ARI Research Note 88-107

A Review of Models and Procedures for Synthetic Validation for Entry-Level Army Jobs

Jennifer L. Crafts, Philip L. Szenas, Wei Jing Chia, and Elaine D. Pulakos
American Institutes for Research

for

Contracting Officer's Representative
Jane M. Arabian

Selection and Classification Technical Area
Lawrence M. Hanser, Chief

Manpower and Personnel Research Laboratory
Newell K. Eaton, Director

December 1988

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

American Institutes for Research

Technical review by

Paul Towhig
Ilene Gast

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Informational Service (NTIS).

FINAL DISPOSITION: This 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 views, opinions, and findings in this report are those of the author(s) and should not to be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) --	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 88-107
6a. NAME OF PERFORMING ORGANIZATION American Institutes for Research 6b. OFFICE SYMBOL (If applicable) WRC	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) 1055 Thomas Jefferson Street, N.W., Suite 200 Washington, DC 20007	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Defense Supply Service 8b. OFFICE SYMBOL (If applicable) DSS	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA 903-87C-0525
8c. ADDRESS (City, State, and ZIP Code) Room 1D-245, The Pentagon Washington, DC 20310-5220	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 6.27.85 | PROJECT NO. A791 | TASK NO. 2.3.1 | WORK UNIT ACCESSION NO. C1

11. TITLE (Include Security Classification)
A Review of Models and Procecures for Synthetic Validation for Entry-Level Army Jobs

12. PERSONAL AUTHOR(S)
Crafts, Jennifer L., Szenas, Philip L., Chia, Wei Jing, Pulakos, Elaine D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
	FROM 87-4 TO 88-8	1988, December	102

16. SUPPLEMENTARY NOTATION
Jane M. Arabian, Contracting Officer's Representative.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Job component models; Job taxonomy; Expert judgments; Synthetic validation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
This review presents relevant literature in the areas of synthetic validation, job component models, and expert judgments. Synthetic validation is a logical process for inferring test battery validity from elemental test validities for job components. The review describes a number of key linkages required for synthetic validation, presents a model of the steps to establish linkages, reviews and evaluates synthetic validation studies in terms of how completely linkages were established and empirically supported, and concludes that sound techniques exist to investigate synthetic validation as a basis for test development for entry-level Army jobs. The review introduces issues important to the choice of a job component model that will describe more than 250 entry-level Army jobs and defines four basic ways to describe and classify jobs. Models representing each approach are discussed and evaluated in terms of applicability for the Army's synthetic validation efforts. Issues related to reliability and validity of expert judgments are discussed.

(Continued)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Jane Arabian	22b. TELEPHONE (Include Area Code) (703) 274-8275	22c. OFFICE SYMBOL

DD Form 1473, JUN 86 — Previous editions are obsolete.

ARI Research Note 88-107

19. ABSTRACT (Continued)

The report concludes with suggestions for development of job component models, judgment procedures, and synthetic validation procedures specific to requirements of the Synthetic Validation Project.

ACKNOWLEDGMENTS

The authors sincerely thank members of the Synthetic Validation Project Core Development Team for their thorough and helpful reviews of the initial draft of this paper: Jane Arabian, John Campbell, Gene Hoffman, Jeff McHenry, Norm Peterson, and Laurie Wise. We would also like to acknowledge the invaluable contributions of George Wheaton, who reviewed and edited the final draft, and Cyndi Owens-Kurtz, who helped review and summarize articles for this review.

A REVIEW OF MODELS AND PROCEDURES FOR SYNTHETIC VALIDATION FOR ENTRY-LEVEL ARMY JOBS

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

A REVIEW OF MODELS AND PROCEDURES FOR SYNTHETIC VALIDATION FOR ENTRY-LEVEL ARMY JOBS

BACKGROUND AND OVERVIEW

The concept of validity has changed in the last ten years. Guion (1976) admonished that the inability to generalize validity precludes the development of general principles and theories necessary to elevate the field of personnel psychology beyond mere technology to the status of a science. The traditional view of validity--that employment test validities are situation-specific--received wide support in the literature. Wernimont (1962; cited in Cascio, 1987) demonstrated that the validity of a weighted application blank shrank from .74 to .07 in only three years. Such evidence helped to perpetuate the belief that demonstrated validity of a test for a group of jobs in one company would not be the same for the same type of jobs in another company, or even similar jobs within the same company.

Schmidt and Hunter (1977; 1984; Schmidt, Hunter, & Pearlman, 1981) have conducted research that has changed the view of situational validity. These authors have shown that much if not all of the variation in test validities across locations may be due to such statistical artifacts as range restriction, criterion and predictor unreliability, and small sample sizes. Thus, rather than supporting the situation-specific hypothesis, this research supported the generalization of test validities for similar jobs across companies and time.

Although numerous studies have supported the notion that validity is generalizable, there has been surprisingly little work on development of a related methodology--synthetic validation. Synthetic validation is a set of methodological techniques for inferring test-battery validity from predetermined test validities for specific components of the total job. Since its inception, synthetic validation has been a much discussed but little practiced technique. The recent neglect of synthetic validation can possibly be attributed to the complexity of implementing synthetic validation procedures. A successful synthetic validation project requires: (a) a procedure for describing each job under consideration in terms of generalizable job components that cover all important aspects of each job; (b) the identification of predictor measures that relate to job components and procedures to calculate the validity of these predictors for each job component; (c) procedures to combine the component validities into an overall validity estimate; and (d) means to evaluate these overall validity estimates against some external criteria. Due to the cost involved in conducting synthetic validation research, particularly in developing the job component taxonomy and developing external criteria, a study is most feasible when extensive empirical validity data exist within a large organization (or across several similar organizations), and the goal is to extend these data for other jobs not covered in prior research. With all these qualifications in mind, it is not surprising that a large-scale synthetic validation project would be undertaken to build upon and extend the results of the largest selection and classification research project in the history of industrial and organizational psychology.

Project A

The Army Research Institute (ARI) is currently involved in a large-scale, multi-year research program known as Project A. The intent of Project A is to develop and improve the selection and classification system for initial assignment of enlisted persons to Military Occupational Specialties (MOS) or jobs. One goal of Project A is to determine whether the Armed Services Vocational Aptitude Battery (ASVAB)--the instrument currently used for selection and placement--is valid for predicting actual job performance, in addition to predicting success in training. A major premise of this effort is that recruits should be selected on the basis of knowledge, skills, abilities, and other personal characteristics (KSAOs) that are required for successful performance on the job. Toward this end, ARI has undertaken Project A to examine the predictive validity of an expanded set of measures--including psychomotor and perceptual tests and interest, biographical, and temperament measures--for predicting job performance in 21 different jobs. The predictive validity of the new and existing selection measures is being assessed against a comprehensive set of job performance measures, as well as against success in completing advanced training for the job.

