Navy Personnel Research, Studies, and Technology Division
Bureau of Naval Personnel (NPRST/BUPERS-1)

Millington, TN 38055-1000 **NPRST-TN-07-12 August 2007**

Revision and Expansion of Navy Computer Adaptive Personality Scales (NCAPS)

Robert J. Schneider, Ph.D.
Kerri L. Ferstl, Ph.D.
Janis S. Houston, Ph.D.
Walter C. Borman, Ph.D.
Personnel Decisions Research Institutes, Inc. (PDRI)

Ronald M. Bearden, M.S.
Amanda O. Lords, Ed.D.
Navy Personnel Research, Studies, and Technology

NPRST-TN-07-12
August 2007

Revision and Expansion of Navy Computer Adaptive Personality Scales (NCAPS)

Robert J. Schneider, Ph.D.
Kerri L. Ferstl, Ph.D.
Janis S. Houston, Ph.D.
Walter C. Borman, Ph.D.
Personnel Decisions Research Institutes, Inc. (PDRI)

Amanda O. Lords, Ed.D.
Ronald M. Bearden, M.S.
Navy Personnel Research, Studies, and Technology

Reviewed and Approved by
Jacqueline A. Mottern, Ph.D.
Institute for Selection and Classification

Released by
David L. Alderton, Ph.D.
Director

Navy Personnel Research, Studies, and Technology (NPRST/BUPERS-1)
Bureau of Naval Personnel
5720 Integrity Drive
Millington, TN 38055-1000
www.nprst.navy.mil

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
31-10-2006	Technical Note	01/01/2006 - 03/31/2006

4. TITLE AND SUBTITLE

Revision and Expansion of Navy Computer Adaptive Personality Scales (NCAPS)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Robert J. Schneider, Kerri L. Ferstl, Janis S. Houston, Walter C. Borman, Amanda O. Lords, Ronald M. Bearden

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

PREVISOR, Inc.
Personnel Decisions Research Institutes, (PDRI)
100 South Ashley Dr. Union Center, Suite 775
Tampa, FL 33602

8. PERFORMING ORGANIZATION REPORT NUMBER

NPRST-TN-07-12

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

Navy Personnel Research, Studies, and Technology (NPRST/PERS-1)
Bureau of Naval Personnel
5720 Integrity Dr.
Millington, TN 38055-1000

10. SPONSOR/MONITOR'S ACRONYM(S)

NPRST

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

12. DISTRIBUTION/AVAILABILITY STATEMENT

A - Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report documents Phase 3 of the development of "Navy Computer Adaptive Personality Scales" (NCAPS). This phase of the instrument development includes the analyses and the recommendations regarding revision and enhancement of the "Adaptability/Flexibility", the "Stress Tolerance", and the "Self-Reliance" scales. Furthermore, it also documents the development of the "Leadership Orientation", the "Self-Control/Impulsivity", and the "Perceptiveness/Depth of Knowledge" scales. A total of 390 new items were generated for the three new NCAPS attributes. The NCAPS item pool currently measures 13 non-cognitive constructs, with a total of 1,884 items.

15. SUBJECT TERMS

personality assessment, interrater reliability, behaviorally-anchored rating scales, whole person assessment, computerized adaptive testing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Genni Arledge
UNCLASS	UNCLASS	UNCLASS	UNCLASS	39	19b. TELEPHONE NUMBER *(Include area code)* 901-874-2115 (882)

Foreword

This report documents the development of the Navy Computer Adaptive Personality Scales (NCAPS). NCAPS is a computer adaptive personality measure being developed for use in the selection and classification of Sailors for entry level Navy enlisted jobs. This important research program will overhaul and improve the Navy's enlisted selection and classification process. The over program—Whole Person Assessment—is designed to replace the current classification algorithm with a more flexible and accurate. Consequently, it will allow us to de-emphasize the almost exclusive focus on mental ability by including personality and interest measures in making classification decisions. Collectively, these efforts will transform and modernize enlisted classification by making it applicant-centric while improving job satisfaction and performance, reducing attrition, and increasing continuation behavior.

NCAPS uses a cutting-edge technological approach to personality measurement that is designed to mitigate many problems that plague traditional instruments, which rely upon Likert rating scales. Likert scales contain sets of homogeneous items, which are subject to both directed faking and socially desirable responding. To minimize these problems, NCAPS incorporates a paired forced-choice item format, uses a complex item response theory (IRT) adaptive selection and scoring algorithm, and intersperses item content. The complexity and novelty of the design constraints requires a series of interrelated research projects. This report covers how the personality constructs were selected, items were developed and scaled, and the results from an initial test of the validity of NCAPS.

The research was sponsored by the Office of Navy Research (Code 34) and funded under PE 0602236N and PE 0603236N.

David L. Alderton, Ph.D.
Director

Executive Summary

This report documents Phase 3 of the development of *Navy Computer Adaptive Personality Scales* (NCAPS), an innovative computer adaptive, paired-comparison measure of personality traits. Phase 1 involved identification, development, and pilot testing of the first three NCAPS scales: Achievement, Stress Tolerance, and Social Orientation. Phase 2 involved identification and development of seven additional NCAPS scales and initial validation of NCAPS. This Phase 3 report documents (a) analyses and recommendations regarding revision of certain existing NCAPS scales to enhance their validity; and (b) development of three additional scales to be incorporated into NCAPS: Leadership Orientation, Self-Control/Impulsivity, and Perceptiveness/Depth of Knowledge.

Though initial NCAPS results were quite promising, a few scales performed less well than expected. We therefore conducted supplemental analyses of the Phase 2 validity data set in an attempt to improve the measurement quality of existing NCAPS scales. Review of facet-level validities, scatter plots, and other relevant statistics led to the following recommendations:

1. Remove the "Works with Different People" facet from the Adaptability/Flexibility scale;
2. Remove the "Puts Aside Worries/Guilt" facet from the Stress Tolerance scale; and
3. Truncate the Self-Reliance scale so that it only includes items:
   a. at trait levels ranging from 2.0–5.7 (on a 2–8 point scale); and
   b. that are not similar in content to items at trait levels above 5.7 (to avoid compromising validity and/or unidimensionality).

A conversion formula was derived to place the truncated Self-Reliance scale scores on the same 2–8 metric as the other nine existing NCAPS scale scores.

The three new scales were selected for inclusion in NCAPS based on: (a) Phase 2 literature review and expert rating of task results linking personality traits to Navy success for enlisted personnel; and (b) the professional judgment of NPRST psychologists regarding the Navy's current selection and classification requirements.

Scale development activities for the three new traits to be incorporated into NCAPS included the same basic steps as for previous NCAPS scale development work: facet identification, item writing and review, scaling the items in terms of their trait levels and relevance to their targeted traits, and final review of items to ensure adequate trait level coverage. A total of 390 new items were generated for the three new NCAPS attributes. The NCAPS item pool now measures 13 non-cognitive constructs, with a total of 1,884 items.

Contents

List of Figures

List of Tables

Introduction

In response to the realization that cognitive ability alone is not an adequate predictor of all of the outcomes important to the modern Navy, an effort was initiated to add one or more measures of other characteristics to the Armed Services Vocational Aptitude Battery (ASVAB; U. S. Department of Defense, 1984) for selection and classification purposes. The decision to develop a personality inventory as a potential complement to the ASVAB in Navy selection and classification followed from work presented in Borman, Hedge, Ferstl, Kaufman, Farmer, and Bearden (2003) and Ferstl, Schneider, Hedge, Houston, Borman, and Farmer (2003), and was conducted under the auspices of the Navy Personnel Research, Studies, and Technology (NPRST) Division, Bureau of Naval Personnel.

