

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
The 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 the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the 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.					
1. REPORT DATE (DD-MM-YYYY) 30-06-2006		2. REPORT TYPE Final		3. DATES COVERED (From - To) Dec 2001 - Mar 2006	
4. TITLE AND SUBTITLE Psychometric Issues Related to the Five-Vector Model				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER N00014-02-1-0179	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Robert L. Linn CRESST/University of Colorado at Boulder Stephen B. Dunbar University of Iowa				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UCLA CSE/CRESST 300 Charles E. Young Dr. North 300 GSE&IS/Mailbox 951522 Los Angeles, CA 90095-1522				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S) ONR	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A Approved for Public Release Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The Navy's five-vector model provides a comprehensive specification of the requirements for performance-based advancement of military personnel. The effectiveness of the model for making performance-based advancement decisions will depend heavily on the psychometric quality of the individual measures and on the psychometric models that are used to combine the information from the measures that comprise each of the vectors of the model. This report considers the vectors one at a time and then discusses challenges in combining information across the five vectors to make performance-based advancement decisions. Analysis results suggest that it is likely that the measurement information will vary across and within vectors in terms of validity and reliability, that the profile of information is likely to be uneven across vectors, and the information from one vector may even conflict with that from another vector. These properties can be expected to make the tasks of combining and integrating the information across vectors quite challenging.					
15. SUBJECT TERMS five-vector model, psychometric issues					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code)

**Psychometric Issues Related
to the Five-Vector Model**

Deliverable – June 2006

Knowledge, Models and Tools to Improve
the Effectiveness of Naval Distance Learning

Eva L. Baker
CRESST/University of California, Los Angeles

Office of Naval Research
Award # N00014-02-1-0179

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
301 GSE&IS, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

20060703030

Copyright © 2006 The Regents of the University of California

The work reported herein was supported under Office of Naval Research Award Number N00014-02-1-0179, as administered by the Office of Naval Research. The findings and opinions expressed in this report do not necessarily reflect the positions or policies of the Office of Naval Research.

TABLE OF CONTENTS

Abstract.....	1
Professional Development.....	2
Validity	3
Reliability	4
Personal Development	4
Personal Development Constructs	5
Leadership and Professional Military Education.....	6
Certification and Qualifications	8
Performance Continuum	9
Integrating Information from the Five Vectors	9
References	11

PSYCHOMETRIC ISSUES RELATED TO THE FIVE-VECTOR MODEL

Robert L. Linn

CRESST/University of Colorado at Boulder

Stephen B. Dunbar

University of Iowa

Abstract

The Navy's five-vector model provides a comprehensive specification of the requirements for performance-based advancement of military personnel. The effectiveness of the model for making performance-based advancement decisions will depend heavily on the psychometric quality of the individual measures and on the psychometric models that are used to combine the information from the measures that make up each of the vectors of the model. The psychometric challenges related to the five vectors have some things in common, but each of them also has some unique considerations. We begin by considering the vectors one at a time, and then discuss challenges in combining information across the five vectors to make performance-based advancement decisions. The results of our analysis suggest that it is likely that the measurement information will vary across and within vectors in terms of validity and reliability, that the profile of information is likely to be uneven across vectors, and the information from one vector may even conflict with that from another vector. These properties can be expected to make the tasks of combining and integrating the information across vectors quite challenging.

The Navy's five-vector model provides a comprehensive specification of the requirements for performance-based advancement of military personnel. The model identifies five broad aspects of performance that are required for the assessment of advancement "potential" of personnel. The five vectors are referred to as (1) professional development, (2) personal development, (3) leadership and professional military education, (4) certification and qualifications, and (5) performance.

Each of the five vectors can be further decomposed into a number of dimensions or constructs. The leadership and military education vector, for example, is conceptualized as including five general constructs called leading change, leading people, working with people, resource stewardship, and accomplishing mission. Each of these general constructs, in turn, comprises a number of more specific characteristics such as creativity and innovation, external awareness, conflict management, and financial management.

Building a personnel-evaluation system based on the five-vector system will require the development and management of a wide range of measures and data collection procedures. The effectiveness of the resulting system for making performance-based advancement decisions will depend heavily on the psychometric quality of the individual measures and on the psychometric models that are used to combine the information from the measures that comprise each of the vectors of the model. The psychometric challenges related to the five vectors have some things in common, but each of them also has some unique considerations. Hence, we will begin by considering the vectors one at a time. We will then discuss challenges in combining information across the five vectors to make performance-based advancement decisions.