Much of the expected gain from an improved selection and classification system is based on the assumption that somewhat different KSAOs are required for performance in different MOS. Without such differences, improvements might be made in overall selection procedures, but there would be very little prospect for gains in our ability to classify recruits into different MOS. Preliminary results from Project A have been consistent with the assumption that different KSAOs are required for different MOS (Campbell, 1986; Wise, Campbell, & Peterson, 1987).

Project A Predictor Domain

Peterson, Hough, Dunnette, Rosse, Houston, and Toquam (1987) described the Army's practice of almost exclusively hiring inexperienced and untrained individuals as one that creates certain unique demands on the development of a comprehensive selection test battery. The implications of these policies are that "...a highly varied set of individual differences variables must be put into use if there is to be a reasonable chance of improving the present level of accuracy of predicting training performance, job performance, and attrition/retention...," and that "...new predictor measures must be appropriate for selecting persons who do not have the training and experience to begin immediately performing their assigned jobs" (p. 1). These considerations led to adopting a construct-oriented approach to predictor development with considerable emphasis on content. The Project A predictor space consists of six predictor domains: (a) general cognitive ability, (b) spatial ability, (c) perceptual-psychomotor ability, (d) temperament/personality, (e) vocational interest, and (f) job reward preference. There are 24 predictor composite scores within these six domains that summarize 75 individual test or scale scores.

Project A Performance Factors

Wise, Campbell, McHenry, and Hanser (1986; Campbell, Hanser & Wise 1986) describe an empirical model of the Project A criterion space utilizing data from hands-on job sample tests, multiple choice job knowledge tests, Army-wide and job specific Behaviorally Anchored Rating Scales (BARS), and administrative archival data (e.g., awards, reprimands). The model consists of five factors:

- Core Technical Proficiency, the soldier's proficiency on tasks central to his or her particular MOS

- General Soldiering Proficiency, the soldier's proficiency on common tasks required of all soldiers

- Effort and Leadership, the degree to which the soldier exerts effort over the full range of job tasks and provides leadership to peers

- Maintaining Personal Discipline, the degree to which the soldier adheres to Army regulations, demonstrates integrity, etc., and

- Physical Fitness and Military Bearing, the degree to which the soldier stays in good physical condition and maintains appropriate military appearance and bearing.

The first two factors have been called "can do" performance components because the measures composing these factors (e.g., hands-on and job knowledge tests) can be thought of as assessing the best performance of which the examinee is capable ("maximal" performance). The last three factors, which are assessed via performance ratings and administrative measures, have been deemed "will do" performance components. These measures provide an indication of an employee's "typical" performance over time and hence provide a measure of work effort as well as ability.

Project A Validity Results

Campbell (1986) and McHenry, Hough, Toquam, Hanson, and Ashworth (1987) have conducted extensive validity analyses that describe the relationships among the Project A predictor and criterion domains. With the exception of the ASVAB scores, all data on which these analyses were based were collected as part of a large-scale concurrent validation effort. The results of these analyses support two conclusions. First, the ASVAB is extremely useful in predicting "can do" job performance. Second, the Army can improve the prediction of job performance (particularly the "will do" factors) by adding noncognitive predictors to its battery of cognitive predictor tests.

The average validity of the ASVAB for predicting Core Technical and General Soldiering Proficiency across nine MOS was .63 and .65, respectively. As McHenry et al. (1987) indicate, "These validity coefficients [are] extraordinarily high, especially when one considers that the ASVAB was administered to the subjects on average two years prior to the collection of job performance data" (pp. 13-14). Little incremental validity for predicting "can do" performance was obtained by adding any of the new predictors.

However, the incremental validities gained by the addition of the new Project A noncognitive predictor measures were substantial for predicting the "will do" performance criteria. For these three performance criteria--Effort and Leadership, Personal Discipline, and Physical Fitness and Military Bearing--the incremental validities from the full set of new predictors were .13, .21, and .22, respectively, over the ASVAB alone. From Project A, the Army has achieved a greater understanding of the relationships between the predictor and criterion domains which will enable the Army to improve its selection and classification system.

The Army Synthetic Validation Project

Although Project A has been described as the "most significant effort in the measurement and interpretation of human differences yet undertaken," where every "major issue in the science and practice of making personnel decisions is being addressed" (Hakel, 1986, p. 373), additional work remains. The Project A research includes a sample of 21 MOS, carefully selected to be representative of the entire population of Army MOS. Ultimately, however, the Army must develop selection measures for 250-plus entry-level MOS. Further, new MOS are continually being created and selection procedures will be needed for these jobs as well. Thus a second large-scale research effort, called The Army Synthetic Validation Project, was undertaken. The purposes of this project are:

- to evaluate the application of synthetic validation procedures for identifying appropriate composites of selection tests for all 250-plus entry-level Army enlisted MOS

- to develop procedures for setting selection test standards that are linked to standards for job performance (Wise, Campbell & Arabian, 1987).

We have surveyed the literature relevant to the first goal of the project in order to facilitate the synthetic validation effort. The results are presented in the current review which focuses on three major conceptual issues in developing a synthetic validation methodology: (a) evaluating different synthetic validation techniques with special emphasis on how they link job components to predictors; (b) constructing a job component model(s); and (c) reviewing the issues in using expert judges to provide both job component criticality and predictor-job component validity judgments. Literature related to the second objective of the project, the development of standard-setting procedures, also is being surveyed and is summarized in a companion literature review, A

Review of Procedures for Setting Job Performance Standards (Pulakos, Wise, Arabian, Heon & Delaplane, in review).

The Structure of the Review and Specific Army Needs

Comprehensive reviews of both synthetic validation research (Mossholder & Arvey, 1984) and job component models (Fleishman & Quaintance, 1984) have recently been conducted. A retracing of those lines was not thought necessary. Thus, in this review we evaluate major research in terms of its application to the Army's needs. Below, we present a brief summary of the structure and major sections of the review.

Synthetic Validation Research

We begin the second chapter with a discussion of the definition of synthetic validity. We then present a conceptual model of the synthetic validation paradigm, and outline the four stages required to identify predictors and job components, and to establish relationships between them (for sets of jobs). Next, we present general synthetic validation methodologies that vary in the degree to which the described linkages are established. We then describe specific research studies, and evaluate them in terms of the four linkages in the synthetic validation model. Since there are other comprehensive reviews of the synthetic validation literature (e.g., Mossholder & Arvey, 1984; Trattner, 1982), we have not attempted to describe all synthetic validation studies in this section. Rather, we have selected a subset of the studies that are either exemplary or are of special interest for the Army's purposes.