NPRST sought to develop an innovative approach to personality assessment using state-of-the-science psychometric methodologies and personality research with the potential for increasing reliability, validity, and utility of personality assessment. This effort resulted in development of an instrument called Navy Computer Adaptive Personality Scales (NCAPS).

NCAPS is based on the Computer Adaptive Rating Scale (CARS) methodology developed by Borman and his colleagues within the performance rating domain (Borman, Buck, Hanson, Motowidlo, Stark, & Drasgow, 2001). NCAPS initially presents item-pairs representing two levels of a trait, one below the scale midpoint and the other above it. The paired-comparison approach was used to provide a better approximation of interval-level measurement than traditional personality instruments, which arguably provide only ordinal level data (Thurstone, 1927). Depending on which item an examinee chooses as more self-descriptive, NCAPS revises the examinee's estimated trait level using Bayes model estimation (Stark & Drasgow, 1998), and then selects two additional items whose trait level values bracket the revised estimated trait level in a way that maximizes trait-level information in an item response theory (IRT) sense. The examinee's selection of the more self-descriptive item for the second paired-comparison results in further revision of the examinee's estimated trait level and the selection of two more statements that once again bracket the (now updated) estimate of the examinee's trait level, and maximize information. Up to 15 item-pairs are presented per trait.

This report documents Phase 3 of the development of NCAPS. Phase 1 was documented in Houston et al., (2003). That report describes development and pilot testing of the first three NCAPS scales: Achievement, Stress Tolerance, and Social Orientation. Phase 2, documented in Houston, Borman, Farmer, and Bearden (2005), involved identification and development of seven additional NCAPS scales and initial validation of NCAPS. This report first documents analyses and recommendations regarding revision of existing NCAPS scales to enhance their validity. It then describes development of three more scales to be incorporated into NCAPS: Leadership Orientation, Self-Control/Impulsivity, and Perceptiveness/Depth of Knowledge.

Revision of Existing NCAPS Scales

The Houston et al. (2005) report describes results of an initial criterion-related validity analysis of the current 10-scale version of NCAPS. In this section, we describe the use of that data set to explore revision of those scales to enhance their validity. In order to clarify our discussion, however, we first provide some additional background.

Two versions of NCAPS were developed in Phases 1 and 2. These were labeled "Adaptive" and "Traditional" NCAPS. Adaptive NCAPS is the CARS-based personality instrument described above. A traditionally formatted version of each NCAPS scale was also developed and administered to examinees for comparison purposes and evaluation of the construct validity of Adaptive NCAPS. Traditional NCAPS consists of 205 items, selected from the total NCAPS item pool to be representative with respect to content and trait level. Examinees responded to Traditional NCAPS items using a 5-point Likert-type scale ranging from "strongly disagree" to "strongly agree."

Computer-based versions of both Adaptive and Traditional NCAPS were administered to 305 Navy enlisted personnel in late 2004. Performance ratings on a subset of these examinees were obtained from their peers and supervisors. Ratings were obtained using 7-point behavior summary scales on 10 dimensions found to be important to work performance in naval enlisted positions: (1) Cooperating/Working Well with Others, (2) Task Proficiency and Productivity, (3) Adaptability/Flexibility, (4) Initiative and Self Development, (5) Knowledge/Support of Unit/Command Objectives, (6) Problem-Solving and Decision-Making, (7) Integrity/Honesty, (8) Work Ethic, (9) Communicating Effectively, and (10) Overall Potential. A unit-weighted composite of these dimensions was computed based on factor analysis results showing that a single factor could account for the intercorrelations between these 10 dimensions in both peer and supervisor rating data (Schneider, Borman, & Houston, 2005).

Criterion-related validities of Traditional and Adaptive NCAPS scales against peer and supervisor ratings reported by Schneider et al. (2005) are shown in Table 1.

Table 1
Uncorrected zero-order correlations between existing Traditional and Adaptive NCAPS scales and peer and supervisor ratings of overall performance

Existing NCAPS Scale	Uncorrected Unit-Weighted Overall Performance Composite (Peer Ratings)		Uncorrected Unit-Weighted Overall Performance Composite (Supervisor Ratings)	
	Traditional	Adaptive	Traditional	Adaptive
Adaptability/Flexibility	.17	.12	.12	.10
Attention to Detail	.24	.24	.12	.17
Achievement	.25	.27	.07	.35
Dependability	.31	.20	.10	.23
Dutifulness	.21	.14	.11	.09
Social Orientation	.21	.14	.02	.22
Self-Reliance	.19	.03	.10	.05
Stress Tolerance	.26	.21	.03	.18
Vigilance	.19	.17	.03	.13
Willingness to Learn	.18	.07	.29	.19

Note. For peer ratings, n = 195 for Adaptive NCAPS correlations and n = 190-197 for Traditional NCAPS correlations; correlations $\geq$ .14 are statistically significant at $p < .05$. For supervisor ratings, n = 85 for Adaptive NCAPS correlations and n = 78 for Traditional NCAPS correlations; for Adaptive NCAPS, correlations $\geq$ .18 are statistically significant at $p < .05$, one-tailed, and, for Traditional NCAPS, correlations $\geq$ .19 are statistically significant at $p < .05$, one-tailed.

In order to determine the degree of overlap between the personality scales measured by NCAPS and overall performance, we computed a unit-weighted composite of the 10 NCAPS scales in both the Traditional and Adaptive formats. The Traditional and Adaptive NCAPS composites had uncorrected correlations with the unit-weighted, peer-rated Overall Performance composite of .30 and .24, respectively (both $p < .05$). When corrected for criterion unreliability, those validities rose to .39 and .32, respectively. We also regressed the unit-weighted, peer-rated Overall Performance composite on the 10 NCAPS scales. The shrunken multiple correlations (i.e., the estimated population cross-validated multiple correlations) were .20 for Traditional NCAPS and .23 for Adaptive NCAPS. After correcting for criterion unreliability, these values rose to .26 for Traditional NCAPS and .30 for Adaptive NCAPS.

We did a similar analysis for supervisor-rated criteria. In that analysis, the Traditional and Adaptive NCAPS composites had uncorrected correlations with the unit-weighted Overall Performance composite of $r = .13$ (*n.s.*) and $r = .27$ ($p < .05$), respectively (the difference between these two correlations is statistically significant at $p < .01$). When corrected for criterion unreliability, those validities rise to $r = .18$ and .37, respectively.[1]

While the foregoing analyses show that NCAPS validity results were very promising, they also show that certain NCAPS scales (e.g., Adaptability/Flexibility, Self-Reliance) did not do quite as well as expected. We therefore sought to improve the measurement quality of existing NCAPS scales, focusing special attention on under-performing scales.