Professional Development

The professional development vector depends on the training system starting with recruit training and continuing through training for each of the 90 separate specialty ratings. Ratings of progress have already been developed for 17 of the ratings and a few on-deck ratings have also been developed. For the ratings of progress and ratings on deck the main psychometric challenge is to obtain data and conduct analyses to determine the psychometric quality of the measures that have

been put in place. For the remaining ratings, the challenge is to develop valid and reliable progress ratings.

Validity

Validity is the most fundamental concern for any measure. Measures are not simply valid or invalid. Indeed, validity is neither an all or none characteristic nor a property of the measure per se. Rather, validity is the degree to which inferences and interpretations of measurement results are justified by supporting evidence and logical analysis. The validation process should begin by identifying the inferences and uses of measurement results that are intended. The process of developing measures needs to be pursued in a manner that is most likely to yield results that can support the desired inferences and uses.

Past experience in developing measures for use in workplace settings suggests that the place to start is with job analyses to identify the knowledge, skills, and abilities (KSA) needed to successfully perform the job. Job analysis information is particularly relevant for the development of measures that support valid inferences about professional development within a specialty rating. The measurement approach needed varies according to the nature of the KSA. Valid measures of core knowledge about such things as electric circuits, for example, can be most efficiently and effectively obtained by means of paper-and-pencil tests. Measurement of some skills, such as the repair of an engine, on the other hand may require hands-on performance measures or high-fidelity simulations. Similarly, the skill required to hit a target or the ability to read and interpret a radar screen generally require hands-on measures of performance or high-fidelity simulations.

Although hands-on measures have greater fidelity for the tasks that are actually performed on the job, they also present a number of psychometric challenges. It is often impossible, for example, to standardize the task. Consequently, the relative difficulty of the task varies from one person to the next. That is, there is a confounding of the measured skill with the difficulty of the task. Such a confounding of difficulty with skill is a potential source of invalidity of the measure that must be considered in evaluating the quality of the measurement results. Simulations can sometimes be constructed that have a high degree of fidelity for critical aspects of the task and have the advantage of greater standardization. For example, simulations that require the demonstration of trouble-shooting skills can present the same fault to each person thereby avoiding the confounding of fault

difficulty with the person's skill. On the other hand, simulations cannot measure the individual's skill in actually manipulating physical objects as would be required, for example, to replace a part.

Hands-on measures often require that the performance be rated by human judges. Ratings of the same performance may vary from one judge to the next. Variability of results to rater introduces a potential source of invalidity that needs to be evaluated. Judgments may also be subject to systematic biases. Some judges have a tendency to be more lenient than others. Even for something that appears to be as straight forward as scoring marksmanship based on the holes in various score rings of a target, it has been found that the distribution of scores tends to have spikes at the minimum score required to qualify as marksman or sharpshooter, suggesting that judges give the benefit of doubt to personnel who are near the cutting score. Similar tendencies need to be considered whenever judges are aware of pre-specified minimum levels of performance required to qualify.

Reliability

Reliability is the second psychometric consideration in evaluating the technical quality of a measure. Measures of knowledge, skills, and abilities consist of samples from a domain of possible measures. A paper-and-pencil test of knowledge could have a variety of different test items, for example, and the question is how consistent the scores would be if different forms of the test consisting of different samples of items were administered. Where scores depend on rating by judges, it is also important to evaluate the degree of consistency of ratings provided by different judges.

Personal Development

Many of the psychometric challenges that arise in the measurement of professional development also are relevant to measures of personal development. The issues discussed above concerning validity and reliability of ratings provided by judges are relevant to judgments of personal development constructs as well as measures of professional development. A number of additional psychometric challenges arise in the measurement of constructs that comprise the personal development vector, however.