Later in Chapter II we discuss the legal standards for establishing synthetic validity and describe how the various synthetic validation approaches and techniques conform to principles stated in the *Uniform Guidelines on Employee Selection Procedures* (1978). In the final section of Chapter II, we summarize the literature on synthetic validation and draw some conclusions relevant to the Army's needs.

Job Component Models

We begin the third chapter by defining what we mean by job component models. Next, we describe four conceptual approaches to classifying job component models. These are a task description approach, a behavior description approach, a behavior requirements approach, and an ability requirements approach. We outline six criteria against which to evaluate various job component models. These criteria are: (a) reliability, (b) validity, (c) comprehensiveness, (d) acceptance/ease of use, (e) ability to provide linkages between job components and Project A predictors, and (f) conformity to the Army structure of job performance. In the final section of the chapter, we describe and evaluate specific models in terms of their applicability to the Army.

Utilizing Expert Judges

The fourth chapter is divided into two main sections. The first section is a review of literature addressing issues related to use of expert judges to provide job component criticality judgments. These issues include: (a) effect of judge type on reliability and validity of judgments; (b) the relationship between individual difference variables and judgment outcomes; (c) the effect of training and/or rating information on judgment outcomes; and (d) calculation of reliability and validity estimates. The second section of the chapter addresses expert judgment of test validity. Salient issues include the accuracy of the judged test validities and judge qualifications.

Summary and Future Directions

We provide a general summary of the literature review in the fifth and final chapter. We also provide recommendations for how the Army might proceed to investigate the viability of a synthetic validation approach.

SYNTHETIC VALIDATION RESEARCH

Definition

Synthetic validity had its formal beginnings over thirty-five years ago in a paper presented by Lawshe (1952) at a symposium on industrial psychology and small business. Lawshe introduced synthetic validity as a means for establishing the validity of a battery of predictor tests when empirical criterion-related validity studies of predictor measures are impractical. Lawshe broadly defined synthetic validity as "the inferring of validity in a specific situation" (1952, p. 32). Balma (1959) subsequently described three distinct stages of the synthetic validation process: (a) logical analysis of jobs into their elements, (b) determination of test validity for these elements, and (c) combination of elemental validities into a whole.

Guion (1965a) has pointed out that "synthetic validity" is actually a misnomer because validity cannot be created. However, the term is a compact way of describing "synthesis of a valid test battery." The term synthetic validation then is the "logical process of inferring test-battery validity from predetermined validities of the test for basic work components" (Mossholder & Arvey, 1984, p. 323).

Linkages

Guion (1976) described three steps in the synthetic validation process: (a) identify job components that are common across a range of jobs, (b) identify test correlates of job components, and (c) combine the correlates of critical job components into a test battery for that job. Peterson et al. (1982) pointed out that Guion also stated that the evaluation of a synthetic model is ultimately based on performance on a composite criterion, thus implying an additional step. This step involves the construction of a single score from the battery of predictors, that is, a predicted performance score.

Synthetic validation thus requires four steps, three of which require data to establish linkages. The four steps are shown in Figure 1. Synthetic validation begins with identification of a set of job elements (or components) for describing a range of jobs (step 1). The components should cover all important aspects of performance in all jobs under consideration. The second step requires linking each job component to one or more predictor measures (linkage 1 in Figure 1). In the third step, job components that are critical for a job are identified (linkage 2). The "synthesis" occurs in the fourth step. Predictors of critical job components are combined into a test battery for a job or a family of jobs (linkage 3). A specific predictor test composite is defined for each job by using job component criticality weights to combine the prediction equations for the job components judged relevant to the job.

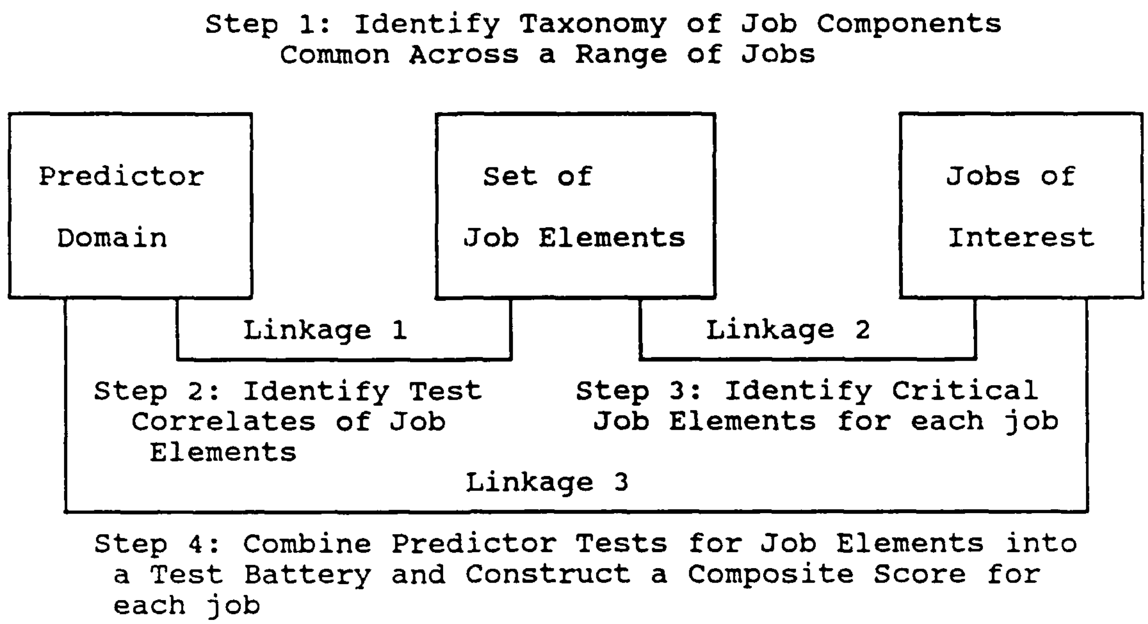

Figure 1. Synthetic Validation Steps and Linkages.

Evaluation of Previous Synthetic Validation Research

There are a limited number of synthetic validation applications reported in the literature. Mossholder and Arvey (1984) and Trattner (1982) have provided comprehensive reviews of these applications. The present review will cover some of the same investigations, but with an emphasis on the more recent applications of synthetic validation and developments in judgment research that have important implications for the linkages in the synthetic validation model.