One possible way of doing this was to compute item-level validities against the unit-weighted peer- and supervisor-rated overall performance criteria and eliminate items with low validities. We decided against this approach, however. First, the reliability of single personality items is low, which makes validity coefficients hard to interpret. One might argue that satisfactorily high validity coefficients against both peer *and* supervisor ratings would mitigate those interpretational difficulties. The problem with this argument is that:

1. The two validity coefficients are not statistically independent, since peers and supervisors rated the same examinees.
2. Peer and supervisor ratings are not highly correlated (r = .37), which means that very few item-level validities would meet even modest validity requirements in both the peer and supervisor data sets. Indeed, if we were to apply a requirement that an item will be dropped if its validity against both supervisor and peer ratings is below r = .05, we would end up dropping substantially more items than we would retain.
3. The use of item-level validities would limit scale revision to Traditional NCAPS items only, since Adaptive NCAPS presents item-pairs, drawn from a much larger pool of items.

Another approach—and the one we decided to use—would be to examine facet-level validities. The use of facet-level validities has the advantage of allowing us to look at validities based on higher-reliability subsets of NCAPS scales than individual items and to generalize from Traditional NCAPS items to the Adaptive NCAPS item pool. It should be noted that the reason facets were created was merely to guide item writing efforts, and not for use as sub-scales. As such, some of the facets have only two or three items in Traditional NCAPS scales, with correspondingly limited alpha coefficients. In those cases, facets are not useful guides to scale revision for essentially the same reason that individual items are not useful guides to scale revision, and were therefore not used.

---

[1] We did not use multiple regression to evaluate the overlap between the predictor space and the supervisor-rated criterion space because the more limited sample size associated with the supervisor rating data was not sufficient to support the sample size requirements of multiple regression.

Removal of an entire facet of an NCAPS scale should require strong evidence that the facet has little or no predictive power. The bar for removal of a facet should therefore be set reasonably high. As such, we determined that, for a facet to be considered for removal from NCAPS, it must have the following characteristics.

- At least four items
- An alpha coefficient > .40
- No statistically significant correlation either with the unit-weighted Overall Performance composite or any individual performance rating scale, in either the peer- or supervisor-rating data sets

Table 2 presents facet-level information to facilitate this analysis, and shows that very few facets meet these criteria for removal. Within the Adaptability/Flexibility scale, however, the Works with Different People facet is a good candidate for removal. It has five items, with an alpha coefficient of .51, and does not correlate significantly with any performance variable in either the supervisor or peer rating data. Moreover, it differs conceptually from the other three Adaptability/Flexibility facets in that it involves adapting to people, as opposed to non-interpersonal phenomena (e.g., tasks, jobs, and situations). It is also noteworthy that the Works with Different People facet is the only one of the four Adaptability/Flexibility facets that is not even marginally correlated (i.e., at $r > .10$ and $p < .10$) with the Adaptability/Flexibility performance dimension in either the peer or supervisor rating data. Finally, there are 191 items presently in Adaptability/Flexibility scale item pool, 36 of which make up the Works with Different People facet. This leaves 155 items, which is more than sufficient to populate an NCAPS scale. On the basis of the foregoing, we recommend that the Works with Different People facet be dropped from the NCAPS Adaptability/Flexibility scale.

Another facet that appears to be a prime candidate for removal from NCAPS is the Puts Aside Worries/Guilt facet of the Stress Tolerance scale. This facet is comprised of six items, with an alpha coefficient of .70, but does not correlate significantly with the overall performance composite or any individual performance variable in either the peer or supervisor rating data. The NCAPS Stress Tolerance scale item pool presently has 119 items, 25 of which make up the Puts Aside Worries/Guilt facet. This leaves 94 items, which we believe will be sufficient to populate the NCAPS Stress Tolerance scale. On the basis of the foregoing, we recommend that the Puts Aside Worries/Guilt facet be dropped from the NCAPS Stress Tolerance scale.

Table 2
NCAPS Facets: Number of items, alpha coefficient, and zero-order correlations with supervisor- and peer-rated work performance criteria

			Cooperating/ Working Well with Others		Task Proficiency and Productivity		Adaptability/ Flexibility		Initiative and Self-Development		Knowledge/Support of Unit/Command Objectives		Problem Solving and Decision Making		Integrity/Honesty		Work Ethic		Communicating Effectively		Overall Potential (Global Rating)		Criterion Composite (Unit-Weighted Composite)	
NCAPS Facet	**k**	**α**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**	**Sup**	**Peer**
Adaptability/Flexibility																								
Willing to Change Task/ Project Approach	4	.45	.10	.16	.11	.11	.22	.17	.05	.15	.04	.10	.12	.06	.10	.10	.09	.07	.10	.15	.13	.14	.13	.15
Likes Variety	4	.42	-.01	.16	-.10	.15	-.03	.12	-.01	.13	.08	.11	-.09	.09	-.03	.11	.00	.06	-.15	.17	-.13	.11	-.05	.15
Work with Different People	5	.51	.05	-.02	.03	.03	.02	-.07	-.08	-.09	.17	-.05	.03	-.06	.04	-.01	-.11	-.08	.00	.08	.01	-.01	.02	-.04
Adapt to New Situations	5	.65	.09	.15	.27	.25	.10	.21	.16	.17	.20	.15	.11	.18	.04	.19	.19	.09	.11	.22	.12	.21	.18	.22
Attention to Detail																								
Exacting/Precise	5	.62	.10	.16	.24	.21	.03	.19	.11	.19	.08	.21	.06	.12	.06	.12	.16	.10	.02	.22	-.04	.15	.12	.21
Spot Imperfections/Errors	4	.67	.02	.18	.14	.24	.12	.24	.02	.19	-.06	.23	.04	.16	.03	.12	.06	.11	.03	.21	-.04	.25	.06	.23
Neat/Organized	7	.64	.14	.18	.11	.18	.00	.19	.13	.14	.14	.20	.11	.06	.21	.07	.14	.07	.00	.16	.04	.13	.15	.17
Achievement																								
Ambitious	3	.4	.07	.22	.26	.26	.09	.14	.13	.20	.07	.18	.10	.18	.08	.14	.06	.12	.17	.27	.07	.20	.15	.24
Challenging Goals	2	.22	.02	.11	.06	.11	-.03	.01	.00	.02	.00	.01	-.03	.02	-.13	-.01	.00	-.01	-.08	.09	-.11	.02	-.03	.05
Confident in Abilities	2	.35	.01	.22	.07	.25	-.02	.15	-.05	.18	.00	.07	-.07	.20	-.07	.16	.00	.10	-.01	.18	-.07	.17	-.02	.21
Persists Despite Obstacles	2	.33	-.07	.11	.04	.17	-.11	.07	-.14	.08	-.08	.04	-.17	.03	-.08	.09	-.05	.02	-.03	.08	-.14	.02	-.10	.10
Strives for Excellence	3	.43	.13	.24	.10	.24	.11	.20	.05	.23	-.12	.22	.01	.10	.05	.17	.17	.21	.05	.18	.11	.11	.08	.25
Works Hard/Long Time	3	.43	.12	.14	.12	.23	.00	.18	.01	.20	.02	.09	.07	.13	.13	.10	.13	.12	.02	.21	-.04	.16	.09	.20

Table 2 (continued)

