Study Report 2007-05

U.S. Army Classification Research Panel: Conclusions and Recommendations on Classification Research Strategies

John P. Campbell & Rodney A. McCloy
Human Resources Research Organization

S. Morton McPhail
Valtera

Kenneth Pearlman
Independent Consultant

Norman G. Peterson
SPR Center

James Rounds
University of Illinois

Michael Ingerick
Human Resources Research Organization

United States Army Research Institute for the Behavioral and Social Sciences

May 2007

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

Barbara A Black — Michelle Sams

BARBARA A. BLACK **Acting Technical Director**	**MICHELLE SAMS** **Director**

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical Review by

Richard Hoffman, U.S. Army Research Institute
Peter M. Greenston, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Study Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Study Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Study Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) May 2007	2. REPORT TYPE Final Report	3. DATES COVERED (from. . . to) March 2006 - December 2006

4. TITLE AND SUBTITLE U.S. Army Classification Research Panel: Conclusions and Recommendations on Classification Research Strategies	5a. CONTRACT OR GRANT NUMBER DASW01-03-D-0015, DO #29
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Campbell, John P. & McCloy, Rodney A. (Human Resources Research Organization); McPhail, S. Morton (Valtera); Pearlman, Kenneth (Independent Consultant); Peterson, Norman G. (SPR Center); Rounds, James (University of Illinois); and Ingerick, Michael (Human Resources Research Organization)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization (HumRRO) 66 Canal Center Plaza, Suite 400 Alexandria, Virginia 22314	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) 2511 Jefferson Davis Highway Arlington, VA 22202-3926	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Study Report 2007-05

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representatives: Tonia Heffner and Peter Greenston
Subject Matter POC: John P. Campbell

14. ABSTRACT *(Maximum 200 words)*:

As the U.S. Army transforms to meet the needs of the future force, the importance of classifying recruits to entry-level jobs will only increase, as will research to enhance the classification process (e.g., development and validation of non-cognitive predictors). A critical component to ensuring the success of this research, and its implementation, is having meaningful and reliable criterion data. Since the late 1980s, however, collecting criterion data for a sufficient number of jobs to meet the Army's classification research needs has proven a challenge. To find solutions to this challenge, the U.S. Army Research Institute (ARI) contracted with the Human Resources Research Organization (HumRRO) to convene a six-member Classification Research Panel. Overall, the Panel recommended that the solution to this challenge ultimately rests on (a) a solid, job analysis system, (b) a method for generalizing (or transporting) validity information across MOS (i.e., for the purposes of estimating classification efficiency for the entire system), and (c) a supporting relational database that collects and stores occupational/job analysis and other relevant data (e.g., criterion-related validity estimates) over time. This report summarizes the Panel's recommendations. It concludes with a roadmap outlining a comprehensive, long-term solution to the Army's criterion challenge.

15. SUBJECT TERMS

Personnel selection and classification; Occupational/job analysis; Generalizing validity; Synthetic validity; Criterion measurement; Classification efficiency; Criterion-related validation.

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	64	Ellen Kinzer Technical Publications Specialist (703) 602-8047

Standard Form 298

Study Report 2007-05

U.S. Army Classification Research Panel: Conclusions and Recommendations on Classification Research Strategies

John P. Campbell & Rodney A. McCloy
Human Resources Research Organization

S. Morton McPhail
Valtera

Kenneth Pearlman
Independent Consultant

Norman G. Peterson
SPR Center

James Rounds
University of Illinois

Michael Ingerick
Human Resources Research Organization

Selection and Assignment Research Unit
Michael G. Rumsey, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

May 2007

Army Project Number	**Personnel Performance**
622785A790	**and Training Technology**

ACKNOWLEDGEMENTS

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and the Human Resources Research Organization (HumRRO) wish to thank the members of the Army Classification Research Panel (ACRP):

John Campbell, *University of Minnesota*
Rod McCloy, *Human Resources Research Organization (HumRRO)*
S. Morton McPhail, *Valtera*
Kenneth Pearlman, *Independent Consultant*
Norman Peterson, *SPR Center*
James Rounds, *University of Illinois*

All members contributed equally to this report. Michael Ingerick (HumRRO) served as lead coordinator for the ACRP and edited this report.

U.S. ARMY CLASSIFICATION RESEARCH PANEL: CONCLUSIONS AND RECOMMENDATIONS ON CLASSIFICATION RESEARCH STRATEGIES

EXECUTIVE SUMMARY AND RECOMMENDATIONS

Background

For the Army, classifying recruits into entry-level jobs represents an essential personnel management function. As the Army transforms to meet the needs of the future force, the importance of classifying recruits to entry-level jobs will increase, as will research critical to enhancing the classification process (e.g., development and validation of non-cognitive predictors, revisions to the existing Armed Services Vocational Aptitude Battery [ASVAB]). A critical component to ensuring the success of this research, and its implementation, is having meaningful and reliable criterion data.

Since the late 1980s, however, collecting criterion data for a sufficient number of jobs to meet the Army's classification research needs has proven a challenge. To find solutions to this challenge, the U.S. Army Research Institute (ARI) contracted with the Human Resources Research Organization (HumRRO) to convene a six-member Classification Research Panel of experts in the areas of personnel selection/ classification, occupational/job analysis, job clustering, criterion measurement, and psychometrics. The Panel's mission was to make recommendations addressing how the Army should

- Obtain criterion data for a sufficient number of MOS in an on-going, systematic fashion to support Army classification research.

- Ensure that the differential validity of new predictors, once established, can be generalized (or transported) to other MOS in the same job family.

Meeting the Army's Needs: Conclusions and Recommendations

Meeting the Army's needs for criterion data is a complex matter. Overall, the Panel concluded that the solution ultimately rests on

- A solid, job analysis system.
- A method for generalizing (or transporting) validity information across MOS (i.e., for the purposes of estimating classification efficiency for the entire system).
- A supporting relational database that collects and stores occupational/job analysis and other relevant data (e.g., criterion-related validity estimates, Soldier-level predictor and criterion data) over time.

Consistent with these conclusions, the Panel made the following recommendations.

Occupational/Job Analysis

Recommendation 1: An Army-specific job analysis system, supported by a relational database for systematically storing and organizing job data, is needed. Among other features, this system should

- Use a common language, customized to the Army context for describing similarities and differences in MOS.
- Consist of a master library of descriptors representing targeted work- and worker-oriented domains critical to the Army's classification research needs, and sufficient for describing any MOS. Specifically:
  - Performance requirements (at a minimum, defined at two levels of specificity)
  - Work/job context
  - Machine-tools-equipment-technology
  - Occupation-specific knowledges and skills
  - Abilities
  - Personal characteristics (specifically interests, values, and temperament)
- Include cross-MOS descriptors (i.e., descriptors that can be applied across MOS) for use in making comparisons and linkages across MOS.
- Specify descriptors, in particular performance requirements, at varying levels of generality that can be organized hierarchically to support the Army's needs for job information at multiple levels of aggregation.

Recommendation 2: Where advisable, investigate the potential for describing MOS in a new way(s) that sufficiently captures cross-MOS differences, and does so in a more efficient and cost-effective manner than might otherwise be possible using an existing system (e.g., O*NET's Detailed Work Activities; for examples, see Appendix A).

Recommendation 3: Similarly, where advisable, investigate the potential for using linkages among descriptors to generalize job data collected from one descriptor to others, as a means to maximize the Army's return from the effort and resources expended. One feasible possibility, and one for which there is evidentiary support, is with performance requirements and interests. This would require additional research, conceivably following the pilot testing and refining of a prototype job analysis approach.

Recommendation 4: Pilot work to develop and refine the proposed job analysis system, as outlined above, is needed and should receive the highest priority, as should construction of the supporting database. For this pilot, the Army need not start from "scratch." Existing descriptor taxonomies from one or more of these systems could inform the development of taxonomies for the proposed system, as could past job analysis work conducted for the Army (e.g., SYNVAL, PerformM21, Select21). Once successfully piloted, the next step would be to populate the database by collecting data on a larger sample of MOS.

Recommendation 5: Specifying non-technical performance requirements would be greatly facilitated by using pre-specified taxonomies to stimulate SMEs' formulation and assessment of these requirements. Candidate taxonomies, or the information needed to improve upon them (e.g., to make them more Army specific), exist and can be found from (a) research on critical incidents (i.e., where non-technical requirements are demonstrated or called for) and Army-wide performance dimensions, (b) the leadership and team literatures, and/or (c) existing job analysis taxonomies and instruments (e.g., PAQ). Similar work to develop or refine existing taxonomies on the predictor side (i.e., interests, values, temperament) would also be useful in this regard.

Recommendation 6: Except for differences in the content used to prompt SMEs, the specification of non-technical performance requirements should follow the same approach as the specification of technical requirements, unless pilot work suggests otherwise. If this is the case, carefully considered modifications to traditional approaches or the use of alternative analysis approaches (e.g., role-based job analysis, team task analysis) could prove useful. Because of this, a flexible approach should be taken in specifying non-technical performance requirements such that these requirements can be specified differently, as needed.

Generalizing (or Transporting) Validity

Recommendation 7: An approach to generalizing (or transporting) validity that is empirically based (in some form), and linked to the recommended job analysis database, should be employed.

Recommendation 8: Several specific approaches for generalizing (or transporting) validity information were identified that could meet the Army's needs. They were:

- A full validity (or test) transportability approach.
- A full hierarchical linear modeling (HLM) approach.
- A combined validity (or test) transportability-HLM approach.
- An incremental, rational synthetic validity-validity transportability-HLM approach.
- Standard job component validation (JCV) approach.

The Army need not a make final decision on which approach to pursue at this point in time. Because the first four approaches operate on and make use of the same data (i.e., from 20-30 criterion-related validation studies), they could be pursued and tested simultaneously, provided sufficient resources are available.

Job Clustering

Recommendation 9: Priority should be placed on solutions that systematically cluster MOS, either separately or jointly (e.g., a multi-tier solution), on the basis of performance requirements and select KSAO descriptors. Because "KSAOs" can cover a wide range of descriptors, great care needs to be paid to the specification and selection of KSAO descriptors for use in clustering. To start, KSAOs should at least be partitioned into three predictor domains: (a) occupation-specific knowledges and skills (KSs), (b) specific abilities (As), and (c) interests/values/temperaments (Os). Solutions that consider both performance requirements and select KSAO descriptors simultaneously could prove to be particularly advantageous. Multiple solutions should be generated and compared, as data become available, so that the evaluation of which solution works best for meeting the aforementioned purposes can be examined empirically.

Recommendation 10: To meet the Army's needs, clustering solutions having an empirical basis (even if supplemented by expert judgments), systematically derived and supported by job analysis data, should be employed. Because the collection of the data needed to validate these solutions (i.e., predictor score profiles, criterion-related validity estimates) will take time to accumulate, the following interim approach is recommended as a starting point:

- Generate an initial cluster solution using general performance requirements-based descriptor scores, collected from the job analysis.
- Use ARI and other psychologists to rate MOS on select KSAOs (e.g., specific abilities, interests, values, temperament) to provide an initial database of scores on these descriptors.
- Provided no validity estimates are available, examine predictor score profiles for each cluster to obtain information on the (a) differences between MOS within and across clusters, and (b) integrity of the clusters and the predictor-based profiles.

Criterion Measurement

Recommendation 11: Using Army-specific job analysis data, the Army should pursue (a) strategies for collecting adequate criterion data for a sufficient sample of MOS and (b) development of criterion measures, or refinement of existing ones, that sufficiently differentiate across MOS.

Recommendation 12: The Army should consider administering a complete set of criterion measures (e.g., JKT, ratings, retention) to focal MOS (i.e., those MOS most representative of a cluster), while administering a reduced set of criteria to non-focal MOS. Decisions on which MOS are focal and which criterion measures to include would best be guided by Army-specific job analysis data, MOS clustering results, Army priorities, and existing theory on predictor-criterion relations.

Recommendation 13: The Army should pursue the use of end-of-training criteria, particularly knowledge tests and peer (and possibly instructor) ratings. Further, the Army should continue to assess the relations between end-of-training criteria and post-training criteria measuring the same, or similar, criterion dimensions.

Recommendation 14: Using Army-specific job analysis data and the results of the MOS clustering as recommended earlier, the Army should explore the feasibility of mid-range criterion tests (or test components), specifically for end-of-training tests.

Recommendation 15: Should the preceding recommendation prove infeasible, the Army-specific job analysis data could be used to maximize the resources used for developing end-of-training knowledge tests. For example, following a "top down" approach to criterion development, the performance requirements taxonomy developed as part of the proposed job analysis system could serve as a general test plan template and the MOS-specific data as a kind of weighting scheme for the general plan. Doing so would enable more incremental development approaches where similarities in test content can be seen, and capitalized on, ahead of time. Alternatively, the job analysis data could be used to weight existing end-of-training criterion tests to enhance their validity.

Recommendation 16: The Army should pursue the use of behaviorally anchored job performance ratings. To minimize halo (and other biases) and maximize the construct validity of these ratings, the Army should (a) specify the performance dimensions to be rated as clearly and distinctly as possible (i.e., so that the scales can be explicitly distinguished from each other), (b) provide raters with the best available training, (c) standardize the rating process to promote consistent implementation across raters and ratees, and (d) ensure that those providing ratings (supervisors and/or peers) have had sufficient opportunity to observe a Soldier.

Recommendation 17: Having Army-specific job analysis data is essential, as they would greatly facilitate (a) the discovery, selection, and specification of MOS-specific and cross-MOS performance dimensions, technical and non-technical, to be assessed by ratings; and (b) the development of experimental, alternative rating formats (and other assessment methods) that provide more realistic and meaningful operationalizations of non-technical performance dimensions in ways that partial out technical performance requirements (e.g., least preferred co-worker scale).

Recommendation 18: When validating and establishing the classification potential of non-cognitive predictors, the Army should employ (a) ratings of MOS-specific and cross-MOS non-technical performance dimensions, and (b) occupational and organization retention-related criteria.

Recommendation 19: Although objective retention and attrition criteria have been and can be highly inaccurate (i.e., if relied on exclusively without consideration of other measures), research could be conducted to render them useable for validation purposes. Doing so, however, would require a significant initial effort either to shape up the official coding for Soldiers' reasons for staying-leaving, or to devise a method to recode those reasons reliably and accurately. Alternatively, the Army could pursue new, alternative possibilities for collecting reasons (e.g., exit surveys) that could then be instituted and stored for future validation work.

Estimating Classification Efficiency

Recommendation 20: To empirically estimate and evaluate the potential classification gains for the entire system accruing from the use of new, alternative predictor batteries (e.g., consisting of new ASVAB subtests or measures of non-cognitive predictors), collect criterion-related validity estimates for a sufficiently representative clustering of MOS (20-30 clusters), specifically estimates from at least one focal MOS in each cluster. These validity estimates need not be obtained in a single study, but can be collected and accumulated over time. Such an incremental approach permits the successive refinement of previously derived estimates of classification gains, and the prediction equations on which they are based, as more data become available.

Recommendation 21: When estimating a predictor battery's classification efficiency, careful consideration needs be paid to the sampling of MOS in the criterion-related validations studies on which the estimates will be based, and the implications this sampling carries for inferences drawn from these estimates. Having job analysis data, as recommended, to cluster MOS and to identify focal MOS, would be useful in this regard.

Recommendation 22: To understand the impact of sample size on estimates of classification efficiency, and its implication for drawing conclusions, make use of formula and/or Monte Carlo-based approaches for modeling error in key parameters (e.g., prediction equations). For an example, see Rosse, J. P. Campbell, and Peterson (2001).

Recommendation 23: When estimating predicted criterion scores, make use of data on multiple predictors-criteria to obtain more accurate estimates of Soldiers' actual performance/ satisfaction. This can be accomplished by modeling relations among criteria and/or predictors when advisable (i.e., the interrelations reflect systematic and theoretically-relevant sources of variance), and incorporating these interrelations when estimating Soldiers' predicted criterion scores.

Recommendation 24: For the purposes of choosing which predictor battery (or batteries) offers the greatest potential to enhance classification, make use of analytic solutions (e.g., Horst, 1954; Sager, Peterson, Oppler, Rosse, & Walker, 1997), or explore alternative to these solutions, to investigate differential validity and to diagnose potential classification efficiency.

Recommendation 25: When validating and investigating the classification potential of non-cognitive predictors, the Army should, at a minimum, include (a) criterion measures assessing non-technical, "will do" performance dimensions and (b) non-performance criteria (e.g., MOS satisfaction, P-O fit, retention, attrition). Regarding the latter, careful consideration needs to be paid to the nature of the method used. For example, because the effects of non-cognitive predictors on objective retention (or attrition) criteria are indirect, such criteria cannot be relied upon exclusively when estimating classification efficiency (i.e., mediators or moderators need to be modeled as well). Otherwise, one is likely to underestimate the classification potential of non-cognitive predictors.