Reviewers have used several different frameworks to categorize synthetic validation research for discussion and evaluation. For example, Guion (1976) categorized the research according to the strategy used to validate predictors against performance. When a criterion-related validation strategy was used, Guion termed this "direct approach," whereas the J-coefficient (e.g., Primoff, 1972) was called a content valid approach. Mossholder and Arvey (1984) categorized each investigation according to the level of research effort. Thus, the more comprehensive, programmatic approaches, such as the J-coefficient (e.g., Primoff, 1972) and the Position Analysis Questionnaire (McCormick, Jeanneret & Mecham, 1972), received greater attention in that review.

In our review, we have evaluated research investigations on the basis of how completely they address the linkages in the synthetic validation model (Figure 1). We have focused our review on how tne three linkages described above are established and supported.

J-Coefficient Research

Primoff (1955, 1957) introduced the J-coefficient approach to synthetic validation. The J-coefficient is a surrogate index for the test-job performance relationship (typically expressed as a Pearson product-moment correlation).

The first step in computing the J-coefficient is to identify a set of job elements that are common to both the predictor tests and job performance. These job elements may be knowledges, skills, abilities, traits, interests, or any other individual difference variable (Primoff, 1957). Next, the importance of each element for overall job performance is determined and importance weights are assigned to each element. Overall job performance is defined as a weighted composite of the job elements and then estimates of the correlations among the job elements are used to compute the correlations between the overall performance composite and each individual job element (performance-element correlations, R_{ye}). Third, standardized regression weights are obtained for predicting each element from the predictor tests (test-element weights, B_{xe}). Finally, the J-coefficient is computed by summing the cross products of the test-element weights and the performance-element correlations:

$$J_{xy} = \text{sum}(B_{xe}R_{ye}).$$

We present alternative formulas in Appendix A.

Urry (1978) extended the application of the basic J-coefficient formula. He presented derivations of the J-coefficient formula for synthetic validation when tests are equally weighted and when tests are optimally weighted. Trattner (1982) has shown that the J-coefficient formula is identical to the formula for the correlation of a test score with the weighted sum of standardized performance scores.

Other research has empirically supported the agreement between J-coefficients and validity coefficients. For example, Dickinson and Wijting (1976) cited literature showing: (a) a correlation of .56 between J-coefficients and validities of concurrent validation studies for 17 journeymen occupations; (b) J-coefficient of .48 for industrial planners, as compared to empirical validity coefficient of .51; and (c) J-coefficient of .47 for insurance clerks, as compared to validity coefficient of .51. They also cited Air Force policy-capturing research that found validity coefficients underestimated by J-coefficients by an average of only 2%.

The J-coefficient approach for synthetic validation clearly incorporates all three linkages discussed earlier. The job elements, established from job analysis, are linked to predictor measures (B_{xe}) as well as to job performance measures (R_{ye}). Thus, B_{xe} and R_{ye} establish linkages 1 and 2 in Figure 1. The J-coefficient formula combines criterion tests of critical elements to yield a composite performance score and thus satisfies linkage 3.

A key issue for the J-coefficient approach is the source of data used to support each of the linkages. Although component relations (B_{xe} or R_{ye}) for J-coefficient can be obtained empirically by a series of validation studies (e.g., Primoff 1955, 1959), judgmental methods have been shown to yield comparable values. In a recent study, Hamilton and Dickinson (1987) investigated several different sources for estimating component relations, including supervisor, incumbent, co-worker, and test expert ratings. They used policy-capturing and other statistical techniques to derive performance-element and test-element weights from these ratings. The different rating sources and statistical techniques for estimating correlations and weights allowed Hamilton and Dickinson to estimate the J-coefficient in 17 different ways. These 17 J-coefficients were compared to validity coefficients obtained in an empirical validity study.

In general, the obtained J-coefficients correlated highly with the validity coefficients (Hamilton & Dickinson, 1987). Supervisors and incumbents were superior to test experts in providing test-element correlation estimates. However, the finding requires some qualification. First, most of the test experts were graduate students who did not have a great deal of testing experience. These results were similar to those obtained by Hirsh, Schmidt, and Hunter (1986) who also found less accurate judgments from psychologists with lesser experience. Second, in contrast to the test experts, supervisors and incumbents did not make direct estimates of the test-element correlation. They provided performance ratings on elements; then these ratings were correlated with test scores from personnel records. In the estimation of the element-performance relationship, policy-capturing, co-worker, and importance ratings yielded accurate J-coefficients, especially when the test-element estimates also were based on the same source (Hamilton & Dickinson, 1987).

Mossholder and Arvey (1984) cited three major issues with the J-coefficient approach: (a) the esoteric development of the methodology, (b) the complex mathematical basis for its use, and (c) measurement issues. The first two issues are more cautionary than substantive. The measurement issues warrant further discussion since neither obscurity nor complexity implies a flawed approach.

The J-coefficient approach raises a number of measurement issues. The first concerns appropriate procedures for obtaining stable test-element correlations. Often, it is infeasible to obtain empirical correlations for every test-element combination. Moreover, small sample sizes may result in very unstable estimates of true correlations. However, procedures for cumulating research results or validities (e.g., meta-analysis) and for obtaining expert judgment of test validities have yielded more stable results. A second concern is the accuracy of the J-coefficient. Although research has shown that under certain conditions the J-coefficient is equal to a validity coefficient (i.e., a Pearson R between a test battery and job success), the equality will hold up only when all critical job elements are included in the J-coefficient. Further, for J-coefficients to generalize across jobs, the intercorrelations among the elements and the test-element correlations must be consistent across jobs (Trattner, 1982). Both of these conditions can be empirically tested prior to the computation of the J-coefficient and will reveal the extent to which synthetic validation is feasible. In addition, there are other assumptions, such as linearity, that also must be satisfied to ensure that the J-coefficient is an accurate index of validity. Given the paucity of data, it is difficult to assess how the J-coefficient is affected by the violation of any of these assumptions. However, the research cited above (Dickinson & Wijting, 1976; Hamilton & Dickinson, 1987) suggests that J-coefficient procedures can yield predicted performance scores that are highly correlated with empirically-derived predicted performance scores, and can result in accurate estimates of empirical test validity.

The Job Effectiveness Prediction System

The Job Effectiveness Prediction System (JEPS) was a major, four-year research program that applied the synthetic validation approach to develop prediction equations for 8,000 clerical positions in 100 insurance companies (Peterson, Rosse, & Houston 1982). The three linkages in the synthetic validation model are described in the three major steps of the study.

Identify critical job elements (linkage 2). A Job Activity and Performance Description Questionnaire (JAPDQ) was used to gather information about jobs. The JAPDQ contains descriptors for 46 task performance elements that can be grouped into (a) responsibility, (b) extra effort, (c) thoroughness and quality of work, (d) acceptance of criticism, supervision, and change, (e) efficiency and organization, (f) job coaching. The 8000 incumbents rated each of the JAPDQ elements on importance and time spent. Job analysts used the ratings to identify elements that were critical for success in each position.