ENCAPS Facet	k	α	Cooperating/ Working Well with Others		Task Proficiency and Productivity		Adaptability/ Flexibility		Initiative and Self-Development		Knowledge/Support of Unit/Command Objectives		Problem Solving and Decision Making		Integrity/Honesty		Work Ethic		Communicating Effectively		Overall Potential (Global Rating)		Criterion Composite (Unit-Weighted Composite)	
			Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer
Dependability																								
Orderly/Planful/Prioritizes	4	.64	.24	.26	.37	.26	.25	.23	.24	.24	.08	.20	.13	.16	.17	.13	.23	.16	.13	.22	.23	.21	.27	.26
Reliable/Efficient with Time	4	.61	.14	.27	.15	.27	-.03	.23	.06	.25	-.07	.22	-.03	.15	.15	.21	.19	.16	-.04	.26	.05	.17	.08	.28
Not Easily Distracted/Bored	4	.64	-.01	.28	.18	.35	.03	.26	-.02	.23	-.06	.18	.01	.18	-.06	.23	.06	.19	-.11	.27	-.04	.21	.00	.30
Doesn't Procrastinate	3	.47	-.04	.14	.13	.17	.00	.12	.01	.15	-.04	.14	-.03	.09	-.08	.12	.06	.10	-.06	.15	-.08	.11	-.01	.16
Dutifulness																								
Sense of Duty/Moral Obligation	3	.27	.00	.09	.15	.15	.03	.07	.01	.04	-.02	.02	-.09	-.01	-.05	.03	-.13	.03	.01	.06	-.04	.09	-.02	.07
Accepts Authority/Follows Rules	6	.68	.26	.12	.19	.21	.05	.15	.12	.10	.00	.08	.03	.08	.17	.09	.14	.07	-.04	.20	.06	.04	.14	.15
Honest/Trustworthy/Fulfills Obligations	6	.61	.08	.20	.14	.28	.02	.21	.06	.21	-.05	.12	-.05	.20	.06	.19	.11	.15	.03	.14	-.03	.20	.06	.24
Accepts Responsibility	4	.47	.21	.17	.19	.22	.08	.08	.06	.19	.02	.09	-.01	.10	.10	.15	.18	.10	-.04	.15	.02	.09	.11	.18
Social Orientation																								
Affiliation	11	.76	.07	.18	.08	.09	.07	.18	-.02	.11	.07	.15	.01	.07	.03	.13	-.06	.08	-.08	.14	-.02	.14	.02	.16
Agreeable	4	.39	-.02	.13	-.01	.07	-.05	.06	-.08	.03	.00	.11	-.21	.03	-.05	.11	-.12	.07	-.06	.13	-.12	.02	-.09	.10
Likes Teamwork	3	.35	.07	.12	.23	.08	.13	.15	.08	.07	.19	.10	.03	.00	.11	.07	.07	.07	.09	.17	.15	.03	.14	.11
Team Player	5	.64	-.02	.17	.09	.17	-.03	.21	-.07	.09	.01	.10	-.11	.10	-.03	.13	-.01	.05	-.06	.18	-.05	.16	-.04	.16

Table 2 (continued)

ENCAPS Facet	k	α	Cooperating/ Working Well with Others		Task Proficiency and Productivity		Adaptability/ Flexibility		Initiative and Self-Development		Knowledge/Support of Unit/Command Objectives		Problem Solving and Decision Making		Integrity/Honesty		Work Ethic		Communicating Effectively		Overall Potential (Global Rating)		Criterion Composite (Unit-Weighted Composite)	
			Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer	Sup	Peer
Self-Reliance																								
Not Dependent	6	.47	.04	.10	.21	.14	.00	.12	.19	.10	.12	.02	.14	.17	.03	.15	.21	.14	.09	.05	.12	.20	.15	.14
Self-Sufficient/ Resourceful	10	.69	-.01	.15	.13	.24	.01	.14	.12	.16	.06	.08	.12	.17	-.13	.13	.12	.08	.04	.10	-.06	.16	.06	.17
Stress Tolerance																								
Composure	10	.77	.09	.33	.22	.30	.05	.32	.18	.27	.11	.22	.08	.27	-.01	.27	.19	.19	.02	.30	.10	.28	.13	.34
Accepts Criticism	2	.23	.04	.05	.07	.01	-.16	.01	.03	.01	.02	.01	-.01	-.03	-.13	-.02	.04	-.07	-.04	.12	-.09	-.02	-.02	.01
Puts Aside Worries/Guilt	6	.70	-.02	.12	.05	.07	-.01	.06	.01	.06	.09	.08	.06	.02	-.10	.09	-.01	.02	.02	.08	-.03	.06	.01	.08
Willingness to Learn																								
Willing to Learn/Actively Seeks Learning Opportunities	5	.60	.22	.25	.08	.17	.07	.13	.01	.20	.03	.18	-.11	.06	.04	.18	.13	.14	.07	.26	.08	.16	.08	.22
Learns from Mistakes	4	.40	.27	.10	.30	.16	.23	.11	.21	.15	.23	.28	.25	-.03	.20	.11	.21	.06	.29	.16	.23	.11	.31	.15
Takes Good Advice	4	.56	.04	.16	.03	.13	.02	.06	.05	.11	.11	.14	.04	.03	.18	.08	.15	.02	.09	.15	.01	.08	.10	.12
Asks Clarifying Questions	5	.53	.36	.07	.28	.11	.30	.05	.34	.15	.28	.12	.22	.15	.20	.12	.26	.06	.22	.17	.32	.16	.35	.14

Note. *k* is number of items in facet. For peer ratings, n = 187-198, correlations $\geq$ .14 are statistically significant at $p < .05$, and correlations $\geq$ .12 are statistically significant at $p < .10$. For supervisor ratings, n = 78-86, correlations $\geq$ .21 (if n = 86) or .22 (if n = 78) are statistically significant at $p < .05$, and correlations $\geq$ .18 (if n = 78) or .19 (if n = 86) are statistically significant at $p < .10$.

One NCAPS scale that had surprisingly low validity was the Self-Reliance scale. Interestingly, however, each of the two facets that comprise Self-Reliance has statistically and practically significant correlations with multiple performance variables in peer and/or supervisor rating data. Given that our facet analysis did not provide a means of improving measurement of Self-Reliance, we further investigated the psychometric properties of that scale—especially the Adaptive version—in an attempt to determine why it did not do a better job predicting work performance. We also sought to determine why Adaptive Self-Reliance had validities that were much lower than Traditional Self-Reliance.

We began by examining scatter plots with Adaptive Self-Reliance plotted against the peer- and supervisor-rated Overall Performance composites. The scatter plot involving supervisor-rated performance revealed an interesting pattern, and is shown in Figure 1.

Data involving the supervisor ratings were of particular interest since we believe that the supervisor ratings were more accurate than the peer ratings, despite their more limited sample size (Schneider, Borman, & Houston, 2005). Figure 1 shows that Adaptive Self-Reliance is more predictive at lower trait levels and less predictive at higher trait levels (i.e., the data points are a better approximation of a line at lower trait levels). To evaluate this assertion more precisely, we computed validity coefficients at several trait levels for Adaptive and Traditional NCAPS against supervisor and peer ratings. Those results are shown in Table 3.