Recommendation 26: When using multiple criteria, a critical issue will be how to treat the multiple, and potentially competing, goals underlying these different criteria (e.g., increased technical performance, increased non-technical performance, greater retention) in the optimization process (i.e., for purposes of estimating classification efficiency). Research investigating multi-stage or multi-track classification models would be useful in this regard, as would policy capturing studies to scale the relative value to the Army of gains on each criterion. One solution to this would be to start by specifying the desired levels of gain (i.e., from use of non-cognitive predictors over and above the ASVAB) that are practically significant to the Army and then determine the relative weighting that would best achieve such gains.

Towards a Comprehensive Solution: An Agenda and Roadmap

Based on the most critical recommendations, the Panel proposed a near-term agenda and roadmap for solving the criterion challenge. According to the Panel, the activities requiring the Army's most immediate attention and resources are:

- Piloting an Army-specific job analysis approach on 3-5 MOS.
- Constructing and populating a supporting relational database to collect and organize job analysis data systematically, along with other relevant personnel research data over time and on an on-going basis.

To facilitate the execution of these activities, the Panel outlined the major tasks and steps to be taken and what should result from their completion.

U.S. ARMY CLASSIFICATION RESEARCH PANEL: CONCLUSIONS AND RECOMMENDATIONS ON CLASSIFICATION RESEARCH STRATEGIES

CONTENTS

List of Tables

List of Figures

U.S. ARMY CLASSIFICATION RESEARCH PANEL: CONCLUSIONS AND RECOMMENDATIONS ON CLASSIFICATION RESEARCH STRATEGIES

Introduction

Why the Army Classification Research Panel Was Formed

For the Army, classifying recruits into entry-level jobs represents an essential personnel management function. The effective matching of recruits to jobs benefits both the Army and the individual Soldier (Lightfoot & Ramsberger, 2000; Rosse, J. P. Campbell, & Peterson, 2001; Zeidner & Johnson, 1994; Zeidner, Johnson, & Scholarios, 1997). For the Army, classification can reduce training costs, minimize first-term attrition, increase job performance, and promote retention. For the individual Soldier, it ensures placement into jobs that best emphasize their abilities, knowledge, skills, interests, and potential. As the Army transforms to meet the needs of the future force, the importance of classifying recruits to entry-level jobs will increase, as will research critical to enhancing the classification process (e.g., development and validation of non-cognitive predictors, revisions to the existing Armed Services Vocational Aptitude Battery [ASVAB]). An essential component to ensuring the success of this research, and its implementation, is having meaningful and reliable criterion data.

Since the end of the Skill Qualification Test (SQT) program in 1989, however, collecting criterion data – in particular job-specific performance data – has been challenging. There is presently no large-scale, operational program for collecting job-specific criterion data on a regular, systematic basis. Consequently, over the last 15-plus years, the Army's criterion collection efforts have been driven by discrete research projects where collecting criterion data for even a small number of Military Occupational Specialties (MOS) has proven difficult. Because of the large number and diversity of entry-level jobs, and the difficulty and expense of collecting criterion data for a sufficiently representative sample of these jobs, collecting the criterion data needed to support the Army's classification research program will continue to pose a challenge.

To find solutions to this challenge, the U.S. Army Research Institute (ARI) contracted with the Human Resources Research Organization (HumRRO) to convene a six-member Classification Research Panel of experts in the areas of personnel selection/classification, occupational/job analysis, job clustering, criterion measurement, and psychometrics. The Panel's mission was to generate innovative, scientifically sound, and technically feasible recommendations for solving this challenge. More specifically, these recommendations would address how the Army should

- Obtain criterion data for a sufficient number of MOS in an on-going, systematic fashion to support Army classification research.

- Ensure that the differential validity of new predictors, once established, can be generalized (or transported) to other MOS in the same job family.

The Research Panel met formally two times over a 6-month period. During and between those meetings, the Panel reviewed how the Army and the other Services currently classify recruits to jobs; formulated and discussed possible strategies to meet the Army's needs; and developed recommendations for a technically sound and feasible approach to a comprehensive classification research and development program.

Overview of Research Panel Report

This report presents the major conclusions and recommendations of the Army Classification Research Panel. The report is organized as follows. First, a brief overview of the Panel's recommendations is presented. Second, the conclusions and recommendations generated by the Panel, organized by focus, are summarized. The report concludes with a near term agenda and roadmap for implementing the Panel's most critical recommendations.

Meeting the Army's Needs: An Overview of the Panel's Recommendations

Meeting the Army's needs for criterion data is a complex matter. Overall, there are several general approaches to solving this challenge:

- Use existing criterion measures "as is."
- Base the choice of existing criterion measures on previous selection and classification research.
- Use occupational/job analysis data to expand on, or refine, existing criterion dimensions and measures.
- Use occupational/job analysis data to define the criterion space and then develop, or select, relevant criterion measures of targeted dimensions. Once relevant criterion measures have been developed (or selected), find ways to institutionalize them.

These approaches differ in their technical soundness, feasibility, and the resources required to implement them, with the latter approaches being more technically sound but more costly. For the Army, it is essential that the proposed solution(s) effectively balance these two goals.

Consistent with this imperative and the need for a comprehensive solution, the Panel considered a number of critical issues and generated recommendations encompassing the core "building blocks" of a personnel classification research program:

- Occupational/job analysis
- Generalizing (or transporting) validity
- Job clustering
- Criterion measurement
- Estimation of classification efficiency

In addition, because of its special importance in the Army's classification research agenda, the Panel considered the implications that the use of multiple criterion dimensions and non-cognitive predictors would have on these recommendations.

Figure 1 provides an overview of the Panel's recommendations. The figure shows that, in the Panel's estimation, the ultimate keys to meeting the Army's needs are (a) a solid, job analysis system and (b) a method for generalizing (or transporting) validity information across MOS (i.e., for the purposes of estimating classification efficiency for the entire system). In particular, having job analysis data is critical and represents an essential first step. As the figure illustrates, these data are critical because they underlie and supply information needed to implement other strategies (i.e., clustering of MOS, criterion measurement, estimating classification efficiency) useful in supporting and advancing the Army's personnel management objectives.[1]

Supporting these strategies is a relational database. The primary purpose of this database is to capture and store the job analysis data needed to support and advance the Army's personnel management objectives, in particular classification, in an ongoing fashion. When combined with other relevant data (e.g., criterion-related validity estimates, Soldier predictor-criterion data), these job data can be used to generate and refine solutions to the Army's classification needs. Populating this database would follow an incremental approach; that is, the Army would start with existing information, then update or supplement that information as new data are available. As the Army's experiences demonstrate, obtaining all the needed criterion data will be difficult in the context of a single study. Because of this, having this database would enable the Army to refine and optimize one or more of these strategies successively over time, as more data become available. This incremental approach balances demands on Army resources while providing the Army with sound and viable solutions to its classification research needs.

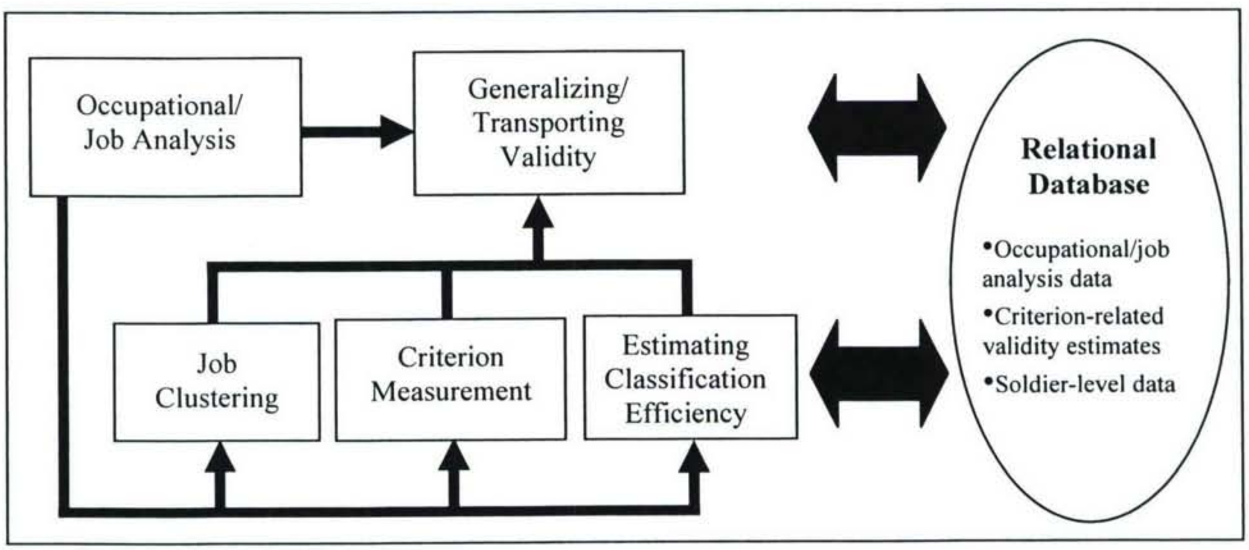

Figure 1. Overview of the Panel's recommendations.

[1] The Army currently maintains an occupational analysis program for collecting job analysis data. It should be understood that the proposed work is not intended to replace, but rather to build on the system currently in use. Consistent with this intention, and with the Army's objectives for the Panel, the recommendations outlined in this report aim to make the existing system more efficient, more flexible, easier to maintain and upgrade, and more directly useful for personnel classification.

Occupational/Job Analysis

<u>*Issue:*</u>

How critical is having job analysis data to supporting the present and future needs of an Army classification research program?

Ultimately, no matter the criterion measure(s) used, conducting the needed criterion-related validation studies to support the Army's classification program will be a large and resource intensive undertaking.[2] Because of this, the Army needs to be in a position to maximize the investments made in its classification research program. Having an appropriate job analysis system (and data) is absolutely essential in this regard. Most critical to the Army's needs, it requires a job analysis system that would enable (a) the discovery and specification of critical MOS-specific and cross-MOS criterion dimensions useful for differentiating MOS and (b) the empirical estimation and determination of the limits of the generalizability of cross-MOS dimensions (i.e., for purposes of transporting or generalizing validity information). Having such a job analysis system, and the information it provides, has proven critical to other large-scale organizations facing the same challenge (i.e., a large number of jobs and insufficient resources for collecting criterion data for all jobs).

Consistent with this, the Panel concluded

<u>*Conclusion:*</u> Although there are multiple ways to approach the problem and focusing on criterion measures may seem the most desirable, the ultimate key to meeting the Army's needs rests on a solid, job analysis system, and a supporting database for organizing and storing critical job information over time. Such a system would provide the Army with the data necessary to meet its needs. Specifically, it would enable the Army to

- Cluster MOS to support the sampling of MOS for criterion-related validation studies and for estimating and determining the limits of generalizing (or transporting) validity information across MOS.
- Generalize, or transport, validity information for a sample of MOS to other MOS for purposes of selecting predictor batteries that maximize classification efficiency.
- Demonstrate the relevance of, or refine, existing criteria for use in criterion-related validation studies.
- Develop criteria that target critical MOS-specific and cross-MOS dimensions useful for differentiating MOS.
- Document changes in MOS over time and their implications for the use (or continued use) of previously collected validity information.

[2] The number of criterion-related validation studies needed to ensure sufficient representation of the population of Army MOS and to achieve a maximal level of classification efficiency (however optimized) is likely to be upwards of 20-30.

Issue:
To meet the Army's needs, what are the essential features and characteristics that a job analysis system should exhibit?

To meet the Army's present and future needs for its classification research program, a specific job analysis system is required. The essential features and characteristics of such a system are summarized in Table 1.

Table 1. Essential Features and Characteristics of an Army Job Analysis System

- ✓ Uses a common language, customized to the Army context for describing similarities and differences in MOS.
- ✓ Consists of descriptors representing a targeted set of work- (i.e., performance requirements, work/job context, machine-tools-equipment-technology) and worker-oriented (i.e., KSAOs) domains critical to the Army's classification research needs.
- ✓ Includes descriptors that are relevant across MOS, such that a particular requirement can appear in multiple MOS *if* the MOS do in fact share a similar requirement (at some level of generality).
- ✓ Specifies descriptors, in particular performance requirements, at varying levels of generality arranged hierarchically, and in accordance with well-defined rules.
- ✓ Where advisable, maximizes (valid) linkages among descriptors representing different domains so that job analysis data collected from one domain can be generalized to others.
- ✓ Built on descriptor taxonomies developed, or refined, using a combined top-down and bottom-up approach.[3]
- ✓ Supported by a relational database that systematically organizes and stores job analysis data, and permits the linkage of these data to other critical information (e.g., criterion-related validity estimates).
- ✓ Is automated and easy to use, particularly by subject matter experts (SMEs).
- ✓ Follows an incremental approach, whereby job analysis data are periodically collected and updated over time (as needed).

In general, the work- and worker-oriented domains most critical to the Army's classification needs, and which the job analysis system should target, are summarized in Table 2. This list is not intended to imply that the Army needs to measure all domains from the start. Should priorities need to be set, performance requirements should receive the highest priority, followed by work/job context, personal characteristics (interests/values/temperament), occupation-specific knowledges and skills, machine-tools-equipment-technology, and abilities. Further,

[3] The recommended approach for developing, or refining, these descriptor taxonomies is covered in greater detail later in this report (see pp. 36-39).

Table 2. Work- and Worker-Oriented Descriptor Domains Critical to the Army's Classification Research Needs

Domain	Definition/Specifications	Representative Examples
Performance Requirements	• Behaviorally based descriptions of what an incumbent in an MOS does and potentially what gets done as a result (i.e., outputs). • At a minimum, defined at two levels of generality.	*General Requirements* • Operates and maintains a motor vehicle • Transports cargo and personnel *Specific Requirements* • Loads/unloads passengers for transport in truck. • Performs tiedown procedures.
Work/Job Context	• Descriptions of the context in which the performance requirements take place. • Could be defined at multiple levels of specificity. • Could be defined in conjunction with performance requirements.	• Situational constraints (e.g., time pressure) • Physical conditions • Trainability of occupation-specific knowledges and skills
Machine-Tools-Equipment-Technology	• Descriptions of the machine(s), tools, equipment, and/or technology used to execute the performance requirements. • Could be specified in conjunction with performance requirements.	• M16A2 rifle • .50 caliber machine gun • Hoist • Wrench • Air compressor
Occupation-Specific Knowledges and Skills	• Descriptions of the occupation-specific knowledges and skills required to successfully execute one or more performance requirements. • Could be defined at multiple levels of specificity. • Could be defined in conjunction with performance requirements.	• Close combat • Basic electronic design and repair • Basic mechanical knowledge and repair
Abilities	• Descriptions of the abilities required to successfully execute one or more performance requirements and to persevere in the MOS. • Definitions would include information on level of complexity (high, medium, low) needed.	• Cognitive ability • Physical ability • Psychomotor ability
Personal Characteristics	• Descriptions of the personal characteristics required to successfully execute the performance requirements and to persevere in the MOS. • Could be defined at multiple levels of specificity.	• Interests • Values • Temperament

it may be possible that the specification of one or more of these descriptors domains could be combined in new, creative ways that both maximizes resources and results in multi-faceted descriptors that prove more useful for meeting the Army's needs than traditional descriptors. For example, work/job context, machine-tools-equipment technology, and occupation-specific knowledges and skills can be viewed as natural extensions of performance requirements and thereby could be represented in, or incorporated into, an expanded specification of performance requirements that includes these domains (e.g., O*NET's Detailed Work Activities; for examples, see Appendix A).

Taken as a whole, these features would enable the Army to meet its classification research needs, as previously stated. In addition, having such a system would enable the Army to describe MOS in new, creative ways that sufficiently capture cross-MOS differences – and do so in a more efficient and cost-effective manner than might otherwise be possible using an existing system. Similarly, by taking advantage of (valid) linkages among the selected descriptors, the Army could generalize job data collected from one set of descriptors to others lacking data. As discussed in regard to the next issue, several existing job analysis systems exhibit one or more of the aforementioned features, but none does so as a whole.

Consistent with the preceding discussion, the Panel recommended the following:

Recommendation 1: An Army-specific job analysis system, supported by a relational database for systematically storing and organizing job data, is needed. Among other features, this system should

- Use a common language, customized to the Army context for describing similarities and differences in MOS.
- Consist of a master library of descriptors representing targeted work- and worker-oriented domains critical to the Army's classification research needs, and sufficient for describing any MOS. Specifically:
    - Performance requirements (at a minimum, defined at two levels of specificity)
    - Work/job context
    - Machine-tools-equipment-technology
    - Occupation-specific knowledges and skills
    - Abilities
    - Personal characteristics (specifically interests, values, and temperament)
- Include cross-MOS descriptors (i.e., descriptors that can be applied across MOS) for use in making comparisons and linkages across MOS.
- Specify descriptors, in particular performance requirements, at varying levels of generality that can be organized hierarchically to support the Army's needs for job information at multiple levels of aggregation.

Recommendation 2: Where advisable, investigate the potential for describing MOS in a new way(s) that sufficiently captures cross-MOS differences, and does so in a more efficient and cost-effective manner than might otherwise be possible using an existing system (e.g., O*NET's Detailed Work Activities; for examples, see Appendix A).