Identify predictors correlated with performance on each element (linkage 1). Each of the 8000 incumbents completed a battery of predictor tests. In addition, the researchers obtained supervisory performance ratings for each performance area. They used these data to form a matrix of empirical predictors (12) by performance areas (46). In addition, 14 testing experts estimated the correlation of each of the 12 tests with each of the 46 performance areas. The means of the empirical and the judged correlations were used as weights for the tests in predicting performance in the 46 areas. The linkage between predictors and performance areas was thus established using both judgmental and empirical data.

Using these correlation weights, prediction equations were formed to produce a predicted performance score for each of the 46 areas for each incumbent. Correlations of predicted scores with supervisors' ratings of actual job performance yielded a measure of validity of the weighting system based on empirical and judgmental data. In addition, the predicted performance scores derived via synthetic validation procedures were compared with predicted performance scores based on optimal regression weights obtained from regression of effectiveness ratings on the predictors. Validities resulting from the two procedures differed by no more than .01.

Formation of a composite prediction equation for overall job performance (linkage 3). The overall performance composite for each job was formed by weighting the predictor composites for each of the 46 performance areas according to criticality ratings for each performance area. Test batteries were reduced to as few tests as possible without compromising validity. The predictor composite score was subsequently renormed to take account of job complexity. Job complexity was defined as the degree to which a job requires high levels of ability, skill, education, and specific training. Three test batteries were developed. A general screening battery was used when an applicant was hired into a "pool" rather than for a specific job. A second battery measuring numerical ability, reading comprehension, and comparing and checking was developed for clerical jobs. The final battery, including numerical ability, reading comprehension, and coding and converting tests, was developed for technical and professional jobs. Correlations between the batteries and performance criteria ranged from .23 to .26.

When corrected for criterion unreliability, the validities (Pearson r's) for the three batteries were slightly over .30. Therefore, the validity results showed that a synthetically formed predictor composite yielded useful prediction of job success. Validity estimates from the synthetic validation system were almost identical to those obtained from a system based on multiple regression equations formed by regressing a composite overall performance criterion on all the predictor tests. It must be noted that the Peterson et al. (1982) synthetic model produced this level of validity without using a technique that capitalized on sample characteristics. Further, the technique included a separate prediction equation for each job that was based directly on job analysis information.

The results of the Peterson et al. (1982) JEPS study are encouraging. The study demonstrated the feasibility of applying synthetic validation to a wide range of jobs across many organizations. The study also showed the utility of using both judgmental and empirical data to establish the four linkages required by the synthetic validation model. The study also demonstrated that judgmental data are highly accurate surrogates for empirical data.

Guion

Guion (1965a) examined the utility of synthetic validation for use in a small business where empirical validation was particularly difficult. As a result of job analyses of four classes of jobs (i.e., clerical, managerial, sales, and warehouse), he developed a list of seven job elements: (a) salesmanship, (b) creative business judgment, (c) routine judgment, (d) customer relations, (e) leadership, (f) detail work, and (g) work organization.

The president and vice president of the small company rated the relevance of each job element for each position in the company. They also ranked the performance of all employees on each of the seven elements plus overall performance. Guion gave the 48 employees of the company a battery of predictor tests that yielded 19 separate scores. He correlated the predictor test scores with the seven job element criteria. He selected the two tests yielding the highest multiple correlation with each job element and generated a prediction equation for each job element.

The new selection measures were tried out on 13 new employees. They were given the new test battery, and expectancy scores were computed for each job element rated as relevant to their job. The expectancy scores were converted to integers and summed to produce an overall performance expectancy. Guion compared these scores to predicted performance scores obtained from traditional test validation procedures. To obtain the "traditional" predicted performance scores, multiple regression was used to obtain an optimum prediction equation using the predictor data and overall performance rankings from the 48 subjects in the original sample. The regression equation then was used to compute predicted performance scores for the 13 new employees. The salient question was whether the prediction of overall effectiveness based on the seven job elements would compare favorably with the traditional type of prediction. For both the synthetic and traditional validation methods, each new employee was classified as either "predicted superior" or "other." The predictions were compared with actual performance ratings. The traditional model correctly classified 46 percent of the 13 new employees, while the synthetic validation approach correctly classified 76 percent. Owing to the small sample size, this difference was not statistically significant.

Through this application, Guion (1965a) showed all four steps for synthetic validation and supporting data for the linkages. Jobs were analyzed into elements, test correlates of the elements were identified and combined into test batteries, and synthetic validity was compared with empirical validity. Although the small sample limited generality of the results, the research application raised the exciting possibility of generalizing test validity for similar work components via synthetic validation.

Lawshe & Steinberg

Lawshe and Steinberg (1955) described a procedure for identifying the critical ability requirements for clerical positions and providing guidelines for "weighting" test scores. Their goal was to assemble subtests for selecting candidates for clerical positions. Their method assumes that individuals with high scores on an aptitude will be the best performers in positions with high requirements for that aptitude.

The researchers use the Purdue Clerical Adaptability Test (PCAT) to measure aptitude. The PCAT contains seven subtests: (a) spelling, (b) arithmetic computation, (c) vocabulary, (d) arithmetic reasoning, (e) memory for instruction, (f) checking, and (g) copying.

The researchers defined aptitude requirments in terms of a 139-item clerical operation checklist. The checklist served as the source of job components requiring different types of aptitude. The researchers performance a content mapping exercise of requirements and aptitudes. They judged which of the seven aptitudes related to each of the 139 job components.

The researchers hypothesized that there would be a positive relationship between aptitude level and number of related aptitude requirements of a job. That is, it was expected that individuals with a higher aptitude score on spelling would be working in a job composed of relatively more spelling operations. To test this hypothesis, the researchers studied 262 clerical positions in 12 companies. Using the 139-item checklist, supervisors designated the five most important operations in each position that differentiated it from other positions. Job incumbents then took the PCAT and their subtest scores were compared with the number of critical aptitude requriements falling in the same performance area (e.g., spelling, vocabulary, etc.).

For spelling, arithmetic computation, vocabulary, and arithmetic reasoning, results confirmed that the number of operations judged to be critical in a performance area was associated with subtest performance in the same area. (Non-significant relationships were obtained between aptitude and aptitude requirements for the remaining three subtests.) These results were transformed into expectancy tables showing an individual's chances of obtaining a median aptitude score or better for up to five critical operations within the same performance area. Operationally, the selection decision would be based on (a) identifying the five most critical operations for a job, (b) determining the number of operations that are measured by each PCAT subtest, and (c) considering subtest scores only when those critical operations are related to high scores.