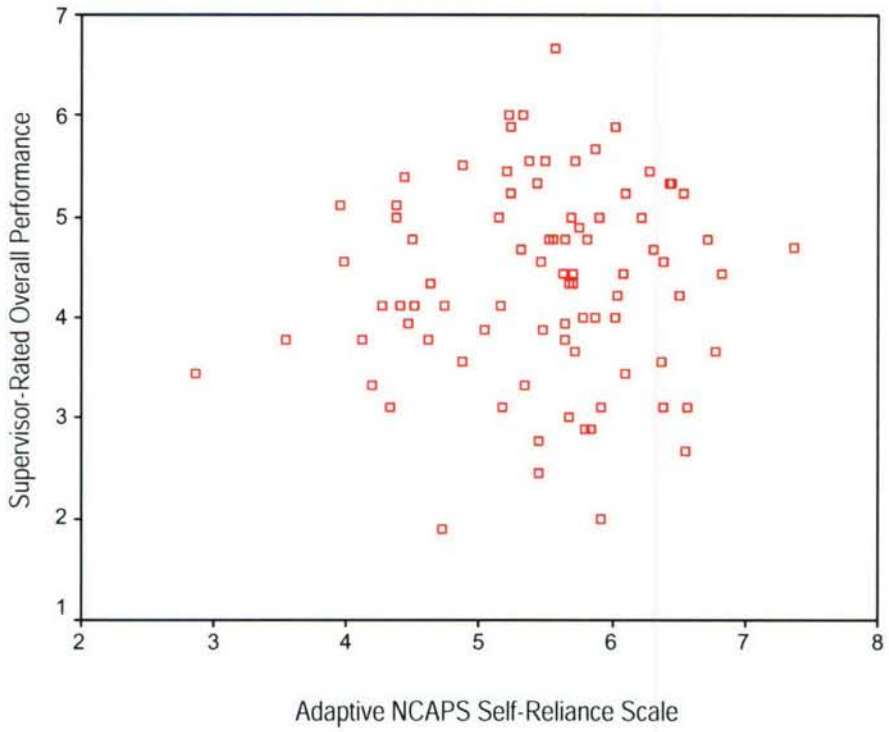

Figure 1. Scatter plot: Adaptive NCAPS Self-Reliance scale against supervisor-rated Overall Performance composite

Table 3
Criterion-related validities of Adaptive and Traditional NCAPS Self-reliance scales against peer- and supervisor-rated overall performance at various trait levels

	Adaptive NCAPS			Traditional NCAPS		
Percentile	Trait Level	Correlation with Supervisor-Rated Performance	Correlation with Peer-Rated Performance	Trait Level	Correlation with Supervisor-Rated Performance	Correlation with Peer-Rated Performance
40	5.57	.27	-.06	3.18	.15	.05
n		41	82		32	75
50	5.70	.19	-.02	3.25	.13	.13
n		48	99		42	89
60	5.79	.15	.01	3.35	.23	.10
n		55	122		52	110
70	5.96	.06	.05	3.44	.21	.09
n		62	139		60	134

The Adaptive NCAPS validities against the supervisor-rated criterion show exactly the pattern of declining validities suggested by the scatter plot. This led us to look for differences in item content at different trait levels to see why validity declines. What we found was that, at lower levels along the Self-Reliance trait continuum, the items primarily measure various forms of dependence (e.g., need for reassurance, insecurity with respect to one's own competence, excessive reliance on others' advice). At higher trait levels, however, careful inspection of the item content reveals a more mixed set of attributes. Some are positive (e.g., not needing much supervision, confidence in one's ability to make decisions on one's own, attempting to solve problems oneself rather than first going to others for help). Other items at the higher end of the Self-Reliance trait continuum seem less relevant to Navy criteria of interest, or possibly even negative/maladaptive (e.g., preferring to work alone; unwillingness to ask for help, even when doing so might be necessary/important).

The foregoing analysis may explain why the Traditional NCAPS Self-Reliance scale does not show the same pattern of declining validities as Adaptive Self-Reliance as one ascends the trait continuum. Several items in the Traditional NCAPS Self-Reliance scale were eliminated during scale refinement due to low item-scale correlations. These may reflect non-validity-enhancing or maladaptive traits that were largely uncorrelated with the more valid aspects of Self-Reliance. Since no such scale refinement was possible with Adaptive NCAPS, its validity may have suffered in comparison to that of its Traditional NCAPS counterpart. There is no clear-cut explanation for why the peer rating data validities were so much lower. However, for reasons stated above, we put more faith in the supervisor rating data than the peer rating data.

How might this information be used to improve measurement of the Self-Reliance scale? We recommend that the scale be truncated such that the higher trait level items are eliminated from the Adaptive Self-Reliance scale item pool. If this type of truncation is implemented, the next question is: At what trait level should the scale be truncated? Clearly, validity levels get higher at lower trait level percentiles. However, the scale also must have diagnostic relevance for a reasonable percentage of examinees. Based in part on review of the items representing various trait levels, as well as on the need to balance validity and examinee relevance, we recommend truncation of the Self-Reliance scale at the median, which corresponds to a trait level of 5.7 (on the 2–8 Adaptive NCAPS metric). We also recommend elimination of items below 5.7 that reflect the same multidimensional and/or validity-compromising content that many of the items at higher trait levels possess. We have identified 14 such items below 5.7, which leaves 113 items in the truncated version of the Self-Reliance scale. Fortunately, Self-Reliance had a large number of items in its item pool, which enabled us to remove a substantial number of items and still have an adequate supply to populate a truncated Adaptive NCAPS Self-Reliance scale.

Truncating at 5.7, of course, would put the Adaptive Self-Reliance scale on a different metric than the other Adaptive NCAPS scales. We addressed this problem by creating a simple transformation formula, as follows:

1. Compute the difference between 5.7 and 2.0, which represent the highest and lowest trait levels in the truncated scale.
2. Divide this difference by six (3.7/6 = .617).
3. Add .617 to 2.0, to arrive at the truncated scale value that corresponds to a value of 3 in the original, un-truncated (2–8) scale.
4. Add .617 to the sum computed in step 3 to arrive at the truncated scale value that corresponds to a value of 4 in the original, un-truncated scale; repeat this process until truncated scale values corresponding to all values in the original, un-truncated (2–8) scale have been computed.
5. Regress the seven un-truncated scale values (i.e., 2–8) on the seven truncated scale values.

This yields the following formula to convert truncated scale values to the 2–8 Adaptive NCAPS metric:

$$SRL_{full} = 1.621(SRL_{trunc}) - 1.243, \qquad (1)$$

where SRL_{full} is the score on the truncated version of the Adaptive NCAPS Self-Reliance scale, transformed to the 2–8 Adaptive NCAPS metric; and SRL_{trunc} is the score on the truncated version of the Adaptive Self-Reliance scale that is to be transformed.

Summary

In this section, we reviewed the promising initial evidence of the validity of NCAPS reported by Houston et al. (2005). We also noted that some NCAPS scales did not perform as well as hypothesized, and conducted more in-depth investigation to determine whether the validity of certain NCAPS scales could be enhanced. Review of facet-level validities, scatter plots, and other relevant statistics led to the following recommendations:

1. Remove the Works with Different People facet from the Adaptability/Flexibility scale;
2. Remove the Puts Aside Worries/Guilt facet from the Stress Tolerance scale; and
3. Truncate the Self-Reliance scale so that it only includes items:
   - at trait levels ranging from 2.0-5.7
   - that are not similar in content to items at trait levels above 5.7 such that they are likely to compromise validity and/or unidimensionality.

A conversion formula was derived to place the truncated Self-Reliance scale scores on the same 2-8 metric as the other nine existing Adaptive NCAPS scale scores.