Recommendation 3: Similarly, where advisable, investigate the potential for using linkages among descriptors to generalize job data collected from one descriptor to others, as a means to maximize the Army's return from the effort and resources expended. One feasible possibility, and one for which there is evidentiary support, is with performance requirements and interests.[4] This would require additional research, conceivably following the pilot testing and refining of a prototype job analysis approach.

Issue:
Do current job analysis methods used by the Army to design jobs and training programs support the present and future needs of a classification research program? If not, would existing job analysis systems outside of the Army, military or civilian, meet these needs?

The Army currently has methods in place operationally for analyzing and collecting MOS-specific job information, specifically task requirements. At present, this information is primarily collected for training purposes (i.e., Advanced Individual Training [AIT]). Although informative, there are several limitations with these data as currently collected. First, because of the focus on technical training, the information collected focuses almost exclusively on technical tasks. Thus, information on non-technical tasks, select KSAOs, or other relevant descriptors is not collected, at least not systematically and on a recurring basis. Second, these tasks are specified in great detail, often resulting in large task lists of hundreds of tasks for each MOS. Such detailed lists make cross-MOS comparisons (e.g., for the purposes of clustering MOS on the basis of task similarity) difficult. Finally, the process for deriving tasks follows an inductive approach that frequently varies across MOS, further making these cross-MOS comparisons difficult. Although existing information could be useful at some level, the current system (and data) does not appear sufficient to meet the Army's classification research needs.

An alternative would be to make use of an existing job analysis system outside the Army. A number of standalone systems exist and can be found in the other Services or in the civilian sector (e.g., O*NET, PAQ, CMQ).[5] In addition to offering a template for collecting and analyzing MOS, several of the civilian job analysis systems have databases that contain information on thousands of jobs, and in some cases (e.g., PAQ, CMQ), prediction equations that can be used to generalize (or transport) validity to other jobs. Although making use of an existing system has its advantages, none of these systems as a whole meets all the features recommended by the Panel to meet the Army's needs. For example, the job analysis systems in place in the other Services (e.g., U.S. Air Force's CODAP) are similarly focused almost exclusively on technical tasks and at a low level of generality. Many of the major job analysis systems developed outside of the Services (e.g., O*NET, PAQ, CMQ) are more comprehensive (i.e.,

---

[4] In brief, linking these two domains would involve the following: First, starting with a well-developed taxonomy of occupational interests, have incumbents rate the performance requirements on the interests constituting the taxonomy. Second, have subject matter experts (SMEs) rate the whole MOS on the interests. Third, analyze linkages resulting from the two sets of ratings and then formulate a procedure for generalizing, or extending, data collected for performance requirements to interests based on these linkages. For evidentiary support for such an approach, see Prediger (1982); Prediger and Swaney (2004); Rounds, Smith, Hubert, Lewis, & Rivkin (1999). The same procedure could be applied to other worker-oriented domains (e.g., temperament), although in some cases the evidentiary support for doing so is currently less extensive.

[5] O*NET = Occupational Information Network (cf. Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1999). PAQ = Position Analysis Questionnaire (cf. McCormick, Jeanneret, & Mecham, 1972). CMQ = Common Metric Questionnaire.

encompassing descriptors across a range of work and worker-oriented variables) and reflect one or more of the features recommended. However, a number of limitations to using these systems (or components of these systems) in their existing formulations remain. These limitations include, but are not limited to, (a) an insufficient representation of military-related tasks or other occupationally related information specific to a military context, and the Army in particular; and (b) the use of cross-job descriptors (e.g., the O*NET's Generalized Work Activities [GWAs]) that are too general for sufficiently differentiating across jobs in general, and Army MOS specifically.

Making use of existing job analysis data found in databases populated using one of these systems is similarly problematic. For one, doing so would require the Army to either (a) "buy in" to one or more of the existing descriptors constituting these systems, and their aforementioned limitations; or (b) develop a method for otherwise matching (or equating) Army and civilian jobs, and a procedure for evaluating the accuracy of this matching, both of which have historically proven difficult. Further, any criterion-related validity information contained in these databases, and prediction equations for generalizing this information to other jobs, will be incomplete. This is because both are limited to cognitively oriented predictors (e.g., general mental ability). Missing from these databases is validity information for the kinds of non-cognitive predictors (i.e., interests, values, temperament) of current interest to Army.

Thus, the Panel concluded the following:

Conclusion: As a standalone system, no single, existing job analysis system as a whole, military or civilian, will be sufficient for the Army's needs. Similarly, linking to or otherwise using existing job analysis data from civilian databases (e.g., PAQ, CMQ, O*NET) is not advisable and is likely to prove more costly in the long run.

Accordingly, the Panel made the following recommendation:

Recommendation 4: Pilot work to develop and refine the proposed job analysis system, as outlined above, is needed and should receive the highest priority, as should construction of the supporting database. For this pilot, the Army need not start from "scratch." Existing descriptor taxonomies from one or more of these systems could inform the development of taxonomies for the proposed system, as could past job analysis work conducted for the Army (e.g., SYNVAL, PerformM21, Select21). Once successfully piloted, the next step would be to populate the database by collecting data on a larger sample of MOS.

Issue:

Does the validation of non-cognitive predictors (i.e., interests, values, and temperament) raise special considerations and implications for the design and conduct of job analysis?

Capturing critical cross-MOS differences in non-technical performance requirements (e.g., peer leadership, teamwork) carry value for classification and in particular for determining the differential validity of non-cognitive predictors (cf. J. P. Campbell, Russell, & Knapp, 1993; Rosse et al., 2001; Motowidlo & Van Scotter, 1994; Murphy & Shiarella, 1997; Wise, McHenry,

& J. P. Campbell, 1990). Thus, their measurement should be pursued. Until recently, however, non-technical performance requirements and their relations to non-cognitive predictors (i.e., interests, values, and temperament) have been largely overlooked in job analysis efforts, both in military and non-military settings (Raymark, Schmit, & Guion, 1997; Hogan, Davies, & Hogan, in press). This raises questions as to whether establishing the differential validity and classification potential of non-cognitive predictors requires special considerations in the specification of non-technical performance requirements and their linkages to non-cognitive predictors.

Recently, taxonomies of select work- and worker-oriented descriptors relevant to the non-cognitive domain have been developed, as have instruments specifically targeting these descriptors (cf. Knapp & R. C. Campbell, 2006; Raymark et al., 1997; Hogan et al., in press; Peterson et al., 1999; Sager, Russell, R. C. Campbell, & Ford, 2005). These existing taxonomies (and instruments), or modified versions that seek to improve upon them, should prove useful in eliciting and capturing the information needed for specifying (a) non-technical performance requirements in Army MOS, (b) the linkages between these requirements and non-cognitive predictors, and (c) cross-MOS differences in (a) and (b), provided such differences exist. Therefore, there will (and should) be differences in the content used to elicit information from SMEs in the non-technical versus the technical domain. However, it is, and should remain, an open question as to whether the approach for specifying requirements will, or should, differ across the two domains. For example, the emergent type of non-technical job descriptor generated from SMEs, particularly when using a bottom-up approach (e.g., critical incidents or situation descriptions), could be considerably different from that representative of technical performance requirements. Although traditional approaches for specifying technical performance requirements – at least as currently constructed and applied – may not be sufficient for capturing all substantive, cross-MOS differences in non-technical requirements, it is premature to presume that (a) alternative analysis approaches would prove more effective, or (b) that traditional approaches, with carefully considered modifications, would not suffice for capturing these differences.

In sum, the Panel concluded

Conclusion: Specifying non-technical performance requirements will require different content to prompt SMEs and elicit the needed information. For purposes of capturing cross-MOS differences, how these requirements are specified may or may not substantively differ from technical performance requirements.

Consistent with the preceding discussion, the Panel recommended

Recommendation 5: Specifying non-technical performance requirements would be greatly facilitated by using pre-specified taxonomies to stimulate SMEs' formulation and assessment of these requirements. Candidate taxonomies, or the information needed to improve upon them (e.g., to make them more Army specific), exist and can be found from (a) research on critical incidents (i.e., where non-technical requirements are demonstrated or called for) and Army-wide performance dimensions, (b) the leadership and team literatures, and/or (c) existing job analysis taxonomies and instruments (e.g., PAQ). Similar work to develop or refine existing taxonomies on the predictor side (e.g., interests, values, and temperament) would also be useful in this regard.[6]

<u>*Recommendation 6:*</u> Except for differences in the content used to prompt SMEs, the specification of non-technical performance requirements should follow the same approach as the specification of technical requirements, unless pilot work suggests otherwise. If this is the case, carefully considered modifications to traditional approaches or the use of alternative analysis approaches (e.g., role-based job analysis, team task analysis) could prove useful.[7] Because of this, a flexible approach should be taken in specifying non-technical performance requirements such that these requirements can be specified differently, as needed.

Generalizing (or Transporting) Validity

<u>*Issue:*</u>

Are strategies for generalizing (or transporting) criterion-related validity estimates that do not require collecting empirical data from Army MOS viable and sufficient for the Army's needs?

Regardless of the criterion measure(s) used, conducting criterion-related validation studies for the full population, or a majority, of MOS is not feasible. Thus, strategies are needed for generalizing (or transporting) validity information collected for a select sample of MOS to other similar MOS (e.g., MOS sharing similar performance requirements) for purposes of obtaining estimates of classification efficiency for the entire system. Although there are strategies that could, at least in the immediate term, minimize the requirement to collect empirical predictor-criterion data from Army MOS, none sufficiently meets the Army's needs.

One such strategy is to make use of existing empirical criterion-related validity estimates, and established prediction equations for generalizing those estimates to other jobs, found in civilian job analysis databases (e.g., PAQ, CMQ). As discussed in the preceding section, such a strategy carries significant limitations. First, these databases are currently exclusively populated with civilian jobs whose criterion space (on which the criterion-related validity estimates are based) may not sufficiently represent military- or Army-specific performance requirements or other criterion dimensions (e.g., retention). Therefore, making use of these validity estimates could result in underestimates of the classification potential of select predictors in an Army context. Second, most of the criterion-related validity estimates contained in these databases, and the prediction equations for generalizing these estimates to other jobs, are limited to general and specific cognitive abilities. Missing in these databases is validity information for the kinds of non-cognitive predictors of interest to Army (i.e., interests, values, and temperament). Consequently, the prediction equations developed for these databases to generalize criterion-related validity information to, or across, Army MOS will be incomplete. A third limitation to making use of the validity information (and equations) contained in these databases, specifically the PAQ or CMQ, is

[6] Even if the development, or refinement, of existing taxonomies does not necessitate an alternative job analysis approach, they could improve the definitions of non-cognitive predictors, or their measurement, in ways that measurably enhance their classification potential.

[7] Preferably, the pilot work to develop and test the proposed job analysis system would be designed to illuminate any potential differences and provide suggestions for how best to proceed.

that it requires the Army to "buy in" to a job component validation (JCV) approach to generalizing validity. Although such an approach is not without its technical merits (cf. Hoffman & McPhail, 1998; McCormick, DeNisi, & Shaw, 1979; McCormick et al., 1972), it is based on a hypothesis – the "gravitational hypothesis" – that might not hold in the Army context. The "gravitational hypothesis" posits that individuals naturally gravitate to jobs commensurate with their abilities, interests, and so forth (McCormick et al., 1979; Wilk, Desmarias, & Sackett, 1995). The Army's current approach to classification greatly emphasizes the Army's needs when assigning recruits to MOS, thus constraining the potential for the "gravitational hypothesis" to operate.

An alternative strategy would be to employ a content-related validation strategy that is comparable to the Army's unit-weighted composites, but is more sensitive to cross-MOS differences (i.e., makes use of integer weights).[8] Such an approach, in combination with job analysis data, could technically produce prediction equations for use in assigning recruits to MOS and estimating classification efficiency. Although such an approach might be preferable to the first, it still carries serious limitations. First, and on a practical note, content validation approaches can be significantly labor intensive and expensive. Thus, such approaches are best for small-scale validation efforts, where there is minimal possibility of ever having empirical criterion-related validity estimates; this is not the case for the Army. Second, relying exclusively on content validity evidence is problematic, because differential validity will be (at least somewhat) independent of predictor content and instead a function of the types of criteria of interest. For example, past research has empirically demonstrated that different validity estimates can be expected for the same predictor battery when used to predict scores on three different criterion measures (e.g., job knowledge test [JKT], hands-on performance test [HOPT], job performance ratings), even when the content of all three criterion measures is identical (e.g., a JKT of 30 key MOS tasks, an HOPT asking a Soldier to demonstrate the capacity to carry out those 30 tasks, and a supervisor's ratings of a Soldier's typical performance on those 30 tasks over the past 6 months) (cf. McCloy, J. P. Campbell, & Cudeck, 1994). Thus, the applicability of prediction equations derived using a content validity approach to any particular predictor-criterion (measure) combination would be under serious question.

Accordingly, the Panel concluded the following:

Conclusion: Consistent with earlier recommendations, linking to or otherwise using the criterion-related validity information, and prediction equations for generalizing this information across jobs, found in existing civilian job analysis databases (e.g., PAQ, CMQ) is not advisable.

Conclusion: Although technically feasible, using a content-related validation strategy is likewise not advisable and would be impractical.

On the basis of these conclusions, the Panel recommended the following:

Recommendation 7: An approach to generalizing (or transporting) validity that is empirically based (in some form), and linked to the recommended job analysis database, should be employed.

[8] An example of such an approach can be found in Arthur, Doverspike, and Barrett (1996).

<u>*Issue:*</u>

What specific approach(es) to generalizing (or transporting) validity information would meet the Army's needs?

Consistent with the preceding discussion, an empirically based approach systematically linked to the recommended job analysis data would best meet the Army's needs. Multiple approaches are available (cf. C. C. Hoffman, Holden, & Gale, 2000; C. C. Hoffman & McPhail, 1998; Hollenbeck & Whitener, 1988; Guion, 1965; Johnson, in press; McCloy, 1994; McCormick, 1959; Peterson, Wise, Arabian, & R. G. Hoffman, 2001; for a recent review, see Scherbaum, 2005). Several specific approaches were identified. They were:

- *A job analysis-based validity (or test) transportability approach (cf. Guion, 1965).* Such an approach would involve using, and be based on, a systems-wide job analysis, such as that being proposed. Specifically, it would consist of the following:

    1. Using job analysis data – either what is available or collected specifically to facilitate the implementation of the proposed approach – assess the degree to which MOS share the same general performance requirements.
    2. Once MOS have been equated on that basis, identify focal MOS (or multiple MOS within a common cluster). These focal MOS (about 20-30) would constitute the sample of MOS for which criterion-related validation studies will be conducted.
    3. Conduct empirical criterion-related validation studies using incumbents from the select sample of focal MOS.
    4. After the studies have been conducted and the criterion-related and differential validity established for some set of predictors (or tests) – in accordance with professional standards – validity estimates could then be transported to the other MOS within the cluster, or however equivalence has been operationalized.

- *A full hierarchical linear modeling (HLM) approach (cf. McCloy, 1994).* In brief, this approach would involve collecting sufficient criterion data from a sample of Army jobs and job analysis data for all jobs to build a single multi-level equation (i.e., persons nested within jobs) that will generate job-specific prediction equations for use in obtaining Soldiers' predicted criterion scores, even for MOS missing criterion data. Implementing this approach would involve:

    1. Obtaining criterion data on a reasonable number (at least 20, although 30 or more would be preferred) of Army jobs – ideally ones that span the identified clusters.
    2. Obtaining job analysis data on the full population of Army jobs to permit identification of variables defining various job characteristics (e.g., cognitive complexity, working conditions, finger/manual dexterity).
    3. Building an HLM that regresses criterion data on individual characteristics (e.g., ASVAB scores, education tier) at Level 1 and the regression parameters for the individual characteristics on job characteristic variables at Level 2.

4. For MOS not included in the estimation, placing the values for the job characteristic variables into the Level-2 equations from the HLM and obtain estimated job-specific regression coefficients that could be used to obtain predicted criterion scores for Soldiers across the MOS missing criterion data.

- *A combined validity transportability and empirically based synthetic validation or HLM approach.* Comparable to the first approach, this approach would consist of the following:

  1. Presume that the clustering exercise produces 20-25 clusters of MOS. Using the job analysis information on which they were based, identify the most representative MOS in each cluster on the basis of their performance requirements.
  2. Conduct criterion-related validation studies for the selected sample of focal MOS.
  3. With data in hand, estimate a prediction equation for each focal MOS. These equations would then be transported to the other MOS in the applicable cluster. Differential assignments would be to the cluster and then made more specific by Army priorities, training seat availability, and applicant preferences.
  4. For target clusters, a (empirical) synthetic validation, or HLM procedure (like that described in the preceding approach), could be used to further differentiate among MOS in a cluster. The empirical criterion-related validity estimates for synthetic validation purposes could be obtained from the validation studies of focal MOS in each cluster, provided a fair number of studies using appropriate criteria could be completed.[9] If there were at least 20-25 such studies, then HLM techniques could be applied as well.