With the exception of linking performance on the test battery to job performance, the study incorporated the other linkages required in synthetic validation. Critical clerical operations were identified, tests were associated with critical operations, and a procedure was prescribed for weighting test scores. The study has merit for demonstrating the conceptual foundation of synthetic validation.

McCormick, Jeanneret, and Mecham

McCormick, Jeanneret, and Mecham (1972) argued that if the same work activity is performed in two or more jobs, then the behavioral requirements of that activity must be the same across all of those jobs. They proposed using the 187 job dimensions measured by the Position Analysis Questionnaire (PAQ) for gauging the similarity of work activities across jobs and for identifying an appropriate test battery for each job.

They selected 179 positions from a sample of 536 jobs for which PAQ dimension scores were available. These particular positions were selected because they corresponded to the job content of 90 other jobs for which General Aptitude Test Battery (GATB) scores were available. In other words, the PAQ and GATB information was obtained from separate sources.

Two criteria were used to establish the aptitude requirements for the 90 jobs. One was mean GATB subtest scores and the other was validity coefficients of the GATB for job performance. In separate regression analyses, the PAQ dimension scores were used to predict mean subtest scores on the GATB and validity coefficients. (To reduce spurious results, PAQ dimensions were limited to the 10 which were most highly correlated with the criterion.) With mean GATB subtest scores as the criterion variable, McCormick et al. (1972) obtained a median multiple correlation of .71 with mean PAQ job dimension scores. With validity coefficients as the criterion variable, the median multiple-R was .47. These results showed that the PAQ dimension profiles differed across jobs. Thus, McCormick et al. concluded that a structured job analysis method, such as the PAQ, may be used to synthetically determine the aptitude requirements for a job.

This approach to synthetic validation raises a number of issues. First, the approach equates jobs with individuals. There easily could be a correlation between means across jobs (i.e., mean PAQ requirement score and mean test score) but no correlation of predictors with performance within jobs. Second, the choice of the research criteria has been questioned. McCormick et al. (1972) themselves acknowledged that test score means are only a gross measure of job success. Others, such as Trattner (1982) and Mossholder and Arvey (1984), agreed that occupational membership is at best a measure of minimal requirements for job performance. Therefore, the PAQ approach may not withstand a legal scrutiny under the Guidelines (Trattner, 1982). Third, key linkages in the synthetic validation model are absent. The data show that across jobs, there is a correspondence between (a) standing on tests and PAQ dimension scores, and (b) validity coefficients and PAQ dimension scores. However, the approach fails to show how tests are to be linked to the job requirements and how job requirements are related to performance.

Fine

Fine's (1963) work with Functional Job Analyses (FJA) illustrates another approach to synthetic validation. The FJA approach allows a task to be analyzed in terms of what workers do. An analyst evaluated a job on each of seven scales: (a) data, (b) people, (c) things, (d) worker instructions, (e) reasoning, (f) mathematics, and (g) language.

The design of synthetic validation research paralleled that of McCormick et al. (1972). The goal was to show that aptitude requirements across jobs can be determined judgmentally by job analysts and empirically by tests. Two trained specialists analyzed 85 jobs in functional terms and estimated the relative amounts and importance of the worker functions as requirements for successful performance in each job. GATB scores from employees in the 85 jobs were used to predict the aptitude requirements for the jobs.

For 63 of the 85 jobs, the analysts' estimation of significant aptitude levels agreed with that obtained from GATB scores. The results indicate that minimum job requirements can be estimated judgmentally by job analysts as accurately as they can be estimated empirically by tests. However, as a test of synthetic validation, the study failed to link aptitude requirements with job success. Also relevant is Mossholder and Arvey's (1984) criticism against using occupational membership as a criterion.

Legal Standards for Establishing Synthetic Validity

Trattner (1982) evaluated the conformity of synthetic validation to the Uniform Guidelines (1978) validation requirements. According to Trattner, the logical and empirical foundations of synthetic validation meet the Guidelines requirements for test validation. Synthetic validation (as defined by Balma, 1959; Guion, 1976) is predicated on analyzing a job into job components, identifying predictors of the components, and establishing links between job components and job success. All these activities embrace the concepts of criterion-related validity and construct validity as they are defined in the Guidelines:

Criterion-related validity. Demonstrated by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior. (Sec. 16F, p. 38307)

...However, if a study pertains to a number of jobs having common critical or important work behaviors at a comparable level of complexity, and the evidence satisfies subparagraphs 14B (2) and (3) above for those jobs with criterion-related validity evidence for those jobs, the selection procedure may be used for all the jobs to which the study pertains. (Sec. 14D(4), p. 38303)

Procedures for determining the relative importance or contribution of job components for a job and the reliance on job analytic information for selecting tests are likely to be defensible on the basis of content validity (Kirkland v. New York Department of Correctional Service, 1974). Therefore, the J-coefficient approach, and the research conducted by Peterson et al. (1982), and perhaps the approach by Guion (1965a) are defensible procedures. On the other hand, the synthetic validation approaches reported by McCormick et al. (1972) utilizing the PAQ, by Fine (1963) using FJA, and by Lawshe and Steinberg (1955) are more vulnerable. These studies used procedures that may be defended as content valid. However, the studies lacked data supporting direct relationships with job success.

Although it appears that the concept of synthetic validation does satisfy key requirements in the Guidelines concerning job analysis and test validation, synthetic validation has not been directly tested in court cases. There has been one related District level court decision that recognized the potential for generalizing test validity across situations. In a case challenging the use of the General Aptitude Test Battery (Pegues v. Mississippi State Employment Service, 1980), the court wrote:

Empirical research has demonstrated that validity is not perceptibly changed by differences in location, differences in specific job duties or applicant populations. Valid tests do not become invalid when these circumstances change. Plaintiffs' allegation that validity is specific to a particular location, a particular set of tasks and to a specific applicant population, or in other words, that a valid test in one set of circumstances is not valid in circumstances not perfectly identical is not true. (488 F. Supp. 239 p. 254)

The GATB is also the object of a study directed by the National Academy of Sciences. The appointed committee will address, among other issues, the issue of using validity generalization evidence to support use of a test for selection. The extent and conditions of validity generalization will likely have some bearing on the use of synthetic validation approaches that depend on the generality of job component predictions across jobs.

Summary and Conclusions of Synthetic Validation Research

The Army has decided to investigate the feasibility of synthetic validation to extend test validity for a limited number of jobs to the entire range of Army enlisted jobs. Toward this end, we have reviewed several synthetic validation applications. Although few comprehensive studies have been conducted, the model can be--and has been--implemented such that the results are both technically sound and legally defensible. Based on our review, we conclude that sound techniques exist to investigate synthetic validation.