Development of New NCAPS Scales

We also developed three new scales to be incorporated into NCAPS: Leadership Orientation (LDR), Perceptiveness/Depth of Thought (PER), and Self-Control/ Impulsivity (SCN). These Phase 3 scales were identified for development based on:

1. Analysis of expert rating task results reported by Houston and Cullen (2005) regarding the relevance of 19 personality constructs[2] (10 of which had already been incorporated into NCAPS) to overall success in the Navy, as well as success in 79 specific enlisted Navy positions.
2. Analysis of literature review reported by Schneider and Waters (2005) on the extent to which the same 19 personality constructs would be likely to be useful selection and classification tools for enlisted Navy positions.
3. The professional judgment of NPRST psychologists regarding the Navy's current selection and classification requirements.

[2] These 19 traits represent a comprehensive "middle-level" taxonomy of personality traits synthesized by Schneider and Waters (2005) for NCAPS development.

We developed the new scales following the same procedures used to develop the existing 10 NCAPS scales (Houston, Borman et al., 2005; Houston, Schneider et al., 2003). Those procedures were as follows:

- Facet identification – Although NCAPS was not intended to include scorable facets, we divided the construct definitions into distinct subcomponents. The resulting facets were used to aid item development.
- Item writing – PDRI researchers wrote new NCAPS items, targeting different trait levels to cover all facets of each target construct.
- Item review – All items were carefully reviewed, resulting in revision, deletion and addition of items.
- Trait level/relevance expert rating task – PDRI personality experts provided ratings used to scale each NCAPS item according to the level the targeted construct that it represents, as well as its relevance to that construct. Items were reviewed in an iterative process, based on these scaling results.
- Finalization of item pool – We conducted a final review of the items, and then recomputed item trait level counts at all trait levels to ensure adequate trait level coverage for each of the three new NCAPS scales.

Each of these activities is described below.

Facet Identification

The Schneider and Waters (2005) 19-trait NCAPS taxonomy was purposely constructed at a moderate level of trait specificity. In other words, we wanted constructs that were broad enough to allow for efficient measurement, but narrow enough not to obscure meaningful distinctions between traits (Ferstl et al., 2003). Thus, NCAPS was designed to yield construct (or scale) scores, but not narrower facet scores.

Although NCAPS does not have scorable facets, it has proven useful in previous NCAPS scale development work to divide construct definitions into their component parts for item writing purposes. In this project, therefore, we again divided each construct definition into facets before writing items. Thus, facets served as a guide for item writers to help them to cover all elements of each trait. After the items were scaled for trait level, we assessed trait level coverage by facet, and then focused on gaps when writing additional items. Definitions and facets for the constructs covered in this project appear in Table 4.

Table 4
Constructs and facets used in item development

Construct	Definition	Facets
Leadership Orientation (LDR)	willing to lead, take charge, offer opinions and direction, and take responsibility for guiding others' actions; able to mobilize others to act; is confident and decisive	LDR1 Willing to lead LDR2 Mobilize others LDR3 Decisive
Perceptiveness/Depth of Thought (PER)	interested in pursuing topics in depth; enjoys abstract thought and has a need to understand how things work; enjoys searching for patterns in data and understanding the "big picture;" knowledgeable about many things; perceptive and insightful	PER1 Need for/possession of in-depth knowledge PER2 Perceptive/Insightful
Self-Control/Impulsivity (SCN)	thinks through possible consequences before taking action; does not act on the "spur of the moment;" has no difficulty controlling emotions and behavior he/she knows to be inappropriate	SCN1 Control emotions SCN2 Control behaviors SCN3 Consider consequences

Item Writing and Review

Four PDRI researchers served as item writers. Each of these researchers had also written items in earlier phases of NCAPS development and they followed the same guidelines and procedures described in the reports documenting those efforts (Houston et al., 2005; Houston et al., 2003). Briefly, each item was to be a statement tapping one facet of a construct at a particular trait level, ranging from 1 to 7. Instructions provided to item writers included construct definitions; a definition of, and scale for, trait level; item formatting specifications; targeted reading level; and the desired (i.e., near-uniform) trait level distribution.

We wrote, reviewed, and scaled items in three rounds. This approach allowed us to ensure that the items were of high-quality and covered trait levels adequately for each construct. Once written, every item was reviewed by two or three other item writers prior to the expert rating task described below.

In Round 1, we wrote and scaled 349 items (108 LDR, 126 PER, and 115 SCN). In Round 2, we wrote and scaled 123 additional items (59 LDR, 32 PER, and 32 SCN). In Round 3, we wrote and scaled 10 more items (2 LDR, 2 PER, and 6 SCN). Thus, a total of 482 new draft items were written.

Trait Level/Relevance Expert Rating Task

Raters

All items written in Phase 3 were rated by PDRI researchers who are experts in the domains of personality research and work performance. Thirteen raters provided trait level ratings in both Rounds 1 and 2. In Round 3, three PDRI project team members scaled the final 10 items added to the item pool using a consensus discussion approach.

Procedure

Raters received a rating form that included rating instructions and all of the items to be rated, classified according to target construct. They did not see target facets or target trait levels for any item, and item order was randomized within each target construct.

The form presented raters with a brief description of NCAPS, though most of the raters were already familiar with the project and had participated in trait level scaling of items developed in the earlier NCAPS phases. Raters were asked to provide two expert ratings for each item: (1) a *Trait Relevance* rating, and (2) a *Trait Level* rating. The Appendix shows instructions for each rating presented to the raters, along with the rating scales used[3].

The Trait Relevance rating was not used in previous phases of NCAPS development. This is because, in previous phases, we were able to use the data from administration of the Traditional NCAPS version of each new scale to evaluate internal consistency, including item-scale correlations. In the present phase, however, traditionally-formatted NCAPS scales were not part of the development plan. To address the construct relevance of our items, we therefore used the alternate approach of asking raters to evaluate each item's trait relevance directly.

After making final decisions about retention of the Round 1 and 2 items (see below), we found a few places where there were fewer available items than we would have liked. Thus, we added a final set of 10 items to fill in the minor trait level gaps that remained. We scaled these Round 3 items using a consensus discussion approach. Three PDRI project team members used the instructions and rating scales described above (except that construct relevance was replaced by facet relevance), along with a subset of previously scaled Round 1 and 2 items with trait levels to provide context/calibration. They first rated trait relevance and trait level independently, and then discussed and reached consensus about the facet relevance and trait level for each of the 10 new items.

[3] It should be noted that, consistent with earlier phases of NCAPS development, trait level was established using a 1-7 scale. The existing NCAPS algorithm, however, requires a 2-8 scale for trait level, which is reflected in our discussion in the previous section of this report. The trait levels associated with each of the new items developed in this project will be converted to the 2-8 scale required by the existing NCAPS algorithm.

Data Screening

Outlier Ratings. The first step in analyzing the trait level ratings was to identify outlier ratings. As in Phases 1 and 2, we defined "outlier" as a rating that was separated from the nearest rating by more than one scale point with a frequency equal to 0. For example, if one rater gave the item a 2 and all the other ratings were 4s and 5s, the 2 was treated as an outlier. Combining the Round 1 and 2 scaling data, there were 6,136 individual ratings. Of these, 40 ratings (0.65%) were outliers. The outliers were assumed to be rater errors. As such, the individual outlier ratings were dropped from the data set before item statistics were computed.