- *An incremental, rational synthetic validation-validity transportability-HLM approach.* In general, this approach would consist of (a) starting with synthetically derived prediction equations based on rational (expert) judgments as illustrated in the Army SYNVAL project (cf. Peterson, Owens-Kurtz, R. G. Hoffman, Arabian, & Whetzel, 1990; Peterson et al., 2001); and then (b) modifying the synthetically derived equations as empirical criterion-related validation studies are completed and the HLM approach can be applied.[10] There are different ways in which this approach could be implemented. One strategy would be as follows:

  1. Create the recommended job analysis system and collect the job analysis data.
  2. When job analysis data are available, create the clusters of MOS.
  3. Collect linkage judgments between the job descriptors and MOS performance dimensions and create synthetic equations for (a) each MOS, (b) each cluster, and (c) one overall equation (as in the Army SYNVAL project).

---

[9] Because of this requirement, the job analysis should be designed with the empirical synthetic validation in mind.

[10] Where feasible, existing empirical estimates (e.g., from Project A or Select21) could be used to supplement the judgmental estimates of criterion-related validity and enhance the initial synthetic equations.

4. Use whatever version of the synthetic equations seems to work best. There is no real way to test this, except to compare the equations in some way (e.g., to the equations estimated by Zeidner and Johnson, cf. Zeidner, Johnson, Vladimirsky, & Weldon, 2000b) for interim use in classification.
5. As the 20 – 30 validity studies are completed, to be conservative, use a validity (or test) transportability approach (i.e., apply the focal MOS equation to the other MOS in the cluster) to replace the synthetic equations.
6. When all the 20-30 studies are completed, implement the combined validity transportability-HLM approach.

- *Standard job component validation (JCV) approach (cf. McCormick, 1959; McCormick et al., 1979).* Comparable to what has been done using the PAQ, this approach would involve using job analysis data to derive components and scores on those components for MOS. This would involve

  1. Using mean predictor (test) scores and validity coefficients – based on past research or collected from new criterion-related validation studies conducted for a selected sample of MOS – and constructing equations (from regressing mean test scores and validity coefficients onto the component scores) for the targeted MOS.
  2. Once established, these component equations could then be used to construct classification/assignment equations for other MOS, or to update existing equations for MOS as the nature of the jobs change.[11]

All five approaches are technically feasible and each has some research base supporting it, albeit to varying degrees. To fully implement any of the five approaches as outlined would require (a) MOS-specific job analysis data (preferably for the full population of MOS), (b) 20-30 criterion-related validity studies, and (c) sample sizes per MOS sufficient for estimating relatively stable regression coefficients (for predicting criterion scores) or validity coefficients for job components within an MOS.[12] Where the five approaches primarily differ is in their sensitivity in capturing cross-MOS differences in validity for the predictor-criterion combination(s) of interest (i.e., some approaches permit the estimation of MOS-specific equations for any number of MOS, whereas others produce equations applicable to a cluster of MOS).

---

[11] For this approach to be feasible, it is critical that the criteria be measured in ways that are congruent with the job analytically derived components (i.e., the criterion dimensions are conceptually and operationally consistent with the components).
[12] The fourth approach could be implemented on the basis of the rationally derived synthetic equations without conducting all of the recommended criterion-related validity studies. As described, the synthetic equations could then be refined, or replaced, as the recommended validity studies are completed.

In sum, the Panel recommended

Recommendation 8: Several specific approaches for generalizing (or transporting) validity information were identified that could meet the Army's needs. They are:

- A full validity (or test) transportability approach.
- A full HLM approach.
- A combined validity (or test) transportability-HLM approach.
- An incremental, rational synthetic validity-validity transportability-HLM approach.
- Standard job component validation (JCV) approach.

The Army need not a make final decision on which approach to pursue at this point in time. Because the first four approaches operate on and make use of the same data (i.e., from 20-30 criterion-related validation studies), they could be pursued and tested simultaneously, provided sufficient resources are available.

Job Clustering

Issue:
What MOS clustering solution(s) would meet the Army's classification research needs? Specifically, on what basis (e.g., similarity in performance requirements, select KSAOs) should MOS be clustered?

The overall purpose for clustering MOS is, broadly speaking, to optimize the classification of Army recruits to MOS. Within this overall purpose, however, there are multiple secondary purposes that can be served from clustering MOS – from facilitating the collection of criterion data to investigating possible enhancements to the operational classification system (e.g., from using a two-stage procedure, whereby recruits are first assigned to broad clusters of MOS on the basis of their interests, then to a specific MOS within a cluster based on their abilities). Table 3 provides a listing of potential purposes, and the descriptor(s) most relevant to each. As evident from Table 3, different purposes imply different descriptors, even if only in how said descriptors are weighted. Because of this, and as demonstrated by past research (cf. Cornelius, Carron, & Collins, 1979; Reynolds, Laabs, & Harris, 1996), very different clustering solutions can result depending on which particular descriptors are used to cluster MOS (Sackett, 1991). For example, clustering MOS exclusively on the basis of similarity in abilities requirements will yield a solution different from clustering on the basis of performance requirements. Due to resource constraints, it is not feasible to pursue all these possibilities in the immediate term. Further, doing so in several cases would be premature (e.g., investigating the classification potential of a two-stage procedure). Thus, prioritization is needed.

As indicated previously, the Army's criterion problem rests in significant part on the large number and diversity of entry-level MOS, and the difficulty and expense of collecting criterion data for a sufficiently representative sample of these MOS. Generally speaking, meeting these needs requires clustering solutions that (a) support the sampling of MOS for the purposes

Table 3. Purposes for Clustering MOS

Purpose	Primary Clustering Descriptor(s)	Comment(s)
1. To create occupational clusters to facilitate the sampling of MOS and the conduct of criterion-related validation studies.	General and specific performance requirements; any/all occupation-specific KSAs.	Could facilitate the conduct of criterion-related validation studies by grouping occupations that are equivalent to enable the Army to obtain a sufficiently representative sample of MOS that would permit generalization to other MOS but minimize difficulties in obtaining the requisite data.
2. To use clustering to explore/detect candidates for "mid-range" criterion measures.	General performance requirements.	For potential use in studying feasibility of developing end-of-training criteria applicable across multiple MOS.
3. To research/explore the utility of non-task descriptor-based MOS sub-clusters, either in and of themselves, or within broader task descriptor-based clusters, especially for prediction of non-performance criteria (i.e., attrition or retention-related). [a]	Any/all KSAOs; also, potentially, organization/job context descriptors (for aspects of research related to person-organization fit issues).	Clustering descriptor(s) would be selected based on specific research question(s) of interest. Could address such questions as, Are there clusters of MOS for which certain types or combinations of KSAOs are predictive of attrition, or of non-task performance (e.g., peer leadership), or of person-organization fit? [b]
4. To research/explore the utility of task descriptor-based MOS sub-clusters within broader non-task descriptor-based clusters.	Specific and/or general performance requirements.	Could address such questions as, What MOS within Aptitude Areas (AA) have similar performance requirements (useful for supporting a validity or test transportability approach)? [c]

Notes. (a) For all purposes listed that involve "researching/exploring the utility of…" it should be understood the central question guiding all such research is the "utility for what?" (i.e., what is the criterion on which to evaluate this research?). In some cases, this criterion could be the actual criterion data collected from the validation studies (i.e., which clustering solution maximizes classification efficiency), provided there are criterion data for a sufficient number of MOS. Alternatively, the criterion could be more qualitative in nature, along the lines of, "Does the clustering solution answer – or shed more light on – the question asked?" These later cases still require advance specification of the parameters for acceptable answers to such questions.

(b) Should aggregate classification gains accrue from the use of alternative criteria (e.g., retention-related) but their maximization result in different MOS assignments than maximizing aggregate technical performance, then additional research would be needed to scale the relative value of gains on each criterion.

(c) Such an approach would also be useful for identifying clustering solutions that maximize technical performance.

Table 3. Purposes for Clustering MOS (cont'd)

Purpose	Primary Clustering Descriptor(s)	Comment(s)
5. To research various forms of a two-stage classification model: Similar to #1 (above), but sub-cluster within existing AA clusters. [d]	Interests, values, temperaments.	A two-stage procedure might assign Soldiers to broad clusters first and then make further differential assignments within clusters. Assignments to broad clusters (i.e., the first stage) could be on the basis of interests and/or fit with the peer leadership and team support performance requirements. Specific assignments (i.e., the second stage) within the broad clusters could be made on the basis of AFQT, AA, Army priorities, or openings. Alternatively, if second stage were counseling/guidance-oriented, could mitigate response bias problems with some of these measures. [e]
6. To research various forms of a "two-track" (as opposed to two-stage) selection/classification model via clustering.	Would depend on premises of the second "track," but most likely applications would probably involve either cross-functional skills (e.g., oral communication, teamwork, critical thinking) or specialized/advanced skills or knowledge, as might be used for an alternative track for experienced applicants, or applicants with relevant certifications.	Could address such questions as, Are there clusters of MOS for which certain specialized predictor types (e.g., cross-functional skill measures, or prior work experience) might be leveraged?

Notes. (d) The Army's current classification process consists of a single stage. Specifically, new Army recruits are shown one or more possible MOS opportunities on a career counselors' display screen, then make a choice. Which MOS opportunities are presented is based on a joint function of recruits' AFQT and AA scores, Army priorities, and technical training seat availability.

(e) Because of the way the Army's basic and technical training are currently structured, it would be difficult to do multi-stage classification at different points in time (like the Air Force) by delaying the final job assignment until after basic training.

Table 3. Purposes for Clustering MOS (cont'd)

Purpose	Primary Clustering Descriptor(s)	Comment(s)
7. To create "new and improved" occupational clusters and occupational mapping information (i.e., horizontal and vertical career paths) as career guidance tools for both recruiters and applicants.	Specific and/or general performance requirements, possibly augmented or supplemented by aptitude, temperament, and/or interest-oriented descriptors.	Could enhance overall classification efficiency via improved matching of applicants to MOS (e.g., in terms of non-aptitude predictors) afforded by more refined and well-grounded occupational structure and mapping of occupational interrelationships.
8. To evaluate the efficacy of different rational clustering procedures (i.e., those based on different methods, procedures, clustering descriptors, SMEs, SME training) using empirical clustering solutions as criteria.	Not applicable-multiple (effects of different clustering descriptors is one of the features being evaluated).	Could address the question of the relative accuracy (and hence value for future potential clustering applications where empirical data are unavailable) of different types of rational clustering methods.
9. To generate hypotheses suggested by the mapping of occupational clusters in terms of different descriptors, and subsequent comparisons among these different clustering solutions.	Not applicable-multiple (effects of different clustering descriptors is one of the features being evaluated).	Useful to see what insights might emerge from a thorough examination and comparison among clustering solutions based on different descriptors, in terms of what variables seem to matter, how the different structures do/do not overlap, and so forth.

Note. Purposes 8 and 9 are exploratory, and therefore non-descriptor-specific.

of collecting, then generalizing (or transporting) criterion-related validity information to other MOS; and (b) facilitate the development (or refinement) of criterion measures, in particular "mid-range" criteria (if feasible), capable of sufficiently capturing cross-MOS differences on classification-relevant criterion dimensions (e.g., technical and/or non-technical performance).[13] Further, because investigating the classification potential of new predictors (e.g., interest, values, and temperament), either separately or relative to the existing ASVAB, carries special importance in the Army's classification research agenda, there is the need for solutions that produce clusters that maximally differentiate MOS on the basis of select predictors (i.e., clusters of MOS exhibiting similar patterns of differential prediction with targeted criteria). Absent such clustering solutions the possible classification gains resulting from the use of new, alternative predictors cannot be evaluated independent of the constraints that might be imposed from using existing MOS clusters (e.g., the existing Army Aptitude Areas). For example, to the extent that clustering on the basis of interests produces a clustering solution different from one based on cognitive aptitudes, and thereby results in different MOS assignments, any effort to estimate the classification potential of a predictor measuring recruits' interests using existing MOS clusters will be an underestimate. To address this requires the derivation of cluster solutions that maximally differentiate MOS on the same basis as what the targeted predictors are measuring.

In sum, the Panel concluded

Conclusion: Different MOS clustering solutions will be more or less advantageous for different purposes. The Army's immediate needs require clustering solutions that (a) support the sampling of MOS for criterion-related validation work, (b) facilitate criterion development, and (c) produce clusters that maximize classification efficiency for select predictors.

Consistent with this, the Panel recommended

Recommendation 9: Priority should be placed on solutions that systematically cluster MOS, either separately or jointly (e.g., a multi-tier solution), on the basis of performance requirements and select KSAO descriptors. Because "KSAOs" can cover a wide range of descriptors, great care needs to be paid to the specification and selection of KSAO descriptors for use in clustering. To start, KSAOs should at least be partitioned into three predictor domains: (a) occupation-specific knowledges and skills (KSs), (b) specific abilities (As), and (c) interests/values/temperaments (Os).[14] Solutions that consider both performance requirements and select KSAO descriptors simultaneously could prove to be particularly advantageous. Multiple solutions should be generated and compared, as data become available, so that the evaluation of which solution works best for meeting the aforementioned purposes can be examined empirically.

---

[13] "Mid-range" criteria refer to criterion measures, or individual components of these measures, that are of sufficient generality that they can reasonably differentiate MOS (or cluster of MOS), but are not so specific that they would only be applicable to a single MOS.

[14] Any further partitioning within each of these domains, or weighting of different descriptors within each to form a composite, could then be derived (a) empirically on the basis of principal component analyses of descriptor scores within each domain or (b) rationally on the basis of which descriptors carry the greatest potential for differentiating MOS for purposes of maximizing classification efficiency.

Issue:
Are existing MOS cluster solutions sufficient to meet the Army's classification research needs?

Several MOS cluster solutions, operational or research-based, currently exist (e.g., Human Resources Research Organization, 2005; Johnson, Zeidner, & Leaman, 1992). In some cases, these cluster solutions were derived rationally, in some cases empirically, and in others a combination of the two. The most prominent of these solutions, and the one the Army currently uses to classify recruits into entry-level MOS, is the existing Aptitude Areas (AA), of which there are nine.[15] The existing nine AA date back to 1972 and were derived from an empirical clustering of Career Management Fields through an iterative process of selecting those candidate composite tests which best explained training performance and combining those CMFs which were "explained" by the same tests (Maier and Fuchs, 1972). Since that time, any (re)grouping of MOS under the nine AA (e.g., resulting from changes or updates in MOS) has been handled rationally.[16] Other, more expanded cluster solutions have been generated for use in Army research (cf. Human Resources Research Organization, 2005; Johnson et al., 1992) but are currently not in operational use.

Although existing cluster solutions are informative and could serve as comparison points for evaluating any alternative solutions, none appears sufficient for the Army's needs as outlined above. The reasons for this are threefold. First, the number of clusters constituting these solutions may be smaller than would be optimal for classification purposes (cf. Scholarios, Johnson, & Zeidner, 1994; Zeidner et al., 1997; Zeidner, Johnson, Vladimirsky, & Weldon, 2000a). Thus, using these cluster solutions "as is" could result in underestimates of a predictor battery's classification potential. Second, the clusters constituting many of these existing solutions, specifically the current AAs, are based almost exclusively on a single type of descriptor (e.g., profiles of specific abilities). Consequently, they might be useful for one of the purposes discussed above, but not all.[17] More specifically, and similar to earlier discussions, because they do not consider descriptors relevant to new predictors of special interest to the Army (i.e., interests, values, and temperament), they are likely to produce biased estimates of said predictors' classification potential. Third, and not unrelated to the preceding point, few of these solutions are systematically based on or informed by job analysis data. Specific to the Army's needs, the absence of job analysis data makes it difficult to determine limits on the generalizability (or transportability) of criterion-related validity information to other MOS. This limitation is particularly problematic for the purposes of estimating the classification potential for new predictors, where the collection of new predictor-criterion data is required. Further, this situation greatly constrains the use of these clusters and limits the flexibility with which they could be refined, and revised, over time as MOS change. For example, having job analysis data would enable the Army potentially to reuse previously collected validity information for use in evaluating clustering solutions meant to serve other purposes but rely on the same job information (see Table 3).

---

[15] The nine AAs are: Combat, Field Artillery, Clerical, Electronics Repair, Mechanical Maintenance, General Maintenance, Operators and Food, Surveillance and Communication, and Skilled Technical.

[16] Recent efforts to cluster MOS empirically on the basis of similarities in their ASVAB-based prediction equations (used to estimate recruits' predicted technical performance in an MOS) generally find support for the rational clustering of MOS constituting the nine AAs (cf. Johnson et al., 1992).

[17] Besides satisfying multiple purposes, cluster solutions based on multiple descriptors tend to yield more statistically stable and viable clusters.

Accordingly, the Panel concluded

Conclusion: Although they might serve as useful comparison points, existing MOS cluster solutions are not sufficient to meet the Army's needs, as outlined above.