Personal Development Constructs

The personal development vector includes six general constructs: (1) lifelong learning; (2) life skills; (3) interpersonal relationships; (4) health, wellness and recreation; (5) financial management skills; and (6) values. The knowledge, skills, and abilities required as part of professional development are relatively concrete and are supported by information that can be obtained from job analyses. The personal development constructs, on the other hand, are relatively abstract and depend in several instances on "softer", more difficult to measure characteristics. The interpersonal relationships construct, for example, is intended to include components of personal and family relations, parenting, anger management, stress management, and emotional awareness. Similarly, the values construct is intended to include character education and development through Navy core values, ethics, perseverance, acceptance, and human spirit. These components are quite abstract, poorly defined, and difficult to measure.

It seems highly unlikely that reliable measures that would support valid inferences about the quality of parenting or of family relations, for example, could be obtained on a routine basis for all sailors. It is conceivable that self-report ratings might be obtained or even ratings provided in response to questionnaires completed by family members or acquaintances. Such ratings would be subject to many potential biases and distortions resulting from the desire to make oneself or a family member look good. Such distortions would seriously undermine the validity of the results. Some aspects of interpersonal relationships, such as anger and stress management might be evaluated in the negative. Disciplinary actions could be obtained from personnel records, for example. The accuracy and comprehensiveness of information in personnel records has often been questioned, however.

Instructional experiences of sailors aimed at character education and the development of Navy core values could be monitored. Participation in an instructional experience, obviously does not guarantee that a person has internalized the core values that the experience is intended to promote. Perseverance, another of the value components identified under the values construct might be indirectly assessed by indicators such as the completion of courses and assignments, but the generalizability of such indicators to other situations would be uncertain.

Indicators of the lifelong learning construct can be readily identified. The accumulation of college credits and college grade-point average provide indicators

of the pursuit of traditional education by sailors. Some aspects of financial management skills could be measured by paper-and-pencil tests. For example, tests that measure an individual's ability to balance a checkbook, compute interest, and evaluate potential purchases, or risks in different types of investment could be constructed. Such a test could provide a reliable measure of a sailors understanding of financial management questions. Such a test of knowledge is likely to have relatively poor validity, however, if the goal is to make valid inferences about the quality of a sailor's day-to-day financial management practices because there is a difference between knowing and doing in areas such as financial management.

The health, wellness, and well-being dimension includes physical fitness, nutrition, and well-being. Physical fitness can be evaluated using standard exercise tasks, e.g., running a mile, pull-ups, or pushups. Physical well-being can also be evaluated through routine physical examinations. Consideration needs to given, however, to how information from physical fitness tests and from physical examinations would be combined in making judgments about overall health, wellness, and well-being. Consideration also needs to be given to how judgments about health, wellness, and well-being would be combined with information about the other constructs that make up the personal development vector.

In summary, with some exceptions, the technical quality of measures of the personal development vector is likely to be poor. Although the constructs associated with this vector represent important goals, it is questionable that cost-effective, valid, and reliable measures can be developed for more than a few aspects of the six general constructs that comprise the personal development vector.

Leadership and Professional Military Education

Leadership is of fundamental importance to the military mission. Some aspects of leadership are dictated by military doctrine. Other aspects of leadership have most commonly been assessed by various types of ratings. The use of ratings scales to measure leadership ability has a fairly lengthy, albeit somewhat controversial history. There is a large literature documenting threats to validity in the use of rating scales to measure characteristics such as leadership (see, for example, Landy and Farr, 1983). Ratings are frequently subject to halo effects, which is the tendency of raters to give similar ratings to a given individual on a wide range of putatively distinct dimensions of performance (e.g., leadership, perseverance, trustworthiness). As was mentioned earlier, ratings may also suffer from leniency sets, that is, the

tendency to give personnel being rated higher ratings than their work deserves. Simple four-point ratings of leadership where the four points of the scale are defined as *poor*, *satisfactory*, *exceeds expectations*, and *outstanding* or similar characterizations have frequently been found to be subject to both halo and leniency effects. Those effects have been attributed to the fact that such scales provide only limited guidance to raters about the nature of the characteristic being rated.

One of the more successful approaches to overcoming the limitations of rating scales for characteristics such as leadership is the behaviorally-anchored rating scale. As the name suggests, behaviorally-anchored rating scales use descriptions of specific behaviors for the scale points rather than general adjectives such as "poor" or "outstanding." Behaviorally-anchored rating scales were used with a good deal of success in the in the job performance measurement project conducted for the army in the 1980s (see, Campbell, McHenry, and Wise, 1987) for measuring characteristics such as leadership, effort, personal discipline, and military bearing.