Rater Screening and Interrater Reliability. Trait level ratings were analyzed for anomalous responding by individual raters. Interrater reliability was very good: The Shrout and Fleiss (1979) Case 2 intraclass correlation (ICC), corrected to a single rater, was .92 in Round 1 and .90 in Round 2. NCAPS methodology requires that trait level ratings of each item be very precise, so we conducted further analyses and used stringent criteria to determine whether the data provided by any of the expert raters should be eliminated from the data set used to estimate the trait level of NCAPS items.

Following procedures from Phases 1 and 2, we compared raters' profiles of trait level ratings to the profile of mean trait level ratings (computed across all other raters). Marked differences between a rater's profile and the mean profile would be evidence of anomalous responding. Corrected correlations with the mean rater profile and distance measures (i.e., Euclidean dissimilarity coefficients and average absolute deviation from the mean rater profile) revealed no evidence of anomalous responding. For example, each rater's trait level ratings correlated in the .90s with the mean of all other raters' trait level ratings and the highest average absolute deviation from the mean rater profile was .44 (mean = .36, SD = .03 for Round 1; mean = .34, SD = .05 for Round 2).

Next, trait relevance ratings were analyzed for signs of anomalous responding by individual raters. Interrater reliability and correlation indices were not very useful for the trait relevance ratings, because the vast majority of items were thought to be "definitely" or "probably" relevant by all raters. As such, there was little variance. However, distance measures, which were more meaningful, showed there was no evidence of anomalous responding. For example, the highest average absolute deviation from the mean rater profile was .47 (mean = .15, SD = .05 for Round 1; mean = .21, SD = .09 for Round 2). Moreover, ICC (2, *k*) was .74, despite the limited variance.

We also checked for evidence of logically inconsistent responding. First, we looked for cases in which raters responded "don't know" or "definitely not" to the question of whether an item was relevant to a trait, but nevertheless rated the item's trait level rather than using the "not applicable" option on the trait level rating scale. Second, we looked for cases in which a rater indicated that an item was "definitely" relevant to a trait, but nevertheless gave a trait level rating of "not applicable." These combinations of trait relevance and trait level ratings would be contrary both to logic and to the instructions given to the SMEs. Only two instances of logically inconsistent responding were present in the data. Both were resolved by asking the rater to re-rate the item.

Item Screening

After dropping individual outlier trait level ratings and deciding to retain all raters' remaining data, we calculated descriptive statistics for the trait relevance and trait level ratings. We used these data to inform item revision and retention decisions.

Trait Relevance Rating Results. Of 482 items, 477 had a trait relevance mean of 3.0 or higher. In other words, raters indicated that 99 percent of the items measured their target traits well enough that they probably or definitely should be kept in the test.

We specified some fairly strict criteria by which we flagged items for further review based on trait relevance ratings. All items meeting one or more of the following criteria were flagged for further review:

- Trait relevance mean < 3.0
- Two (15%) or more raters rated trait relevance < 3 (i.e., less than *probably* relevant)
- Nine (67%) or more raters rated trait relevance < 4 (i.e., less than *definitely* relevant)

Using these criteria, we flagged 44 items (9.1% of the item pool) for further review.

Trait Level Rating Results. Next, we applied criteria to identify potentially problematic items based on trait level. All items meeting one or more of the following criteria were flagged for further review:

- Two (15%) or more raters rated the item *not relevant* to the construct
- Trait level standard deviation ≥ .80
- Trait level range ≥ 5 (range = maximum – minimum + 1)
- Using these criteria, we flagged 59 items (12.2% of the item pool) for further review.

Review of Flagged Items. Eighty-two (17%) of the items were flagged based on one or more of the trait relevance or trait level criteria. Two members of the PDRI project team examined flagged items for content and item statistics, and then reached consensus about whether to keep or drop each item. We eliminated 50 of the 82 flagged items from the item pool.

The remaining 32 flagged items were retained. In most such cases, the item only met one of the six flagging criteria, and often met that criterion by a narrow margin. For example, some items were rated as "not relevant" to the construct by two or more raters, but the item content looked reasonable and the trait level ratings had an acceptably small range and SD. Other items were retained despite having SD ≥ .80, because the SDs were < 1.0, the ranges were acceptable (i.e., < 5), and the content appeared to be fine.

Finalization of NCAPS Item Pool

Final Item Review

After all of the steps described above, there were 432 items in the NCAPS item pool for LDR, PER, and SCN. At this point, we conducted a final review of the item pool, and eliminated 42 additional items. These 42 items had passed all screening criteria, but because there were more items than necessary in some places on the trait continuum, we could afford to be very selective and drop more items. The 42 items removed at this stage were removed for two reasons: (1) there was a very similar item in close trait level proximity, and/or (2) we judged that the item was potentially inappropriate (e.g., too complex) for the NCAPS target population. We were left with a final total of 390 items: 149 for LDR, 117 for PER, and 124 for SCN. The mean trait level across all retained trait level ratings (after excluding outlier ratings) became the final trait level for each of these items.

The statistics for items in the final item pool show that the finalized set of items for LDR, PER, and SCN are both relevant to their targeted trait and precise indicators of their trait level. They have an average trait relevance rating of 3.90 out of 4.0 (SD = 0.13), with a minimum of 3.15 and a maximum of 4.0 and appropriately small trait level standard deviations (mean = 0.53, SD = 0.18, median = 0.51, and maximum = 0.99).

Trait Level Coverage

In order for the adaptive CARS methodology to work properly, it is critical that each construct be represented by a sufficient number of items across the entire trait continuum. This goal informed our item writing throughout the project. To confirm that this goal was achieved, we conducted a final review of the distribution of trait levels represented in the item pool. Each distribution is based on the full and final set of items developed in Phase 3. Table 5 shows trait level distributions by construct and Table 6 shows trait level distributions by facet.

For each of the constructs, item counts are greatest at the highest and lowest trait levels. The middle of each trait level continuum is represented by fewer items, as was the case in Phases 1 and 2. However, previous NCAPS results indicate that the middle of each of the three trait continua is sufficiently represented. In other words, it is not the case that there aren't enough items in the middle of each scale; rather, there are more items than necessary at both ends of each scale.

Table 5
Final Phase 3 scales: Item counts by trait level and construct

Construct	Trait Level						
	1.00 to 1.99	2.00 to 2.99	3.00 to 3.99	4.00 to 4.99	5.00 to 5.99	6.00 to 7.00	Total Item Count
Leadership Orientation (LDR)	29	27	16	19	21	37	**149**
Perceptiveness/Depth of Thought (PER)	19	22	8	12	18	38	**117**
Self-Control/Impulsivity (SCN)	40	23	13	9	22	17	**124**
Total Item Count	**88**	**72**	**37**	**40**	**61**	**92**	**390**

Table 6
Final Phase 3 scales: Item counts by trait level and facet

Construct: Facet	Trait Level						
	1.00 to 1.99	2.00 to 2.99	3.00 to 3.99	4.00 to 4.99	5.00 to 5.99	6.00 to 7.00	Total Item Count
LDR1: Willing to lead	14	12	4	10	9	18	**67**
LDR2: Mobilize others	8	7	7	6	8	13	**49**
LDR3: Decisive	7	8	5	3	4	6	**33**
PER1: Need for/possession of in-depth knowledge	15	8	4	7	12	21	**67**
PER2: Perceptive/insightful	4	14	4	5	6	17	**50**
SCN1: Control emotions	13	12	3	4	9	9	**50**
SCN2: Control behaviors	14	6	5	4	7	4	**40**
SCN3: Consider consequences	13	5	5	1	6	4	**34**
Total Item Count	**88**	**72**	**37**	**40**	**61**	**92**	**390**

Summary

In this section, we described identification, development, scaling, screening, and finalization of 390 items measuring three new NCAPS constructs: Leadership Orientation, Perceptiveness/Depth of Knowledge, and Self-Control/Impulsivity. The NCAPS item pool now measures 13 non-cognitive constructs, with a total of 1,884 items. Table 7 summarizes the development timeline and lists the scales currently in the test.