As an alternative, the Panel recommended

Recommendation 10: To meet the Army's needs, clustering solutions having an empirical basis (even if supplemented by expert judgments), systematically derived and supported by job analysis data, should be employed. Because the collection of the data needed to validate these solutions (i.e., predictor score profiles, criterion-related validity estimates) will take time to accumulate, the following interim approach is recommended as a starting point:

- Generate an initial cluster solution using general performance requirements-based descriptor scores, collected from the job analysis.
- Use ARI and other psychologists to rate MOS on select KSAOs (e.g., specific abilities, interests, values, temperament) to provide an initial database of scores on these descriptors.
- Provided no validity estimates are available, examine predictor score profiles on select KSAOs for each cluster to obtain information on the (a) differences between MOS within and across clusters, and (b) integrity of the clusters and the predictor-based profiles.[18] These predictor score profiles would be formed from predictor data collected from Soldiers representing a reasonable number of MOS.

Criterion Measurement

Issue:
What general approach(es) to criterion measurement would prove most viable and best meet the Army's classification research needs?

Measuring the criterion space for purposes of estimating classification efficiency is a complex matter. The criterion space is multidimensional and multi-faceted, and different criterion dimensions reflect alternative and often competing personnel management goals (Rosse et al., 2001). Because of this, estimates of classification efficiency (e.g., mean predicted performance [MPP]) will vary greatly depending on the criterion measure(s) used in validation studies, even when different measures aim to assess the same criterion dimension (cf. McCloy et al., 1994). Thus, the choice of criterion measures used, their quality, their coverage of the criterion space, and so on carries significant implications. Although there could be practical advantages to the use of simple, less expensive alternatives to traditional criterion measurement methods, these advantages would likely be offset by the same alternative measures' deficiencies

[18] If all clusters and predictor profiles are different from each other and sufficiently homogeneous, then such information would indicate the potential for classification gains. This information could also be useful in prioritizing for which MOS to collect criterion data in validation studies. For example, it might be best to collect criterion data from MOS (a) at the center of a cluster, (b) at mid-distance from the center, and (c) far from the center (i.e., based on clustering results using the performance requirements descriptors).

in capturing substantive MOS-specific and cross-MOS differences (in the criterion space) relevant to classification. Moreover, even these simple, less expensive alternatives will require a non-trivial level of resources, because conducting the needed validation studies for a sufficient sample of MOS will be a considerable undertaking. In sum, criterion measurement will require resources for it to be sufficient, irrespective of the criterion measure(s) employed.

Because of this, the Army needs to be in a position to maximize the investments it makes in criterion measurement. Generally speaking, the key to maximizing this investment rests on a solid, Army-specific job analysis system that (a) supports strategies for collecting criterion data providing adequate coverage of the criterion space for a sufficient sample of MOS, while minimizing development and administrative costs; and (b) facilitates the discovery and specification of critical MOS-specific and cross-MOS criterion dimensions that sufficiently differentiate across MOS (i.e., for purposes of capturing differential validity and estimating classification efficiency for select predictors). This latter information can be used either to (a) refine existing criterion measures or (b) develop new measures (e.g., end-of-training criteria, job performance ratings) that could prove effective, and comparatively more economical, in capturing cross-MOS differences than knowledge tests.

In sum, the Panel concluded

Conclusion: The key to maximizing the Army's investment in criterion measurement rests on a solid, Army-specific job analysis system that (a) supports strategies for collecting adequate criterion data for a sufficient sample of MOS while minimizing costs, and (b) facilitates the discovery and specification of critical MOS-specific and cross-MOS criterion dimensions useful for developing, or refining, criterion measures that effectively capture cross-MOS differences. Simple, inexpensive alternatives to traditional criterion measurement methods (e.g., JKTs) are neither feasible nor advisable.

Consistent with this, the Panel recommended

Recommendation 11: Using Army-specific job analysis data, the Army should pursue (a) strategies for collecting adequate criterion data for a sufficient sample of MOS and (b) development of criterion measures, or refinement of existing ones, that sufficiently differentiate across MOS.

Issue:

What specific strategies for collecting criterion data, informed and supported by Army-specific job analysis data, would provide adequate coverage of the criterion space for a sufficient sample of MOS, while minimizing development and administrative costs?

Consistent with the Army's needs, even though multiple criteria are recommended, the collection of criterion data need *not* be an "all or nothing" proposition (i.e., data on a complete set of criteria – e.g., training, on the job, and so on – need not be collected on all Soldiers representing all focal MOS sampled from the same criterion-related validation study). There are alternative, cost-effective strategies for the collection of criterion data, informed and supported by Army-specific job analysis data and clustering of MOS on the basis of these data, that would provide adequate coverage of the criterion space for a sufficient sample of MOS while minimizing resources. One such strategy would consist of the following:

- *First, develop a sufficiently complete set of criterion measures for a sample of focal MOS, each representative of a job cluster (i.e., from clustering MOS on the basis of the recommended job analysis data), and/or samples of MOS from targeted clusters based on Army priorities*. Candidates for this complete set of criterion measures would include (a) a knowledge test (administered at end-of-training or post-training); (b) MOS-specific, behaviorally anchored ratings scales (completed by supervisors and/or peers); and (c) a retention-related criterion (e.g., satisfaction with MOS).[19] Depending on the nature and types of predictors of interest, one could add a "walk through" demonstration of proficiency (i.e., work sample assessment) on key tasks, scored by an administrator.[20] If not practically feasible, a computer-administered exam might suffice as a substitute for the walk through. A pilot study on a subset of focal MOS (4-6) could be useful for determining which criterion measures would constitute a complete set (i.e., provide sufficient coverage of the criterion space relevant to classification).

- *Second, conduct full-scale criterion-related validation studies on these MOS.* As discussed above, the obtained validity information could then be extended to other MOS using one of the recommended approaches for generalizing (or transporting) validity information (see pp. 13-15).

- *Third, for those MOS not receiving the full-scale validation treatment, more limited studies would be completed employing a reduced (or "bare bones") set of criteria, possibly even limited to end-of-training criteria.* The primary purpose of these studies would be to demonstrate empirically that the generalized (or transported) prediction equations for estimating classification efficiency from the preceding step are not adversely affecting the other MOS by their application.[21] When implementing the "bare bones" approach, using existing theory (or hypotheses) about predictor-criterion relations – either separately or in combination with job analysis data and other information – would be useful for selecting or matching criteria with MOS (e.g., selecting retention-related measures for use in an MOS known to have retention issues).

In sum, the Panel concluded

<u>*Conclusion:*</u> Although multiple criteria are recommended, collection of criterion data need not be an "all or nothing" proposition.

[19] In part, the comprehensiveness of this set of criterion measures will be a function of the number of job clusters.

[20] Including the walk through would be advantageous for capturing the potential of psychomotor abilities for classification.

[21] Note, these studies would not indicate exactly how efficiently the equation predicts performance in these MOS.

From this, the Panel recommended

Recommendation 12: The Army should consider administering a complete set of criterion measures (e.g., JKT, ratings, retention) to focal MOS (i.e., those MOS most representative of a cluster), while administering a reduced set of criteria to non-focal MOS. Decisions on which MOS are focal and which criterion measures to include would best be guided by Army-specific job analysis data, MOS clustering results, Army priorities, and existing theory on predictor-criterion relations.

Issue:
Would end-of-training criteria (knowledge tests and/or ratings) be useful and meet the Army's needs? In what specific ways could job analysis data be used to inform and advance end-of-training criteria?

End-of-training criteria, specifically knowledge tests and peer (and possibly instructor) ratings administered at the end of Advanced Individual Training (AIT), would be useful. The reasons for this are threefold. First, and most practically, access to Soldiers for research purposes may be greatest in a training environment. Because of this, the costs to develop and collect data on a knowledge test in particular are likely to be considerably less, on average, in the training (as opposed to the post-training) environment. Second, well designed and soundly administered end-of-training criteria can capture substantive (and meaningful) variance in the criterion space that is relevant to classification. Third, and not unrelated to the preceding point, although not intended to replace post-training criteria, end-of-training criteria – specifically knowledge tests – might serve as reasonable surrogates for (and yield comparable MOS assignments as) the same criterion measures administered post-training, particularly if combined with other post-training criterion measures that are easier to collect (e.g., MOS-specific technical ratings).[22] Consistent with this point, past research indicates that training and post-training criteria measuring the same, or similar, criterion dimensions are significantly and substantively related (cf. J. P. Campbell, 1987; J. P. Campbell & Knapp, 2001; J. P. Campbell & Zook, 1991).

At present, the major limitation with using end-of-training criteria is that collecting these data will require the development of new criterion measures. This is because AIT schools vary in their use of standardized end-of-training criteria. In addition, the schools differ in the specific training performance information they collect (and how they do so).[23] The Army need not start completely from scratch on this, as knowledge tests developed for post-training administration for several MOS exist that could be feasibly repurposed for end-of-training use. Having Army-specific job analysis data could greatly facilitate the development of additional end-of-training criteria, or the refinement of existing criterion measures. For example, although developing "cross-MOS" knowledge tests (i.e., tests applicable to multiple MOS with similar performance requirements) has generally proven infeasible, perhaps "mid-range" tests could emerge from analyzing and clustering the job data, as recommended.[24]

[22] It should be noted that this point is contingent on the use of the same predictor(s).

[23] However, it need not be the case that criteria must be developed for every single MOS.

[24] Consistent with the earlier discussion on "mid-range" criteria, "mid-range" knowledge tests are criterion tests, or individual test components, that are of sufficient generality that they can reasonably differentiate MOS (or cluster of MOS), but are not so specific that they would only be applicable to a single MOS.

In sum, the Panel concluded

Conclusion: For both practical and substantive reasons, well developed and soundly administered end-of-training criteria would be useful for meeting the Army's needs.

Accordingly, the Panel made the following recommendations:

Recommendation 13: The Army should pursue the use of end-of-training criteria, particularly knowledge tests and peer (and possibly instructor) ratings. Further, the Army should continue to assess the relations between end-of-training criteria and post-training criteria measuring the same, or similar, criterion dimensions.

Recommendation 14: Using Army-specific job analysis data and the results of the MOS clustering as recommended earlier, the Army should explore the feasibility of mid-range criterion tests (or test components), specifically for end-of-training tests.

Recommendation 15: Should the preceding recommendation prove infeasible, the Army-specific job analysis data could be used to maximize the resources used for developing end-of-training knowledge tests. For example, following a "top down" approach to criterion development, the performance requirements taxonomy developed as part of the proposed job analysis system could serve as a general test plan template and the MOS-specific data as a kind of weighting scheme for the general plan. Doing so would enable more incremental development approaches where similarities in test content can be seen, and capitalized on, ahead of time. Alternatively, the job analysis data could be used to weight existing end-of-training criterion tests to enhance their validity.

Issue:

Would job performance ratings, including MOS-specific technical ratings, be useful and meet the Army's needs? In what specific ways can job analysis data be used to inform and advance performance ratings?

Behaviorally anchored ratings of Soldier job performance made by supervisors and/or peers, including ratings of Soldiers' performance on MOS-specific technical performance requirements, would be useful. The reasons for this are twofold. First, and comparatively speaking, the resources needed to develop and administer behaviorally anchored ratings, on average, are considerably less than for knowledge tests assessing the same (or similar) criterion dimensions. Second, well designed and soundly administered behaviorally anchored ratings can capture substantive and meaningful sources of variance in the criterion space relevant to classification. In particular, ratings can capture facets of the criterion space potentially useful for classification purposes, specifically non-technical performance (or "will do") dimensions, that are not (and cannot be) assessed by knowledge tests but which might be especially useful for determining and establishing the classification potential of non-cognitive predictors.

Two potential reservations with the use of ratings concern their (a) potential susceptibility to halo and other rater biases, and (b) ability to capture critical MOS-specific and cross-MOS performance dimensions that sufficiently differentiate across MOS (i.e., for purposes of determining and estimating the classification potential of select predictors). Although halo (and other rater biases) can be problematic if left unchecked, there are strategies that have proven effective in minimizing these biases and maximizing the construct validity of ratings when administered for research purposes. Specifically, these include (a) specifying the performance dimensions to be rated as clearly and distinctly as possible (i.e., so that the scales can be explicitly distinguished from each other), (b) providing raters with the best available training, (c) standardizing the rating process to promote consistent implementation across raters and ratees, and (d) ensuring that those providing ratings (supervisors and/or peers) have had sufficient opportunity to observe a Soldier. Of particular importance to minimizing rater biases, raters must be encouraged to (a) accept the research goals of representing the ratees' standing on the dimensions as accurately as possible, as these dimensions are explicitly defined by the rating instrument (and not the raters' own implicit understanding of the dimensions); and (b) take the time to consider each scale carefully and thoroughly. Well designed and delivered rater training, and the use of select data collection method, can be effective in ensuring that raters have sufficient motivation (and time) to make accurate ratings.

The most significant factor affecting the ability of ratings to sufficiently capture MOS-specific and cross-MOS performance dimensions relevant to classification concerns the selection and specification of the performance dimensions (technical or non-technical) to be assessed. Having Army-specific job analysis data, as recommended earlier, would be useful in ensuring that the dimensions selected, and how they are specified, sufficiently differentiate MOS for this purpose. Similarly, these data could facilitate the development of experimental, alternative rating formats (and other assessment methods) that provide more realistic and meaningful operationalizations of non-technical performance dimensions in ways that partial out technical performance requirements (e.g., least preferred co-worker scale).

In sum, the Panel concluded

Conclusion: For both practical and substantive reasons, well developed and soundly administered behaviorally anchored job performance ratings would be useful for meeting the Army's needs.

Consistent with this, the Panel recommended

Recommendation 16: The Army should pursue the use of behaviorally anchored job performance ratings. To minimize halo (and other biases) and maximize the construct validity of these ratings, the Army should (a) specify the performance dimensions to be rated as clearly and distinctly as possible (i.e., so that the scales can be explicitly distinguished from each other), (b) provide raters with the best available training, (c) standardize the rating process to promote consistent implementation across raters and ratees, and (d) ensure that those providing ratings (supervisors and/or peers) have had sufficient opportunity to observe a Soldier.

Recommendation 17: Having Army-specific job analysis data is essential, as they would greatly facilitate (a) the discovery, selection, and specification of MOS-specific and cross-MOS performance dimensions, technical and non-technical, to be assessed by ratings; and (b) the development of experimental, alternative rating formats (and other assessment methods) that provide more realistic and meaningful operationalizations of non-technical performance dimensions in ways that partial out technical performance requirements (e.g., least preferred co-worker scale).

Issue:
Does the validation of non-cognitive predictors (e.g., interests, values, temperament) raise special considerations and implications for criterion measurement?

Validating and evaluating the classification potential of non-cognitive predictors carries implications for criterion measurement. With the exception of select interests, non-cognitive predictors have generally not emerged as significant contributors to classification efficiency, particularly over and above specific aptitudes (cf. Rosse et al., 2001; Scholarios et al., 1994). In part, this finding is potentially attributable to the nature and type of criterion measures used in these studies, which has almost exclusively been a JKT or a composite representing MOS-specific technical performance (e.g., based on JKT scores and MOS-specific technical performance ratings). Past research conducted within jobs, however, indicates that non-cognitive predictors are most strongly predictive of (a) non-technical, "will do" performance dimensions (e.g., demonstrating effort, citizenship, peer leadership, teamwork); (b) non-performance criteria, specifically occupational and organizational retention criteria; and (c) indices of career success (cf. Barrick & Mount, 1991; Hogan & Holland, 2003; Hough & Furnham, 2003; Hurtz & Donovan, 2000; Judge, Higgins, Thoresen, & Barrick, 1999; Ozer, & Benet-Martinez, 2006). As indicated previously, estimates of classification potential for the same predictors are contingent on, and will vary as a function of, the criterion measure used and the classification aims to optimize. Therefore, the inclusion of measures capturing these specific facets of the criterion space could be critical when determining and evaluating the classification potential of non-cognitive predictors.

In sum, the Panel concluded

Conclusion: Validating and establishing the classification potential of non-cognitive predictors carries implications for criterion measurement. Specifically, there is a need to include criteria measuring (a) MOS-specific and cross-MOS dimensions of non-technical, "will do" performance; and (b) occupational and organizational retention (e.g., MOS satisfaction).

On the basis of this conclusion, the Panel recommended

Recommendation 18: When validating and establishing the classification potential of non-cognitive predictors, the Army should employ (a) ratings of MOS-specific and cross-MOS non-technical performance dimensions, and (b) occupational and organization retention-related criteria.

Recommendation 19: Although objective retention and attrition criteria have been and can be highly inaccurate (i.e., if relied on exclusively without consideration of other measures), research could be conducted to render them useable for validation purposes. Doing so, however, would require a significant initial effort either to shape up the official coding for Soldiers' reasons for staying-leaving, or to devise a method to recode those reasons reliably and accurately. Alternatively, the Army could pursue new, alternative possibilities for collecting reasons (e.g., exit surveys) that could then be instituted and stored for future validation work.

