes and sex differences in clerical speed. *Intelligence, 14,* 149 - 165.

Oberlander, E. M., Oswald, F. L., Hambrick, D. Z., & Jones, A. (2007). *Indivual difference variables as predictors of error during multitasking* (NPRST-TN-07-09). Millington, TN: Navy Personnel Research, Studies, and Technology

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

Peterson, N. G., Russell, T. L., Hallam, G., Hough, L., Owens-Kurtz, C., Gialluca, K., & Kerwin, K. (1990). Analysis of the experimental predictor battery. In. J. P. Campbell & L. M. Zook (Eds.), *Building and retaining the career force: New procedures for accessing and assigning Army enlisted personnel: Annual Report 1990 FY* (Research Note 952). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Pommerich, M. (2013, October). *Preliminary evaluation of Coding Speed.* Briefing given to the Navy by Defense Manpower Data Center – Personnel Testing Division. Seaside, CA: Defense Data Manpower Center.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Advanced Progressive Matrices (1998 edition)*. Oxford, England: Oxford Psychologists Press.

Ree, M. J., & Carretta, T. R. (1994). Factor analysis of ASVAB: Confirming a Vernon-like model. *Educational and Psychological Measurement, 54*, 459-463.

Robinson, P. (2007). Aptitudes, abilities, contexts, and practice. In R. M. DeKeyser (Ed.), *Practice in a second language* (pp. 256 - 286). NY: Cambridge University Press.

Russell, T. L., Le, H., & Putka, D. (2007). Predictor measure validation methods and Armed Services Vocational Aptitude Battery results. In D. J. Knapp & T. R. Tremble (Eds.), *Concurrent validation of experimental Army enlisted personnel selection and classification measures,* ARI Technical Report 1205. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Russell, T. L., & Peterson, N. G. (2001). The experimental battery: Basic attribute scores for predicting performance in a population of jobs. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 269-306). NY: Erlbaum.

Russell, T. L., Peterson, N. G., Rosse, R. L., Hatten, J. T., McHenry, J. J., & Houston, J. S. (2001). The measurement of cognitive, perceptual, and psychomotor abilities. In J. P. Campbell & D. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 71-109). Hillsdale, NJ: Erlbaum.

Sager, C. E., Peterson, N. G., Oppler, S. H., Rosse, R. L., & Walker, C. B. (1997). An examination of five indexes of test battery performance: Analysis of the ECAT battery. *Military Psychology, 9*, 97-120.

Schmidt, F. L., Dunn, W., & Hunter, J. (1995). *Potential utility increases from adding tests to the Armed Services Vocational Aptitude Battery (ASVAB)* (NPRDC-TN-95-5). San Diego, CA: Navy Personnel Research and Development Center.

Scholarios, D. Johnson, C. & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology, 3,* 412-424.

Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science, 58,* 1438-1457.

Segall, D. O. (1997). The psychometric comparability of computer hardware. In W. A. Sands, B. K. Waters, & J. R. McBride, (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 219-226). Washington, DC: American Psychological Association.

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

Segall, D. O. (2010, October). *Table Processing Speed: A new processing speed test*. Presentation given to the Manpower Accession Policy Working Group. Seaside, CA: Defense Manpower Data Center.

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

Waters, S., Russell, T., Shaw, M., Allen, M., Sellman, W., & Geimer, J. (2009). *Development of a methodology for linking ASVAB content to military job information and training curricula.* (HumRRO Final Report FR-09-64). Alexandria, VA: Human Resources Research Organization.

Weiner, E. L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft,* NASA Conference Report 177528. NASA Ames Research Center.

Wen, Z. (2012). Preview Article - Working memory and second language learning. *International Journal of Applied Linguistics, 22,* 1-22.

Wen, Z. & Skehan, P. (2011). A new perspective on foreign language aptitude research: Building and supporting a case for "Working memory as language aptitude", *ELT Journal Advanced Access*, 15-36.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davies (Eds.), *Varieties of attention* (pp. 63–102). NY: Academic Press.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50,* 449-455.

Wiley, J., Jarosz, A. F., Cushen, P. J. & Colfflesh, G. J. H. (2011). New rules use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 256 – 263.

Wise, L., Welsh, J., Grafton, F., Foley, P., Earles, J., Sawin, L., & Divgi, D. (1992). *Sensitivity and fairness of the Armed Services Vocational Aptitude Battery (ASVAB) technical composites.* (DMDC Tech. Rep. No. 92-002). Monterey, CA: Defense Data Manpower Center.

Wolfe, J. H. Alderton, D. L., Larson, G. E., Bloxom, B. M. & Wise, L. L. (1997). *Expanding* the content of CAT-ASVAB: New tests and their validity. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 239-249). Washington, DC: American Psychological Association.

Distribution List

AIR FORCE PERSONNEL CENTER, STRATEGIC RESEARCH AND ASSESSMENT
BRANCH (HQ AFPC/DSYX)

AIR FORCE – FORCE MANAGEMENT POLICY DIRECTORATE, TRAINING AND
EDUCATION REQUIRMENTS AND RESOURCES DIVISION (HQ AF/A1PT)

AIR UNIVERSITY LIBRARY

ARMY RESEARCH INSTITUTE LIBRARY

ARMY WAR COLLEGE LIBRARY

CENTER FOR NAVAL ANALYSES LIBRARY

CHIEF OF NAVAL PERSONNEL (OPNAV 132G, NAVY SELECTION AND
CLASSIFICATION OFFICE)

DEFENSE MANPOWER DATA CENTER (CHIEF, PERSONNEL TESTING DIVISION)

HUMAN RESOURCES DIRECTORATE TECHNICAL LIBRARY

JOINT FORCES STAFF COLLEGE LIBRARY

MARINE CORPS MANPOWER AND RESERVE AFFAIRS, MANPOWER PLANS
DIVISION, INTEGRATION AND ANALYSIS SECTION (SECTION HEAD)

MARINE CORPS UNIVERSITY LIBRARIES

MILITARY ACCESSION POLICY WORKING GROUP

NATIONAL DEFENSE UNIVERSITY LIBRARY

NAVAL HEALTH RESEARCH CENTER

NAVAL POSTGRADUATE SCHOOL DUDLEY KNOX LIBRARY

NAVAL RESEARCH LABORATORY RUTH HOOKER RESEARCH LIBRARY

NAVAL WAR COLLEGE LIBRARY

NAVY PERSONNEL RESEARCH, STUDIES, AND TECHNOLOGY SPISHOCK
LIBRARY

HQ USMEPCOM (DIRECTOR OF TESTING)

OFFICE OF NAVAL RESEARCH (CODE 34)

OFFICE OF THE UNDERSECRETARY OF DEFENSE (PERSONNEL & READINESS)
(ASSISTANT DIRECTOR, ACCESSION POLICY)

PENTAGON LIBRARY

USAF ACADEMY LIBRARY

US COAST GUARD ACADEMY LIBRARY

US MERCHANT MARINE ACADEMY BLAND LIBRARY

US MILITARY ACADEMY AT WEST POINT LIBRARY

US NAVAL ACADEMY NIMITZ LIBRARY