Table 7
NCAPS scales and development timeline

Development Phase	Year Completed	Scale Names
Phase 1	2003	AV: Achievement SO: Social Orientation ST: Stress Tolerance
Phase 2	2005	ADF: Adaptability/Flexibility ADL: Attention to Detail DEP: Dependability DUT: Dutifulness/Integrity SRL: Self-Reliance WTL: Willingness to Learn VIG: Vigilance
Phase 3	2006	LDR: Leadership Orientation PER: Perceptiveness/Depth of Thought SCN: Self-Control/Impulsivity

References

Borman, W. C., Buck, D. E., Hanson, M. A., Motowidlo, S. J., Stark, S., & Drasgow, F. (2001). An examination of the comparative reliability, validity, and accuracy of performance ratings made using computerized adaptive rating scales. *Journal of Applied Psychology, 86*, 965-973.

Borman, W. C., Hedge, J. W., Ferstl, K. L., Kaufman, J. D., Farmer, W. L., & Bearden, R. M. (2003). Current directions and issues in personnel selection and classification. In J. J. Martocchio & G. R. Ferris (Eds.). *Research in personnel and human resources management.* Stamford, CT: JAI Press.

Ferstl, K. L, Schneider, R. J, Hedge, J. W., Houston, J. S., Borman, W. C., & Farmer, W. L. (2003*). Following the Roadmap: Evaluating Potential Predictors for Navy Selection and Classification.* (Technical Report No. 421). Minneapolis: Personnel Decisions Research Institutes, Inc.

Houston, J. S., Borman, W. C., Farmer, W. L., & Bearden, R. M. (Eds.) (2005). *Development of the Enlisted Computer Adaptive Personality Scales (ENCAPS), Renamed Navy Computer Adaptive Personality Scales (NCAPS)* (Institute Report #503). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Houston, J. S., Schneider, R. J., Ferstl, K. L., Borman, W. C., Hedge, J. W., Farmer, W. L., & Bearden, R. M. (2003). *NCAPS: Development of the Enlisted Computer Adaptive Personality Scales for the United States Navy* (Institute Report #449). Minneapolis: Personnel Decisions Research Institutes, Inc.

Schneider, R. J., Borman, W. C., & Houston, J. S. (2005). Initial validation of ENCAPS. In J. S. Houston, W. C. Borman, W. L. Farmer, & R. M. Bearden (Eds.), *Development of the Enlisted Computer Adaptive Personality Scales (ENCAPS), Renamed Navy Computer Adaptive Personality Scales (NCAPS)* (Institute Report #503). Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Stark, S., & Drasgow, F. (1998, April). *Application of an IRT ideal point model to computer Adaptive assessment of job performance.* Paper presented at the Society for Industrial and Organizational Psychology Annual Conference, Dallas, TX.

Thurstone, L. L. (1927). Psychophysical analysis. *American Journal of Psychology, 38*, 368-389.

U. S. Department of Defense (1984). *Test Manual for the Armed Services Vocational Aptitude Battery (DoD 1340.12AA).* North Chicago, IL: U.S. Military Entrance Processing Command.

Appendix:
Expert Rating Task Instructions

Expert Rating Task Instructions

Trait Relevance Rating

As you know, one of the most important characteristics of personality trait scales is their internal consistency. In past NCAPS development work, we were able to pilot test a Traditional paper-and-pencil version of the scales so that we could drop statements that did not correlate well with their associated scale score. This time, however, we will not be able to pilot test the statements using a Traditional format, and the computer adaptive format of NCAPS does not allow us to compute statement-scale correlations or to evaluate internal consistency reliability. We are therefore asking you to make a Trait Relevance rating for each statement.

You will use the following scale to make your Trait Relevance ratings:

Do you think this statement measures its target trait well enough that it should be kept in the test?	
4	Definitely
3	Probably
2	Probably not
1	Definitely not
d/k	Don't know

This scale will drop down when you click in the trait relevance response box for each statement. When making a trait relevance rating, please consider the following factors:

- Is the statement adequately related to its target trait's definition?
- Are the respondents' scores on the statement likely to be sufficiently related to their overall scale scores on the target trait (i.e., item-total correlations of about .20 or higher)?
- Is the statement's meaning clear and unambiguous?

Trait Level Rating

In order to form appropriate pairs of statements for NCAPS, it is essential that we obtain accurate estimates of the trait level of each statement. Thus, we ask that you rate the level on the target trait (i.e., construct) that is reflected in each of our draft statements.

Please make a *Trait Level* rating using the following scale, which will drop down when you click in the response box:

A person who agrees with this statement has a(n) _____ level of [*the target trait*].

7	Extremely high
6	High
5	Slightly high
4	Moderate
3	Slightly low
2	Low
1	Extremely low
n/a	Not applicable

If you gave a statement a *Trait Relevance* rating of 1 ("Definitely not"), rate that statement's Trait Level as n/a ("Not applicable"). If you gave a statement a *Trait Relevance* rating of 2 ("Probably not") or d/k ("Don't know"), you will also likely rate that statement's trait level as n/a ("Not applicable"). You may, however, choose to rate that statement's trait level (despite your rating its relevance as 2 or d/k) if you think there is sufficient possibility that the statement measures its target trait.

Note that the lowest trait level rating, a "1," indicates that the statement reflects an extremely low level of the target trait, and *not* that the statement is a poor or irrelevant indicator of the target trait.

Distribution

AIR UNIVERSITY LIBRARY
ARMY MANAGEMENT STAFF COLLEGE LIBRARY
ARMY RESEARCH INSTITUTE LIBRARY
ARMY WAR COLLEGE LIBRARY
CENTER FOR NAVAL ANALYSES LIBRARY
DEFENSE TECHNICAL INFORMATION CENTER
HUMAN RESOURCES DIRECTORATE TECHNICAL LIBRARY
JOINT FORCES STAFF COLLEGE LIBRARY
MARINE CORPS UNIVERSITY LIBRARIES
NATIONAL DEFENSE UNIVERSITY LIBRARY
NAVAL HEALTH RESEARCH CENTER WILKINS BIOMEDICAL LIBRARY
NAVAL POSTGRADUATE SCHOOL DUDLEY KNOX LIBRARY
NAVAL RESEARCH LABORATORY RUTH HOOKER RESEARCH LIBRARY
NAVAL WAR COLLEGE LIBRARY
NAVY AERONAUTICAL MEDICAL INSTITUTE
NAVY PERSONNEL RESEARCH, STUDIES, AND TECHNOLOGY SPISHOCK LIBRARY (3)
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