Estimating Classification Efficiency

<u>*Issue:*</u>
What are the viable options for estimating the classification potential of new predictor batteries (e.g., ones consisting of new ASVAB subtests or measures of non-cognitive predictors)? What impact will practical constraints associated with the typical criterion-related validation study, specifically the number of MOS and the sample size per MOS, have on these estimates? How can (and should) these constraints best be dealt with?

To empirically estimate and evaluate the potential classification gains for the entire system accruing from the use of new, alternative predictor batteries (e.g., consisting of new ASVAB subtests or measures of non-cognitive predictors) will require criterion-related validity estimates for a sufficiently representative clustering of MOS, specifically estimates from at least one focal MOS in each cluster.[25] Provided the cluster solutions were based on appropriate job analysis data, and if a very representative focal MOS could be identified in each cluster, and if the same prediction battery could be validated on a sample from each MOS, and if the sample size per MOS was about 300-500, then one could obtain reliable estimates of classification gains over a wide range of simulated, real world conditions using the Enlisted Personnel Allocation System (EPAS). Alternatively, one could use a simulation-based approach (e.g., Zeidner et al., 1997; Zeidner et al., 2000a) to estimate and compare the maximum potential gains that can be achieved if the battery is used in an optimal fashion or under a limited set of operational constraints (e.g., job quotas), relative to a specified alternative (e.g., the existing AA composites). Collecting criterion data to derive these validity estimates, however, for even a handful of MOS has proven challenging. Even with implementing one or more of the recommendations offered above, securing predictor-criterion data from a large number of Soldiers for each focal MOS will be difficult, at least in the context of a single study. This limitation understandably raises questions regarding (a) the implications of these practical constraints, in particular the sampling of MOS and sample size per MOS, on approaches for estimating a predictor battery's classification efficiency; and (b) how best to deal with these implications.

The number and nature of MOS sampled in a study carries significant implications for estimating a predictor battery's classification efficiency, as these estimates (e.g., mean predicted performance [MPP]) are conditional on the number and nature of MOS on which they are based. For example, if one were to employ an expanded sample of MOS, a different sample of MOS, or both, one could obtain differing estimates of classification efficiency, especially at the individual MOS level (cf. Rosse et al., 2001; Scholarios et al., 1994). As a result, the number of MOS studied and their representativeness of the population of MOS as a whole can greatly affect estimates of classification efficiency, as well as the conclusions drawn from these estimates.

---

[25] It is presumed that the new prediction equations for each cluster would be used to make job assignments in ways similar to how the existing equations constituting the nine AA composites are used.

Therefore, the choice of MOS included in criterion-related validation studies is critical unless conclusions will be limited specifically to the MOS included in those studies. However, this is not expected to be the case with the Army.

Another factor that carries significant implications for estimating classification efficiency is the sample size per MOS. The smaller the sample size the greater the error in the estimation of key parameters – in particular the prediction equations used to assign and estimate the predicted criterion scores of recruits for some set of jobs (Rosse et al., 2001). This issue is non-trivial, because instability in these parameters could account, at least partially, for observed differences across MOS in said parameters, irrespective of any substantive differences owing to differential performance requirements (or other cross-MOS differences). Thus, failure to consider estimation error could lead to overestimates of a predictor battery's classification efficiency. In the past, estimation error has been addressed (at least partly) through large-scale simulation-based approaches (cf. Zeidner et al., 1997; Zeidner et al., 2000a), which were developed to address the simplifying assumptions (e.g., the correlations among predicted performance estimates are equal, the prediction equations for each job have equal validity, the population of people being assigned is infinite) made by analytic solutions for estimating classification efficiency (e.g., Brogden, 1959).[26]

Although informative, existing simulation-based approaches for estimating a predictor battery's classification efficiency (cf. Zeidner et al., 1997; Zeidner et al., 2000a) might not be advisable given the parameters of the typical criterion-related validation study. These approaches were developed in, and have been applied to, situations where there were large sample sizes and large numbers of MOS. Further, these approaches do not sufficiently account for all relevant sources of estimation error (e.g., error in the estimation of the prediction equations, particularly at the individual MOS level), whose effects would be compounded in the typical validation study characterized by small sample sizes per MOS and a fixed subset of MOS. Consistent with this, there are indications that estimates of classification efficiency derived from these approaches, specifically MPP, do not closely match the behavior of mean actual criterion performance under situations characterizing the typical validation study (cf. Rosse et al., 2001).

In sum, there are several approaches available for estimating a new predictor battery's classification efficiency. Ideally, over time sufficient empirical data would be available for a representative sample of MOS to permit the use of EPAS or a simulated-based approach (e.g., Zeidner et al., 1997; Zeidner et al., 2000a); which approach is ultimately employed will, and arguably should, depend on the level of fidelity and accuracy desired for real-world decision-making. Should sample size (per MOS) preclude the use of existing simulation-based approaches, alternative approaches are available (e.g., Rosse et al., 2001) that better model the estimation error associated with the sample sizes characterizing the typical validation study. In either case, it should be remembered that the needed criterion-related validity estimates do not have to be collected in a single study. Estimates of potential classification gains can be successively refined over time as more data become available; even estimates based on half of

[26] In brief, these simulation-based methods involve first estimating assignment and predicted criterion score equations empirically using different samples drawn from collected data. The assignment equations are then applied to multiple synthetically generated samples, using a linear programming algorithm to make optimal assignments to the set of jobs under consideration. Once assigned, predicted criterion scores are computed within each sample using the predicted criterion score equations. The average predicted criterion score is then aggregated across samples to obtain a mean estimate of classification efficiency, along with an estimate of its standard error (cf. Zeidner et al., 1997, 2000a).

the recommended MOS clusters (e.g., 12-15) would be informative. Should estimates be needed sooner, one could employ an approach comparable to the fourth approach proposed for generalizing (or transporting) validity (see pages 14-15), whereby one starts with a representative set of rationally-derived prediction equations, supplemented by currently available data, that are then successively refined and replaced over time by empirically-derived equations. A first, and extremely important, step in this estimation process irrespective of the approach taken is to choose which predictor battery, or batteries, offers the greatest potential to enhance classification. To make these selections, analytic solutions (e.g., Horst, 1954; Sager, Peterson, Oppler, Rosse, & Walker, 1997) should be used, or alternatives to existing solutions explored, that enable one to examine and to diagnose a battery's potential classification efficiency.

Consistent with this, the Panel concluded

Conclusion: Estimating classification efficiency (or gain) is a complex matter and can be greatly affected by changes in the number and nature of MOS sampled, sample size per MOS, and so forth. Because of this, initial estimates based on limited data must be refined and improved as more data become available. The Army should strive to meet the data requirements described above, such that simulation-based approaches based on sufficiently large sample sizes and a representative sample of Army MOS can be used to fully estimate classification gains under a variety of conditions.

Accordingly, the Panel made the following recommendations:

Recommendation 20: To empirically estimate and evaluate the potential classification gains for the entire system accruing from the use of new, alternative predictor batteries (e.g., consisting of new ASVAB subtests or measures of non-cognitive predictors), collect criterion-related validity estimates for a sufficiently representative clustering of MOS (20-30 clusters), specifically estimates from at least one focal MOS in each cluster. These validity estimates need not be obtained in a single study, but can be collected and accumulated over time. Such an incremental approach permits the successive refinement of previously derived estimates of classification gains, and the prediction equations on which they are based, as more data become available.

Recommendation 21: When estimating a predictor battery's classification efficiency, careful consideration needs be paid to the sampling of MOS in the criterion-related validations studies on which the estimates will be based, and the implications this sampling carries for inferences drawn from these estimates. Having job analysis data, as recommended, to cluster MOS and to identify focal MOS, would be useful in this regard.

Recommendation 22: To understand the impact of sample size on estimates of classification efficiency, and its implication for drawing conclusions, make use of formula and/or Monte Carlo-based approaches for modeling error in key parameters (e.g., prediction equations). For an example, see Rosse et al. (2001).

Recommendation 23: When estimating predicted criterion scores, make use of data on multiple predictors-criteria to obtain more accurate estimates of Soldiers' actual performance/ satisfaction. This can be accomplished by modeling relations among criteria and/or predictors when advisable (i.e., the interrelations reflect systematic and theoretically-relevant sources of variance), and incorporating these interrelations when estimating Soldiers' predicted criterion scores.

<u>Recommendation 24:</u> For the purposes of choosing which predictor battery (or batteries) offers the greatest potential to enhance classification, make use of analytic solutions (e.g., Horst, 1954; Sager et al., 1997), or explore alternatives to these solutions, to investigate differential validity and to diagnose potential classification efficiency.

Issue:
Does the validation of non-cognitive predictors (e.g., interests, values, temperament) raise special considerations and implications for estimating classification efficiency?

Within jobs, non-cognitive predictors (e.g., interests, values, temperament) have demonstrated non-trivial levels of predictive validity for organizational and occupational retention criteria, particularly occupational entry and commitment (Holland, 1997; Hough & Furnham, 2003). Further, non-cognitive predictors, specifically temperament (e.g., the Big Five), have exhibited greater predictive validity, relative to general mental ability (GMA), in the prediction of select types of performance-related criteria. In particular, temperament predictors have shown greater predictive potential relative to GMA for (a) ratings of non-technical (or "will do") performance (e.g., demonstrating effort, citizenship performance, peer leadership, team support) and (b) indices of career success (cf. Barrick & Mount, 1991; Hogan & Holland, 2003; Hough & Furnham, 2003; Hurtz & Donovan, 2000; Judge et al., 1999; Ozer, & Benet-Martinez, 2006). With the exception of select interests, however, non-cognitive predictors have generally not emerged as a significant contributor to classification efficiency, particularly over and above specific aptitudes (Rosse et al., 2001; Scholarios et al., 1994).

Although this finding might partly reflect the fact that there are no substantive, cross-job differences to detect, there are alternative explanations. These include, but are not limited to, (a) the choice of criterion dimension(s) assessed, and for which classification aims to optimize (e.g., technical versus non-technical, will do" components of performance); (b) the choice of the criterion method (or measure) employed to assess the dimension(s) of interest (e.g., a JKT versus MOS-specific technical ratings); (c) the basis used to cluster MOS, specifically the degree to which clustering incorporates descriptors salient to non-cognitive predictors; (d) the choice of classification models (e.g., two-stage or two-track classification) for optimizing MOS assignments, whose specific formulations may or may not capitalize on the potential of non-cognitive predictors; (d) the level of specificity at which non-cognitive predictors and criteria are measured, and whether those levels match; (e) the level of analysis (i.e., individual, team, or unit) at which non-cognitive predictors and/or criteria are measured; and (f) the greater potential for, and need to model, complex relationships (e.g., indirect, asymptotic, or curvilinear) between non-cognitive predictors and criteria. All of these factors, either individually or in combination, could explain the failure to show significant classification potential from the use of non-cognitive predictors. Comparatively speaking, the first four issues, in particular the choice of criterion dimensions assessed and the method(s) used to measure them, are arguably the most critical and immediately pressing.

When using multiple criteria, how best to combine and treat the multiple goals underlying these criteria in the optimization process (i.e., for purposes of estimating classification efficiency) could also become an issue, and potentially a very critical one. Should optimizing on the basis of alternative, non-technical criteria (e.g., MOS retention) result in different MOS assignments than from maximizing a technical performance criterion, then how multiple (and competing) goals are combined carries significant implications for determining the value added from using non-cognitive predictors in the classification process. Investigating multi-stage or multi-track classification models would be useful in this regard, as would policy capturing studies to scale the relative value to the Army of gains on different criteria.[27]

From this, the Panel concluded

Conclusion: Provided substantive cross-MOS differences exist, when estimating the classification potential of non-cognitive predictors, careful consideration needs to be paid to several factors, especially the nature of the criterion dimension(s) assessed and the method used to measure those dimension(s).

Consistent with this, the Panel recommended

Recommendation 25: When validating and investigating the classification potential of non-cognitive predictors, the Army should, at a minimum, include (a) criterion measures assessing non-technical, "will do" performance dimensions and (b) non-performance criteria (e.g., MOS satisfaction, P-O fit, retention, attrition). Regarding the latter, careful consideration needs to be paid to the nature of the method used. For example, because the effects of non-cognitive predictors on objective retention (or attrition) criteria are indirect, such criteria cannot be relied upon exclusively when estimating classification efficiency (i.e., mediators or moderators need to be modeled as well). Otherwise, one is likely to underestimate the classification potential of non-cognitive predictors.

Recommendation 26: When using multiple criteria, a critical issue will be how to treat the multiple, and potentially competing, goals underlying these different criteria (e.g., increased technical performance, increased non-technical performance, greater retention) in the optimization process (i.e., for purposes of estimating classification efficiency). Research investigating multi-stage or multi-track classification models would be useful in this regard, as would policy capturing studies to scale the relative value to the Army of gains on each criterion. One solution to this would be to start by specifying the desired levels of gain (i.e., from use of non-cognitive predictors over and above the ASVAB) that are practically significant to the Army and then determine the relative weighting that would best achieve such gains.[28]

---

[27] For an illustrative example of such a scaling exercise completed as part of Project A, see Sadacca, J. P. Campbell, DiFazio, Schultz, and White (1990).

[28] It should be noted that combining, or differentially weighting, different criteria may in itself contribute to the maximization of classification gains. For example, for highly (cognitively) complex MOS where skills are in great external demand and not easily trainable, one might estimate the weights that best predict those Soldiers likely to stay in the MOS, whereas for the less complex MOS where skills are easily trainable, one might estimate the weights that best predict non-attrition. Because the two sets of resulting weights are likely to be considerably different, using them jointly could produce greater classification efficiency than the use of either separately, provided they are not negatively correlated.

Towards a Comprehensive Solution to the Criterion Challenge: An Agenda and Roadmap

Overview

As stated previously, the Army Classification Research Panel's mission was to generate innovative, scientifically sound, and technically feasible recommendations that addressed how to

- Obtain criterion data for a sufficient number of MOS in an on-going, systematic fashion to support Army classification research.
- Ensure that the differential validity of new predictors, once established, can be generalized (or transported) to other MOS in the same job family.

Consistent with the need for a comprehensive solution, the Panel considered a number of critical issues and generated recommendations encompassing the core "building blocks" of a classification research program:

- Occupational/job analysis
- Generalizing (or transporting) validity
- Job clustering
- Criterion measurement
- Estimation of classification efficiency

In addition, because of its special importance in the Army's classification research agenda, the Panel considered the implications that the use of multiple criterion dimensions and non-cognitive predictors would have on these recommendations.

As expected, and consistent with the earlier discussion, these recommendations vary in their priority. Figure 2 summarizes the proposed near-term agenda and roadmap for solving the criterion challenge and implementing the Panel's most critical recommendations.

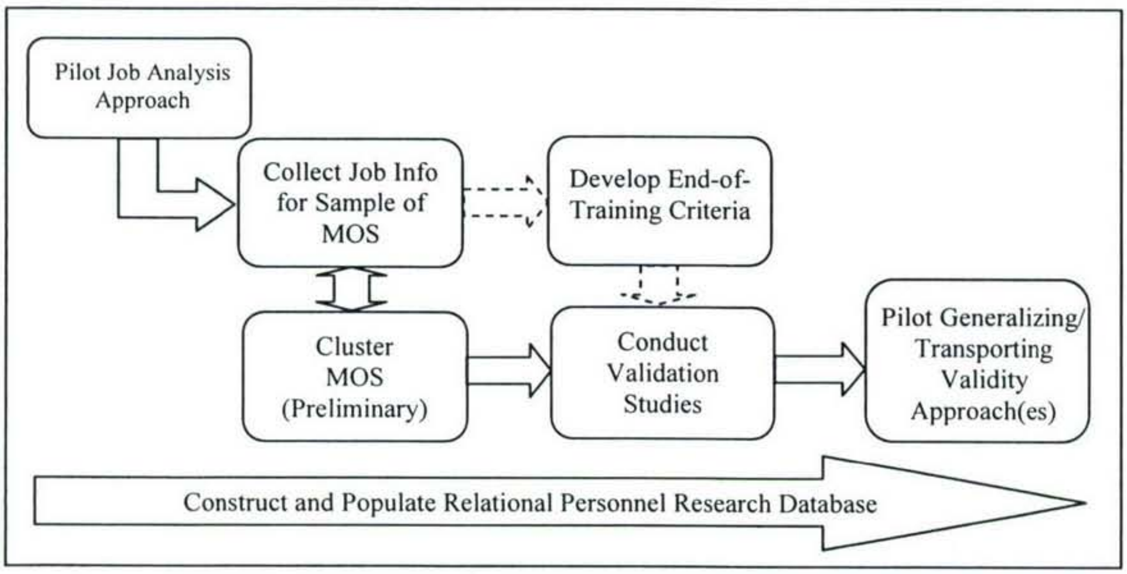

Figure 2. Proposed near-term agenda and roadmap for solving the criterion challenge.