Early, small-scale studies by Guion (1965a) and Lawshe and Steinberg (1955), spelled out the conceptual foundations of synthetic validation. More recently, major job analysis methods or instruments (e.g., PAQ, FJA, and J-coefficient) also have served as the basis for synthetic validation. Of the latter three, only the J-coefficient approach offers a practical, defensible technique for satisfying the four steps required by the synthetic validation model. The synthetic validation studies involving the PAQ and the FJA would be tenable by adding the linkage between job components and job success. Peterson et al. (1982) also have demonstrated a successful large-scale application of synthetic validation procedures.

Successful synthetic validation hinges on establishing support for all three linkages. Further, evidence for the linkages should be established on the basis of sound research design. Such designs should address and adhere to the principles of content, criterion, and construct validation.

The first step requires a comprehensive job component taxonomy to describe work. Successful synthetic validation also requires the proper specification of linkages of job components with predictors and with job performance. Empirical validation of the linkages is desirable. The predictor-job component linkage can be supported via criterion-related validation. The wealth of data available from Project A will provide validity information where appropriate. Multiple predictor measures of a job component are amenable to construct validation. Theoretically, a successful synthetic validation provides not only convergent validity of multiple predictor measures, but also generality of those measures across jobs.

Data supporting the linkages can be obtained by judgmental methods. Expert judgment will assume an increasing role in the synthetic validation process. Research has shown that expert judgments have yielded accurate results (Hamilton & Dickinson, 1987; Hirsh et al. 1986). The Guidelines have endorsed expert judgment for demonstrating content validity of a testing procedure as well as in determining the critical requirements of a job. Experts' estimates may be obtained to complement empirical validities. In addition, judgments of job success will likely be needed to fully demonstrate synthetic validity.

In the next chapter, we address specific linkages in the model presented in Figure 1. We present alternative types of job descriptors and evaluate various job component models to guide development of a job component model for the Army Synthetic Validation Project (step 1). In a subsequent chapter, we present evidence supporting the use of judgmental data to establish linkages 1 and 2.

JOB COMPONENT MODELS

The first step in a synthetic validation exercise is the specification of a set of common job elements or job components that can be used to describe a broad range of jobs. This chapter describes different job component models or descriptive approaches that have been used to specify jobs in terms of their constituent components or elements. The chapter is divided into three major sections. First, we define what we mean by a job component model and outline major differences among various models. Although there are a variety of job component models as well as a number of dimensions on which they can vary (e.g., generality, structure, intended applications, what gets described--e.g., positions, occupations, tasks), we concentrate on the four major variations. These four different kinds of models describe jobs in terms of task characteristics, job behaviors, behavior requirements, and ability requirements. Synthetic validation models almost certainly will use one or more of these approaches to describe the components or elements of jobs. Second, we outline six criteria against which job component models can be evaluated, including specific requirements for an Army job component model. Third and finally, we review and evaluate specific approaches to description in terms of the criteria outlined earlier. To facilitate organization of this discussion, the descriptive approaches are categorized into the four general kinds of job component models mentioned above.

Major Types of Models

A job component model may be conceived of as a set of descriptors having a number of important properties. The descriptors represent elements or components that are important units of jobs in the sense that they have implications for job performance. The elements or components are not job-specific but rather can be used to describe a wide variety of jobs. The elements or components can be linked to the knowledge, skills, aptitudes and other characteristics of personnel who are job incumbents. A job component model often is simply the factors/clusters of behaviors, attributes, or tasks derived from performance data (e.g., the factor analytic work of Fleishman, 1975), or from responses to a structured job analysis instrument such as the Occupational Analysis Inventory (Cunningham, Boese, Neeb, & Pass, 1983) or the Position Analysis Questionnaire (McCormick et al., 1972). Thus, a job component model serves as the basis for describing and differentiating among jobs.

Job component models can vary on a variety of dimensions. A job component model can be developed to meet either of two general objectives; the first is theory development and the second is practical application. When a job component model is intended for theory development, a general descriptive system is derived, often with eventual utility for a number of more specific applications. A potential shortcoming, however, may be that the model is so broadly conceived as to limit differentiation that could be achieved with a more specifically focused model. On the other hand, when specific applications are the driving force behind model development, unique models may result. The potential drawback then is that while the highly focused, relatively unique model may satisfy the purpose for which it was developed, it may be limited in terms of generalizability to other areas.

Another way in which models can vary is that they can be structured either hierarchically or non-hierarchically. An example of a hierarchical model is the Berliner, Angell, and Shearer (1964) model. The system has 47 specific job elements subsumed by six intermediate level elements, which in turn are encompassed by four main elements. Non-hierarchical models contain components of relatively equivalent levels of specificity. The specificity of any given model has implications for its utility; a hierarchical model has potential for different applications within the same job domain.

Different models also can vary in their unit of focus or analysis. Job component models may be based on macro-level data, such that positions, jobs, or occupations are described as whole entities. Alternatively, micro-level data may be used to describe job tasks or other entities subsumed by a job.

Prior reviews have identified four different types of job component models (McCormick, 1979; Fleishman & Quaintance, 1984); we briefly describe each of the four types below. They include task characteristics approaches, behavior description approaches, behavior requirements approaches, and ability requirements approaches. Each of the four different approaches has been used for job description and for investigation of job similarities and differences (Fleishman & Quaintance, 1984; McCormick, 1979). While other types of job-related information can also function as job descriptors, the four commonly used approaches are adequate for most purposes, e.g., job description, categorization, investigation of job similarities and differences, and study of interrelationships. We review each of the four approaches in the following sections.

Task Characteristics Approaches

The work itself, or inherent characteristics of tasks, are the units of interest in the task characteristics approach. Tasks are assumed to impose certain conditions on an individual performing them, such as the goals to be achieved, procedures to be followed, characteristics of the responses elicited from the task performer, and activity content. These task dimensions are distinct from the human responses, processes, and/or abilities that are involved in performance. This approach differs from the other three in that it is not based on either the individual's observable/overt behavior or internal processing. Choice of descriptors and/or a strategy for narrowing the descriptor set are issues to be addressed when using the task characteristics approach. Task characteristics models have been commonly used to investigate the effects of variations in task characteristics (employed as independent variables) on various learning and performance variables (the dependent variables).

Behavior Description Approaches

Observations and descriptions of behaviors exhibited by a worker while working are the basis for the behavior description approaches. Job components are based on observable behaviors, rather than on what individuals should do or might do. Description can vary in level of specificity, e.g., rotates wheel in clockwise direction vs. adjusts. Behavior description approaches are useful for purposes such as documenting task domains for jobs, or cataloging effects of specified environmental variables on specific aspects of performance (Fleishman & Quaintance, 1984). This approach is amenable to use of structured listings of verbs at one or several levels, and thus job analysis terminology can be standardized across job families for descriptive purposes.