The five constructs or dimensions specified under the leadership and professional military education vector suggest the aspects of leadership that need to be considered. As was previously noted, the five dimensions are (1) leading change, (2) working with people, (3) leading people, (4) resource stewardship, and (5) accomplishing mission. The components comprising these five dimensions are listed in Table 1. From a review of items listed in Table 1, it is apparent that the components are quite variable. Some suggest general traits (e.g., creativity or vision), others (e.g., financial management, or team building) suggest more specific knowledge, skills, and abilities, and still others (e.g., technical credibility or responsibility, authority, and accountability) involve general characteristics. One component, combat/crisis leadership, is not realistically measured during peacetime. Nonetheless, the components listed in Table 1 could be used to suggest lists of specific behaviors that that would distinguish persons with different levels of the components. The resulting list of specific behaviors could be used in the development of behaviorally-anchored rating scales.

Table 1

Dimensions of the Leadership and Professional Military Education Vector

Dimension	Component
Leading Change	• Creativity and innovation • External awareness • Flexibility • Service motivation • Strategic thinking • Vision
Working with People	• Influencing/negotiating • Oral communication • Partnering • Political awareness • Written communication
Leading People	• Developing people • Conflict management • Leveraging diversity • Professionalism • Team building • Combat/crisis leadership
Resource Stewardship	• Financial management Leveraging technology • Human resource management
Accomplishing Mission	• Responsibility, authority and accountability • Decisiveness/risk management Continuous improvement • Problem solving • Technical credibility

Certification and Qualifications

The certification and qualifications vector represents a combination of documentation of completed recruit and initial skills training in the service and commercial certifications such as the completion of college course or the receipt of an associate degree from a college. Although the compiling the information regarding certifications and qualifications is relatively straight forward, interpreting the information is another matter. College courses are variable in terms of content coverage and standards and an associate degree from one community college may

mean something quite different than the same degree for the same major from another college. Furthermore, information about an individual's level of professional development from the professional development vector or information about the quality of performance obtained under the performance vector may be inconsistent with the fact that the person has a particular certification.

Performance Continuum

The performance vector may overlap with the professional development vector. It is expected to include both training measures and on-the-job measures. Performance measures obtained during training would include both paper-and-pencil tests of the knowledge of the content of the training course and performance tasks that require trainees to demonstrate skills (e.g., repair a radio, disassemble and reassemble a piece of equipment, load, aim, and fire at a target) and abilities (e.g., read and interpret a radar screen, read and understand a circuit diagram) that the training is intended to develop.

On-the-job performance measures are likely to include both supervisor ratings and hands-on performance tasks. The issues discussed above relating to obtaining information from ratings that support valid inferences about performance apply to measures obtained for purposes of the performance vector. Steps need to be taken to minimize the threats to validity caused by halo effects and leniency tendencies. Behaviorally-anchored rating scales, while costly and time-consuming to develop, have clear advantages over scales that use general labels such as "poor" or "excellent."

Factors that affect the validity of inferences about the quality of performance from hands-on performance tasks were also discussed above. Of particular concern are difficulties in standardizing the difficulty of hands-on performance tasks. In some instances, simulations may be preferable to hands-on tasks because the advantages of standardization of task difficulty outweigh the disadvantages of loss of fidelity to the actual task that must be performed on the job.

Integrating Information from the Five Vectors

The five-vector model is intended to support a performance-based advancement system. Advancement decisions are basically dichotomous while the information from the five-vector model is continuous and multifaceted. It is likely

that the measurement information will vary in terms validity and reliability both across and within vectors. The profile of information is also likely to be uneven across vectors and the information from one vector may even conflict with that from another vector. Consequently, the tasks of combining and integrating the information across vectors in a way that will support the most valid inferences about a sailor's advancement potential will be quite challenging.

References

- Campbell, J., McHenry, J. & Wise, L. (1987). *The modeling of performance*. Paper presented at the mid-year conference of the Society of Industrial and Organizational Psychology, Atlanta, GA.
- Landy, F., & Farr, J. (1983). *The measurement of performance: Methods, theory, and applications*. New York: Academic Press.