As can be seen, the activities requiring the Army's most immediate attention and resources are (in descending order of priority):

- Piloting an Army-specific job analysis approach on 3-5 MOS.
- Constructing and populating a supporting relational database to store and organize job analysis data systematically, along with other relevant personnel research data over time and on an ongoing basis.
- Collecting data for an expanded sample of MOS using the piloted job analysis approach. The selection of MOS would best be guided by establishing (multiple) criteria against which MOS can be prioritized. These criteria should reflect a range of imperatives, from technical to policy to practical (e.g., maximization of cross-MOS differences in select performance and KSAO requirements, sufficient coverage of the Army job space, Army priorities, level of resources and SME effort needed to collect the requisite data). Once a set of criteria has been chosen, MOS could be rated against these criteria both for purposes of collecting job analysis data and for inclusion in criterion-related validation work.[29]
- Clustering the MOS sampled on the basis of the collected job analysis data, then successively refining this clustering as additional data are collected. The resulting clustering solution(s), in combination with other criteria (e.g., Army priorities), would inform which MOS to sample for the criterion-related validation work.
- Using the collected job analysis data and other information, select criterion dimensions and measures (existing or new) for use in the criterion-related validation studies. Included in this activity would be the development of well-standardized end-of-training criteria, specifically knowledge tests and (peer) ratings.
- Conducting criterion-related validation studies for the sample of focal MOS identified from the clustering solution(s) previously generated.[30]
- Using the data collected from the validation studies, piloting one or more of the proposed approaches for generalizing (or transporting) validity information.

Of these, the first two – piloting an Army-specific job analysis approach and constructing a supporting relational database – represent the most essential, as all other activities are based on and make use of information generated from their completion. The remainder of the report outlines the steps to be taken and other requirements to completing these critical activities.

[29] At a minimum, this expanded sample of MOS would need to be larger than the sample of MOS for which the planned criterion-related validation studies would be conducted on. The sample of focal MOS for the criterion-related validation work would preferably comprise 20-30 occupations. Thus, for purposes of clustering MOS to identify the focal MOS to sample for criterion-related validation work, job analysis data would be needed from about 50-60 MOS.

[30] As indicated previously, the Panel estimates this sample would preferably comprise 20-30 MOS and be informed by the previously obtained clustering results. Such a sample is expected to provide sufficient representation of the population of Army MOS and achieve a maximal level of classification efficiency (however optimized), either for use in estimating the classification potential of select predictors or for use in operationally assigning recruits to MOS.

Piloting an Army-Specific Job Analysis Approach

To meet the Army's present and future needs for its classification research program, having job analysis data represents an essential first step. More importantly, the Army's needs require a job analysis system with specific features and characteristics. In general, this system should

- Use a common language, customized to the Army context for describing similarities and differences in MOS. Specifically, this common language would consist of a reasonably comprehensive set of descriptors representing select work- (i.e., performance requirements, work/job context, machine-tools-equipment-technology) and worker-oriented (i.e., select KSAOs) domains critical to the Army's classification research needs, and sufficient for describing any MOS.
- Include cross-MOS descriptors (i.e., descriptors that can be applied across MOS) for use in making comparisons and linkages across MOS.
- Specify descriptors, in particular performance requirements, at varying levels of generality that can be organized hierarchically to support the Army's needs for job information at multiple levels of aggregation.

As mentioned, no existing, standalone job analysis system (e.g., O*NET, PAQ, CMQ) exhibits all of these features. Resultantly, the Army would have to develop such a system, although the Army need not start from scratch. Developing a system with these features carries a number of benefits. In particular, it would enable the Army to describe MOS in ways that sufficiently capture MOS-specific information and in a more efficient and cost-effective manner than might otherwise be possible using an existing system (e.g., see Appendix A). The first step in developing this system is to pilot and provide a proof-of-concept for a job analysis approach that follows the above specifications. Once prototyped, the approach could then be extended to the Army as a whole for use in supporting the Army's classification research program, most immediately to facilitate clustering MOS.

Objectives and Steps in Piloting Approach

The primary objective of this pilot study is to prototype and field test a job analysis approach using 3-5 MOS, resulting in a proof-of-concept that could then be systematically implemented Army-wide and over time at a reasonable cost.[31] Central to the work in this pilot study is the development and evaluation of alternative taxonomies of targeted work and worker-oriented descriptors, in particular performance requirements, customized to the Army for use in describing and analyzing similarities and differences in MOS. When completed, this pilot study should produce the following:

- Descriptor taxonomies for a select set of work- and worker-oriented domains customized to the Army context, hierarchically organized according to well-defined rules. These taxonomies would include cross-MOS descriptors that could be applied

[31] In designing the prototype job analytic approach, the aim is to develop a reasonably sound approach that can be feasibly applied to all relevant MOS, and provided sufficient funds are available, the entire population of MOS.

across MOS, either within or across job families. The domains for which taxonomies would be developed are (in order of priority):[32]

  - o Performance requirements
  - o Work/job context and interests/values/temperament
  - o Occupation-specific knowledges and skills, to include machine-task-equipment-technology
  - o Abilities

- A standardized procedure and specifications for formulating MOS-specific performance requirements that sufficiently balance the resources needed to collect this information against the usefulness of the resulting data.
- Provided resources are available to investigate these issues, specifications on how the requisite job analysis data are best elicited and collected, similarly in ways that sufficiently balance the resources expended against the usefulness of the data collected. In particular, these specifications would address the following:
  - o Which existing source materials (e.g., Soldier manuals, MOS training curricula and objectives, existing task inventories) should be used, and how much weight should be placed on them?
  - o Which SMEs should be consulted (i.e., Soldiers of what rank, levels of experience, exposure to more than one MOS, and so forth)? Should SMEs other than Soldiers (e.g., psychologists) be included, and for what purposes?
  - o What methods of collecting information and/or judgments should be included in the approach? That is, (a) how exactly should SME sessions be run and (b) what are the descriptors to be presented? What methods will be used to elicit quantitative judgments about the descriptors from the SMEs?
  - o Do the answers to these issues differ by descriptor domain?

Further, and more practically, the results of this pilot should make clear the level of effort that will be required in gathering the requisite job analysis data across MOS over time.

The pilot could follow a number of specific designs. In general, conducting the pilot should involve the following activities:

1. *Select sample of 3-5 MOS.* These MOS could be selected on the basis of multiple criteria: (a) maximizing differentiation on performance and select KSAO requirements, (b) Army priorities (e.g., high-density MOS, MOS that are difficult to recruit or train Soldiers for, anticipated future need), (c) amount and quality of existing job information available, (d) resources and level of SME effort required to analyze, and (e) existing plans for including MOS in criterion-related validation work. To capitalize on past effort expended, one possibility would be to focus on the MOS sampled for the PerformM21 project.

[32] As indicated previously and as suggested here in these recommendations, there is the possibility that one or more of the taxonomies developed could incorporate multiple domains into the same taxonomy for purposes of generating descriptors that capture MOS-specific information in new and potentially more powerful and multi-faceted ways than existing taxonomies (cf. Dietrich et al., 2002; National Center for O*NET Development, 2003).

2. *Collect and inventory existing information on MOS-specific task requirements. Formulate a standardized procedure and specifications for generating MOS-specific performance requirements*. Information on existing MOS-specific performance requirements could come from Soldier manuals, MOS training curricula and objectives, existing task inventories, and recent job analysis work (e.g., PerformM21). Provided the information is sufficiently current and comprehensive, this information (once compiled) could be used to formulate a standardized procedure and specifications for generating MOS-specific performance requirements that sufficiently balances the resources expended to collect this information against the usefulness of the resulting descriptors. In formulating this procedure, there are several specific issues to be addressed. They are (a) how to handle the common task domain for non-combat specialties, (b) how to incorporate information about work context (or conditions) under which the tasks must be performed (e.g., task by task, or as a separate category of information), (c) how should theater or mission-specific requirements be handled; and (d) how much and at what level of detail should the specific performance requirements be specified. Another critical issue to be addressed is the specification of non-technical performance requirements (e.g., peer leadership, teamwork), as this information is likely to be missing from existing Army materials and task inventories. A subset of performance requirements for each MOS should be written in several ways so that the resulting task statements vary according to the characteristics listed above. Prototypes for these specific task statements can be found in PerformM21 and Project A. A panel of professional job analysts (e.g., psychologists from ARI) would evaluate the utility of the alternative representations for supporting Army classification research, specifically for clustering MOS. A panel of proponent SMEs should review the alternatives to judge whether one or more of them misrepresent the task content of the MOS. The end products of this activity should be (a) a standardized procedure and specifications for formulating MOS-specific performance requirements, and (b) lists of specific performance requirements, formulated in accordance with the proposed procedure and specifications, for the 3-5 MOS sampled for subsequent use in developing the hierarchical performance requirements taxonomy.

3. *Using a combined top-down and bottom-up approach, determine the preferred method, levels of aggregation, and specifications for clustering tasks hierarchically into a performance-requirements taxonomy.*[33] In brief, completing this activity should consist of the following. First, using existing taxonomies of general technical performance requirements (e.g., task categories from the Army SYNVAL project, PerformM21, Select21) and of non-technical requirements (e.g., O*NET's Generalized Work Activities [GWAs]), existing taxonomies from relevant literature in leadership and teamwork, Army research on critical incidents), have a panel of professional job analysts formulate an initial taxonomy of higher-order, general performance categories, sufficiently comprehensive and customized to the Army context – the resulting taxonomy would probably consist of 50-100 performance categories.[34] Second, have proponent SMEs rationally sort the specific performance requirements (generated in the preceding activity) into this initial taxonomy. Use the results of this sorting exercise to refine the taxonomy. Third, have SMEs (proponent or professional job analysts)

[33] In general, the "optimal" number of levels should be sufficient to capture both the specificity needed for criterion development and the generality needed for generalizing validity and clustering, without being too cumbersome to develop or collect data on.

[34] For listings of the task categories from the Army SYNVAL and PerforrmM21 projects, see Appendix B.

for each MOS sort the specific performance requirements into performance categories that maximize the homogeneity of content within categories and minimize overlap between categories. The category assignments could then be transformed into a matrix of similarities between pairs of tasks and then re-clustered empirically. From these data, performance categories would be developed within MOS. These categories could then be used by SMEs to sort specific performance requirements across MOS. As before, the resulting co-occurrence matrices could be re-clustered empirically. Finally, have SMEs repeat this procedure with the higher-order, general performance categories. When finished, the results from both the top-down and bottom-up approaches would form the basis for the preferred method, levels of aggregation, and specifications (e.g., all general performance categories must subsume at least two specific performance requirements, the number of specific performance requirements should not exceed X) for clustering tasks into a hierarchical taxonomy. The end product of this activity should be a complete method and specifications for clustering tasks hierarchically into a performance requirements taxonomy that can be systematically and consistently applied across MOS.

4. *Using a similar top-down and bottom-up approach, develop taxonomies for the other work and worker-oriented domains as prioritized above, starting with work/job context and interests/values/temperament.* Using a similar top-down and bottom-up approach, a panel of professional job analysts should develop taxonomies for other select domains, as prioritized. Preferably, for each domain, there would be multiple alternative taxonomies (at least three) that could be compared and contrasted.[35] For example, the development of the work/job context taxonomy could start with several existing taxonomies (e.g., O*NET, PAQ). These initial taxonomies could then be refined and customized to the Army context using existing critical incident information collected from previous Army projects, from the formulation of the specific non-technical tasks for the 3-5 sampled MOS, or both. Similarly, a taxonomy of interests, sufficiently comprehensive and specific to the Army, could be developed by first organizing the higher-order, general task categories around an existing interest taxonomy (e.g., RIASEC; Holland, 1997) to formulate interests that differentiate across Army MOS. These initial taxonomies could then be refined using the same critical incident information used for the work/job context taxonomies. For each domain, a panel of professional job analysts should use each taxonomy to judge the requirements for each of the 3-5 sampled MOS. On the basis of this initial evaluation the alternative representations should be revised as appropriate and then used more formally to judge the criticality of the applicable work- or worker-oriented requirements for each of the 3-5 MOS in the pilot. The end product of this activity should be taxonomies for select work- and worker-oriented domains that can be systematically and consistently applied across MOS.

5. *Using the findings and specifications from the pilot study, conduct a field test of prototyped job analysis approach on an expanded sample of MOS.* As indicated above, once the pilot study has been completed, and a supporting relational database for storing and organizing job analysis data has been developed, the next step would be to field test the prototyped job analysis approach on an expanded sample of MOS.

[35] As with performance requirements, the work- or worker-oriented requirements constituting these domains could be defined at multiple levels of specificity, although that may vary depending on the nature of the domain in question.

Constructing and Populating a Supporting Relational Database

Critical to supporting the proposed occupational/job analysis system and the Army's classification research needs in general is a relational database.[36] As discussed, the primary purpose of this database is to capture and store the job analysis data needed to support and advance the Army's personnel management objectives, in particular classification, in an ongoing fashion. When combined with other relevant data (e.g., criterion-related validity estimates, Soldier predictor-criterion data), these job data can be used to generate and refine solutions to the Army's classification needs. More specifically, when populated with these data, the database could be used to [37]

- Pilot and validate a method for generalizing (or transporting) criterion-related validity estimates collected on a sample of MOS to other MOS (e.g., for use in creating job-specific prediction equations for MOS lacking criterion data).
- Cluster MOS on the basis of one or more of the descriptors constituting the piloted job analytic approach (e.g., for purposes of aggregating validity estimates, or prediction equations, for use in estimating classification gains for the entire system or to identify which MOS to sample for criterion-related validation studies).
- Document changes in MOS over time and examine their implications for the use (or continued use) of previously collected validity estimates.
- Aggregate Soldier data on the same predictor-criterion combinations across different studies to increase the sample size available, either within an MOS or across MOS, for purposes of developing job-specific prediction equations or for estimating classification gains.

Populating this database would follow an incremental approach. Generally, this would involve starting with existing information (e.g., Soldier manuals, MOS training curricula and objectives, existing task inventories, criterion-related validity estimates and Soldier-level predictor-criterion data from past research, such as Project A, Select21) and then updating or supplementing that information as new data are collected. Taking an incremental approach (a) balances demands on Army resources while providing the Army with sufficiently sound and reasonably viable solutions to its classification research needs, and (b) allows for these solutions to be successively refined over time as more, or newer, data become available.

In general, constructing and populating the relational database would involve the following activities:

1. *Design the database. Identify the requisite data elements.* As indicated, this relational database aims to serve as the focal point for accumulating occupational/job analysis data and other personnel data critical to meeting the Army's classification research needs. At a minimum, the database should contain (a) MOS-specific and cross-MOS occupational/job analysis data on select work- and worker-oriented descriptors, (b) empirical or synthetically

[36] Like the proposed job analysis approach, this database is not intended to replace any existing Defense Manpower Data Center (DMDC) and/or Army personnel management databases (e.g., ATRRS). Rather, the intention of this database is to serve as a focal point where relevant personnel management data can be combined into a single, integrative source.

[37] It should be noted that the benefits and uses of this database, and in particular the job analysis data, are not restricted to the research applications outlined here.

derived validity estimates and prediction equations (e.g., from Army SYNVAL, Project A, Zeidner and Johnson's Differential Assignment Theory [DAT] research program) at both the MOS and job family level, and (c) predictor-criterion data from individual Soldiers (e.g., from Project A, Select21, the Army Class Concurrent Validation [CV] study).[38]

2. *Develop the database.* Once the data elements have been identified and sufficiently specified, development can begin. Even if not all the data elements previously identified will be populated from the start, it is important that they be built into the database.

3. *Start to populate the database.* As discussed, populating the database would follow an incremental approach (i.e., start with existing information and then update or add data as they become available). Populating the database may require establishing memoranda of agreement with select Army components for the continued and ongoing collection of relevant data. At some point, having a means for standardization and scaling of different predictor-criterion measure combinations will become important. Developing such a mechanism is technically feasible and can be done at reasonable cost.

Piloting the proposed job analysis approach and developing the supporting relational database could be completed simultaneously. As these two activities will likely require different project teams each possessing different skills sets, it is critical that there be coordination between the two teams during the design and development phases.

[38] Predictor-criterion data need not be limited exclusively to individual-level data. Team- or unit-level data (e.g., team or unit effectiveness, team or unit cohesion) could also prove useful.

References

Arthur, W., Jr., Doverspike, D., & Barrett, G. V. (1996). Development of a job analysis-based procedure for weighting and combining content-related tests into a single test battery score. *Personnel Psychology, 49, 971-985.*

Barrick, M. R., & Mount, M. K. (1991). The Big-Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1-26.

Brogden, H. E. (1959). Efficiency of classification as a function of the number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. *Educational and Psychological Measurement*, *19*, 181-190.

Campbell, J. P. (Ed.) (1987). *Improving the selection, classification, and utilization of Army enlisted personnel – Annual report, 1985 fiscal year* (TR 746). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P, & Knapp, D. J. (Eds.) (2001). *Exploring the limits in personnel selection and classification.* Mahwah, NJ: Lawrence Erlbaum Associates.

Campbell, J. P., Russell, T. L., & Knapp, D. J. (1993). *ROADMAP: An agenda for joint-service classification research* (FR 93-23). Alexandria, VA: Human Resources Research Organization.

Campbell, J. P., & Zook, L. M. (Eds.) (1991). *Improving the selection, classification, and utilization of Army enlisted personnel: Final report on Project A* (TR 1597). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Cornelius, E. T., Carron, T. J., & Collins, M. N. (1979). Job analysis models and job classification. *Personnel Psychology*, *32*, 693-708.

Dietrich, E., Hendrickson-Larson, J., Hoppe, R., Paige, B., & Rosenow, S. (2002). *Labor Exchange Skills Project.* Washington, DC: U.S. Department of Labor, Employment and Training Administration.

Guion, R. M. (1965). Synthetic validity in a small company: A demonstration. *Personnel Psychology*, *18*, 49-63.

Hoffman, C. C., Holden, L. M., & Gale, K. (2000). So many jobs, so little "N": Applying expanded validation models to support generalization of cognitive test validity. *Personnel Psychology*, *53*, 955-991.

Hoffman, C. C., & McPhail, S. M. (1998). Exploring options for supporting test use in situations precluding local validation. *Personnel Psychology*, *51*, 987-1003.

Hogan, J., Davies, S., & Hogan, R. (n press). Generalizing personality-based validity evidence. In S. M. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validity evidence*. San Francisco, CA: John Wiley & Sons, Inc.

Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalysis perspective. *Journal of Applied Psychology, 88*, 100-112.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources, Inc.

Hollenbeck, J. R., & Whitener, E. M. (1988). Criterion-related validation for small sample contexts: An integrated approach to synthetic validity. *Journal of Applied Psychology, 73*, 536-544.

Horst, P. (1954). A technique for the development of a differential prediction battery. *Psychological Monographs: General and Applied, 68*, (9, No. 390).

Hough, L. M., & Furnham, A. (2003). Use of personality variables in work settings. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of personality: Vol. 12. Industrial and organizational psychology* (pp. 131-169). Hoboken, NJ: John Wiley & Sons.

Human Resources Research Organization (2005). *Future job clusters* (RP 2005-04). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology, 85*, 869-879.

Johnson, C. D., Zeidner, J., & Leaman, J. A. (1992). *Improving classification efficiency by restructuring Army job families.* (TR 947). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Johnson, J. W. (in press). Synthetic validity: A technique of use (finally). In S. M. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validity evidence*. San Francisco: John Wiley & Sons, Inc.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*, 621-652.

Knapp, D. J., & Campbell, R. C. (Eds.) (2006). *Army enlisted personnel competency assessment program: Phase II report* (TR 1174). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lightfoot, M. A. & Ramsberger, P. F. (2000). *Matching recruits to jobs: The enlisted personnel allocation system (EPAS)* (SR 41). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Maier, M. H. and Fuchs, E.F. (1972). Development and Evaluation of a New ACB and Aptitude Area System (Technical Research Note 239). Arlington, VA: U.S. Army Behavior and Systems Research Laboratory.

McCloy, R. A. (1994). Predicting job performance scores for jobs without performance data. In B. F. Green & A. S. Mavor (Eds.), *Modeling cost and performance for military enlistment* (pp. 61-99). Washington, DC: National Academy Press.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*, 493-505.

McCormick, E. J. (1959). The development of processes for indirect or synthetic validity: III. Application of job analysis to indirect validity. A symposium. *Personnel Psychology, 12*, 402-413.

McCormick, E. J., DeNisi, A. S., & Shaw, J. B. (1979). Use of the position analysis questionnaire for establishing the job component validity of tests. *Journal of Applied Psychology, 64*, 51-56.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. (1972). A study of job characteristics and job dimensions based on the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology, 56*, 347-368.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*, 475-480.

Murphy, K. R., & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology, 50*, 823-854.

National Center for O*NET Development. (2003). *Summary report: Updating the Detailed Work Activities*. Alexandria, VA: Author.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401-421.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., & Fleishman, E. A. (Eds.) (1999). *An occupational information system for the 21st century: The development of O*NET*. Washington, DC: American Psychological Association.

Peterson, N. G., Owens-Kurtz, C., Hoffman, R. G., Arabian, J. M., & Whetzel, D. L. (Eds.) (1990). *Army Synthetic Validation Project: Report of Phase II Results, Volume I* (TR 892). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Peterson, N. G., Wise, L. L., Arabian, J., & Hoffman, R. G. (2001). Synthetic validation and validity generalization: When empirical validation is not possible. In J. P Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 411-452). Mahwah, NJ: Lawrence Erlbaum Associates.

Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior, 21*, 259-287.

Prediger, D. J., & Swaney, K. B. (2004). Work task dimensions underlying the World of Work: Research results for diverse occupational databases. *Journal of Career Assessment, 12*, 440-459.

Raymark, P. H., Schmit, M. J., & Guion, R. M. (1997). Identifying potentially useful personality constructs for employee selection. *Personnel Psychology, 50*, 723-736.

Reynolds, D. H., Laabs, G. J., & Harris, D. A. (1996). Occupational genealogies: Using job family characteristics in personnel research. *Military Psychology, 8*, 195-217.

Rosse, R. L., Campbell, J. P., & Peterson, N. G. (2001). Personnel classification and differential job assignments: Estimating classification gains. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 453-506). Mahwah, NJ: Lawrence Erlbaum Associates.

Rounds, J., Smith, T., Hubert, L., Lewis, P., & Rivkin, D. (1999). *Development of occupational interest profiles for O*NET*. Raleigh, NC: O*NET Consortium.

Sackett, P. R. (1991). Exploring strategies for clustering military occupations. In A. K. Wigdor & B. F. Green, Jr. (Eds.), *Performance assessment for the workplace, Volume II: Technical issues* (pp. 305-332). Washington, DC: The National Academies Press.

Sadacca, R., Campbell, J. P., DiFazio, A. S., Schultz, S. R., & White, L. A. (1990). Scaling performance utility to enhance selection/classification decisions. *Personnel Psychology, 43*, 367-379.

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

Sager, C. E., Russell, T. L., Campbell, R. C., & Ford, L. A. (2005). *Future Soldiers: Analysis of Entry-Level Performance Requirements and their Predictors* (TR 1169). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Scherbaum, C. A. (2005). Synthetic validity: Past, present, and future. *Personnel Psychology, 58*, 481-515.

Scholarios, D., Johnson, C. D., & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology, 79*, 412-424.

Wilk, S. L., Desmarias, L., & Sackett, P. R. (1995). Gravitation to jobs commensurate with ability: Longitudinal and cross-sectional tests. *Journal of Applied Psychology*, *80*, 79-85.

Wise, L. L., McHenry, J. J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. *Personnel Psychology*, *43*, 355-366.

Wise, L. L., Peterson, N. G., Hoffman, R. G., Campbell, J. P., & Arabian, J. M. (Eds.) (1991a). *Army Synthetic Validation Project, Report of Phase III Results, Volume I* (TR 922). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Wise, L. L., Peterson, N. G., Hoffman, R. G., Campbell, J. P., & Arabian, J. M. (Eds.) (1991b). *Army Synthetic Validation Project, Report of Phase III Results, Volume II: Research Instruments* (RP 91-08). Alexandria, VA. U.S. Army Research Institute for the Behavioral and Social Sciences.

Zeidner, J., & Johnson, C. D. (1994). Is personnel classification a concept whose time has passed? In M. G. Rumsey, C. B. Walker, & J. B. Harris (Eds.), *Personnel selection and classification* (pp. 377-410). Hillsdale, NJ. Lawrence Erlbaum Associates.

Zeidner, J., Johnson, C. D., & Scholarios, D. (1997). Evaluating military selection and classification systems in the multiple job context. *Military Psychology, 9,* 169-186.

Zeidner, J., Johnson, C. D., Vladimirsky, Y., & Weldon, S. (2000a). *Specifications for an operational two-tiered classification system for the Army, Volume 1: Report* (TR 1108). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Zeidner, J., Johnson, C. D., Vladimirsky, Y., & Weldon, S. (2000b). *Specifications for an operational two-tiered classification system for the Army, Volume 2: Appendixes* (RN 2000-12). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Appendix A:

Examples of Selected Detailed Work Activities (DWAs) that Combine Different Descriptors, Organized within O*NET's Generalized Work Activities (GWAs)[39]

Job A: Aircraft Engine Specialist

GWA: Repairing and Maintaining Mechanical Equipment

- adjust or set mechanical controls or components [1]
- align or adjust clearances of mechanical components or parts
- assemble, dismantle, or reassemble equipment or machinery
- conduct tests to locate mechanical system malfunction
- diagnose mechanical problems in machinery or equipment
- dismantle or reassemble rigging
- inspect machinery or equipment to determine adjustments or repairs needed
- lubricate machinery, equipment, or parts
- maintain welding machines or equipment
- repair aircraft ignition or ignition systems
- test mechanical products or equipment

GWA: Controlling Machines and Processes

- operate hoist, winch, or hydraulic boom [2]
- set up and operate variety of machine tools
- use electrical or electronic test devices or equipment
- use hand or power tools
- use lifting equipment in vehicle repair setting

Job B: Avionics Technician

GWA: Controlling Machines and Processes

- operate industrial or nondestructive testing equipment
- solder electrical or electronic connections or components
- use precision measuring tools or equipment
- use soldering equipment
- use voltmeter, ammeter, or ohmmeter

GWA: Identifying Objects, Actions, and Events

- understand detailed electronic design specifications [3]
- understand technical information for electronic repair work [3]
- understand technical operating, service or repair manuals [4]

[1] example of general task requirement
[2] example of MTE statement
[3] example of occupation-specific knowledge statement
[4] example of use of information/materials/resources statement

[39] For additional information, see: Dietrich, Hendrickson-Larson, Hoppe, Paige, and Rosenow (2002); The National Center for O*NET Development (2003).

Appendix B:

Performance (Task) Categories from the Army SYNVAL Project[40]

I. Maintenance
   A. Mechanical Systems Maintenance
      1. Perform operator maintenance checks and service
      2. Perform operator checks and services on weapons
      3. Troubleshoot mechanical systems
      4. Repair weapons
      5. Repair mechanical systems
      6. Troubleshoot weapons
   B. Electrical and Electronic Systems Maintenance
      1. Install electronic components
      2. Inspect electrical systems
      3. Inspect electronic systems
      4. Repair electrical systems
      5. Repair electronic components
II. General Operations
   A. Pack and Load
      1. Pack and load materials
      2. Prepare parachutes
      3. Prepare equipment and supplies for air drop
   B. Vehicle and Equipment Operations
      1. Operate power excavating equipment
      2. Operate wheeled vehicles
      3. Operate track vehicles
      4. Operate boats
      5. Operate lifting, loading, and grading equipment
   C. Construct/Assemble
      1. Paint
      2. Install wire and cable
      3. Repair plastic and fiberglass
      4. Repair metal
      5. Assemble steel structures
      6. Install pipe assemblies
      7. Construct wooden buildings and other structures
      8. Construct masonry buildings and structures
   D. Technical Procedures
      1. Operate gas and electric powered equipment
      2. Select, layout, and clean medical/dental equipment and supplied
      3. Use audiovisual equipment
      4. Reproduce printed material
      5. Operate electronic equipment

[40] From Wise, Peterson, R. G. Hoffman, Campbell, and Arabian (1991a). For copy of questionnaire to rate MOS on these performance (task) categories, see Attachment 5 in Wise, Peterson, R. G. Hoffman, Campbell, & Arabian (1991b).

Performance (Task) Categories from the Army SYNVAL Project (cont'd)

6. Operate radar
7. Operate computer hardware
8. Cook
9. Perform medical laboratory procedures
10. Conduct land surveys
11. Provide medical or dental treatment

E. Make Technical Drawings
1. Sketch maps, overlays, or range cards
2. Produce technical drawings
3. Draw maps and overlays
4. Draw illustrations

III. Administrative

A. Clerical
1. Type
2. Prepare technical forms and documents
3. Record, file, and dispatch information
4. Receive, store, and issue supplies, equipment, other materials

B. Communication
1. Use hand and arm signals
2. Read technical manuals, field manuals, regulations, and other publications
3. Use maps
4. Send and receive radio messages
5. Give oral reports
6. Receive clients, patients, guests
7. Give directions and instructions
8. Write documents and correspondence
9. Write and deliver presentations
10. Interview
11. Provide counseling and other interpersonal interventions

C. Analyze Information
1. Decode data
2. Analyze electronic signals
3. Analyze weather conditions
4. Order equipment and supplies
5. Estimate time and cost of maintenance operations
6. Plan placement or use of tactical equipment
7. Translate foreign languages
8. Analyze intelligence data

D. Applied Math and Data Processing
1. Control money
2. Determine firing data for indirect fire weapons
3. Compute statistics or other mathematical calculations
4. Provide programming and data processing support for computer operations

E. Control Air Traffic
1. Control air traffic

Performance (Task) Categories from the Army SYNVAL Project (cont'd)

IV. Combat
  A. Individual Combat
    1. Use hand grenades
    2. Protect against NBC hazards
    3. Handle demolitions or mines
    4. Engage in hand-to-hand combat
    5. Fire individual weapons
    6. Control individuals and crowds
    7. Customs and laws of war
    8. Navigate
    9. Survive in the field
    10. Move and react in the field
  B. Crew-Served Weapons
    1. Load and unload field artillery or tank guns
    2. Fire heavy direct fire weapons (e.g., tank main guns)
    3. Prepare heavy weapons for tactical use
    4. Place and camouflage tactical equipment and materials in the field
    5. Fire indirect weapons (e.g., field artillery)
  C. Give First Aid
    1. Give first aid
  D. Identify Targets
    1. Detect and identify targets

V. Supervision (not included in any of the four other major categories)
    1. Plan Operations
    2. Direct/Lead Teams
    3. Monitor/Inspect
    4. Lead
    5. Act as a Model
    6. Counsel
    7. Communicate
    8. Train
    9. Personnel Administration

Performance Categories from PerformM21 for MOS 63B (Wheeled Vehicle Mechanic)[41]

Performance Requirements

A. Engine – Lubrication, fuel, exhaust, and cooling system

1. Service engine assembly.
2. Replace engine oil filter
3. Correct malfunctions in oil cooler and lines.
4. Troubleshoot and correct malfunctions in fuel system.
5. Correct malfunctions in fuel pump.
6. Replace fuel lines and fittings, fuel filter assembly, fuel tank.
7. Troubleshoot and correct malfunction of glow plug system.
8. Troubleshoot exhaust system and replace muffler and crossover pipe.
9. Troubleshoot cooling system and replace radiator, hoses, lines and clamps.
10. Correct malfunctions of fan, fan drive, and drive belts.

B. Electrical – Engine, instrument panel, wiring harness systems

1. Troubleshoot charging system.
2. Correct malfunctions of alternator.
3. Troubleshoot starter system and replace starter.
4. Troubleshoot malfunctions of electrical system.
5. Replace protective control box.
6. Correct malfunctions of sending units and warning switches.
7. Correct malfunction of batteries.
8. Troubleshoot electrical gauges.
9. Repair engine and chassis wiring harness.
10. Correct malfunctions of 100 amp alternator.

C. Power Train – Transmission, transfer, propeller shafts, axles and components

1. Troubleshoot and service transmission.
2. Replace neutral safety switch.
3. Troubleshoot transfer.
4. Replace propeller shafts, universal joints, and center bearings.
5. Troubleshoot axles.
6. Replace front axle spindle.
7. Replace halfshaft.
8. Correct malfunction of geared hub and knuckle.
9. Adjust geared hub spindle bearing.
10. Replace upper and lower ball joints.
11. Replace CV boot assembly.

[41] For additional information, see Knapp and R. C. Campbell (2006).

Performance Categories from PerformM21 for MOS 63B (Wheeled Vehicle Mechanic) (cont'd)

D. Chassis – Brakes, wheels and hubs, steering, springs and shocks, body, winch components, central tire inflation system (CTIS)

1. Troubleshoot brake system.
2. Replace brake lines and fittings.
3. Replace hand brake shoes.
4. Replace service brake shoes.
5. Replace front and rear brake pads, calipers, and rotors.
6. Replace master cylinder and hydro-boost.
7. Replace air hydraulic cylinder and treadle valve.
8. Replace air compressor and belts; inspect air brake control valves.
9. Correct malfunctions of wheel and tire assemblies.
10. Troubleshoot steering system.
11. Correct malfunction of tie rod assembly.
12. Correct malfunction of drag link assembly.
13. Correct malfunction of power assist cylinder; replace power steering lines and fittings.
14. Replace shock absorbers.
15. Replace seat belts.
16. Troubleshoot winch.
17. Troubleshoot and correct malfunctions on central tire inflation system (CTIS).

E. General Maintenance – Test equipment, tool kits, preventive maintenance

1. Maintain test, measurement, and diagnostic equipment (TMDE)
2. Maintain assigned vehicle.
3. Maintain toolkit.
4. Prepare equipment inspection maintenance worksheet.
5. Perform scheduled preventive maintenance checks and services (PMCS).