Several authors (McCormick, 1959; Prien, 1977; Ash, 1982) have used a variety of terms to define two distinct descriptor bases within behavior description approaches. The first subtype is referred to as "work-oriented" description--where job descriptors describe job content mainly in terms of what is accomplished as the result of work activities. Description is based on observable outcomes or results of tasks/activities done on the job. In contrast, the second subtype, "behavior-oriented" description, describes the activities or the behaviors individuals actually engage in to perform a job.

Behavior Requirements Approaches

Behavior requirements models are also based on behaviors, but rather than emphasizing the actual behaviors performed, the focus is on behaviors that are assumed to be required for performing at a specified criterion level, or behaviors that should occur. In this approach the descriptors essentially are terms describing functions or intervening processes--what activities should occur between an impetus and a further action. Two types of descriptors that are commonly found in this type of model are "decision making" and "problem solving." Examples of applications for these models include procedure development and sequencing of training.

Ability Requirements Approaches

A fourth approach, called the ability requirements approach, is based on the notion that abilities are relatively enduring qualities of a task performer. A basic assumption of this approach is that different types of attributes/abilities exist. Further, it is expected that there are individual differences in levels of these attributes and that these individual differences are related to job-related behaviors and performance. For description and classification purposes, tasks of jobs would be grouped on the basis of similar ability requirements. The nature of the abilities is usually judged or inferred from factor analyses of groups of tests, since the abilities cannot be directly observed.

Multi-Methodological Approaches

Specific descriptive strategies, methods, or job component models used in conducting job analyses yield different types of job information, information that often may prove useful and/or complementary for different purposes. Rather than distinguish among job analysis methods on the basis of their focus or unit of analysis, Ash (1982) advocates categorizing job analysis methods by information obtained: (a) task-based, (b) behavior-based, and (c) attribute-based. Prien (1977) suggested using a multi-methodological approach for selection research, incorporating all three methods for describing jobs. The comprehensive information obtained from such a strategy could then be used for a variety of personnel functions (e.g., employee selection, training, etc.). This suggests the possibility of using a mixed model that would incorporate a combination of descriptor types in order to obtain more comprehensive job descriptive information. Many job component models, including those fitting one of the four main types, contain a mixture of descriptor types (e.g., a number of task or situational characteristics along with a majority of behavioral entities).

Evaluation Criteria

In the previous section, we have defined job component models and identified the four most prevalent types. Irrespective of the intended purpose for developing a job component model and regardless of what type of model may be appropriate, any job component model should meet certain statistical and practical criteria. The specific criteria employed are likely to vary from situation to situation. Various authors (Knerr, Miller, Hill, Nadler, Deppner, Dowell, & Somers, 1982; Fleishman & Quaintance, 1984) have proposed a number of different criteria for evaluating job component models. However, these criteria all fit the more general categories of reliability, validity, and utility.

We have identified six criteria against which to evaluate the potential applicability of various job component models for use by the Army in synthetic validation research. The first four criteria apply to job component model use for job description purposes. The fifth criterion applies to use of model components in establishing linkages with selection tests. The sixth criterion applies to the model of job performance currently being developed by the Army. These criteria, intended to help guide the development of a job component model as well as to evaluate the final product, are: (a) reliability, (b) validity, (c) comprehensiveness, (d) acceptance/ease of use, (e) ability to provide linkages among job components and Project A predictors, and (f) conformity to the Army conceptualization of job performance. The comprehensiveness, acceptance, and linkage criteria can be interpreted as more specific aspects of the broad "utility" category.

Reliability

By "reliability" we mean the consistency with which judges can describe jobs in terms of selected job components or elements. Subject Matter Experts (SMEs) should be able to consistently decompose jobs into elements or components (whether they are task characteristics, behaviors, behavior requirements, or abilities). Thus, descriptors must be defined as precisely and objectively as possible, no matter what level of description is used. The importance of precise descriptor definition is highlighted when using a hierarchical component model, as we expect that the reliability of SMEs' judgments will vary across the different levels. Intra-as well as inter-rater reliability should be assessed.

Validity

Validity refers to the adequacy of the resulting job descriptions in fulfilling the model's purpose. For the Army's purposes, assuming the jobs are defined specifically enough to tie predictors to them, it is desirable that the same predictors predict individual differences in performance on the specified content of each job component.

Comprehensiveness

The third criterion, comprehensiveness, requires that the job components cover all aspects of performance considered important for the domain of jobs under scrutiny. This criterion must be further qualified for the present context. The model should be sufficiently comprehensive to permit use across the range of 250-plus MOS, while at the same time, components must be sufficiently specific to allow SMEs to provide ratings that are as reliable and valid as possible. The Army's focus on the motivational components of a job requires that we judge models missing such motivational elements as incomplete.

Acceptance/Ease of Use in Job Description

Ease of use is the extent to which judgments required by the model are relatively straightforward and easy for raters to make. The degree to which the model is logical and appropriate for the intended purpose(s) also influences acceptability. The appropriateness of a model depends to some extent on its comprehensiveness. Raters will find a model unacceptable and difficult to work with if its components do not provide adequate coverage of all job entities. At the other extreme, raters will have difficulty using a model which is not concise. The model should have mutually exclusive categories, to enhance the reliability with which SMEs use the components to describe jobs. Mutual exclusivity is especially important during initial attempts to break jobs into components.

Ability to Provide Linkages with Project A Predictors

Of major importance in the Army synthetic validation project is the requirement to develop job components that can be linked to Project A predictor constructs. Some models can be linked with these selection tests in a straightforward manner, whereas others (e.g. task characteristics models) require taking additional steps in order to do so. We will categorize models having components that can be directly linked to predictors as having "High" ability to provide linkages with Project A predictor tests. To illustrate, we can directly relate scores on various ability tests to components of models based on ability requirements descriptors. Models that involve more steps or more complex procedures to establish these linkages will be categorized as "Low" on this criterion.

Conformity to the Army Conceptualization of Job Performance

Now that we have outlined the major types of job component models and the criteria considered important for evaluating them, we must discuss an additional requirement necessary for model development in the present context. A job component model to be used for synthetic validation across 250-plus Army MOS requires more than just adherence to the five evaluation criteria described above. An additional specification is that the model and its specific job descriptors should make sense for describing Army jobs. The model should contain components that reflect our current view of the enlisted soldier's performance domain.

Job component models vary in their degree of applicability/ appropriateness for use in the Army. Some specific job components will be relevant for Army jobs while others will not, especially with the Army's previous focus on measurement of first-tour performance. Thus, particular attention will be given to components that are judged useful for describing first-tour performance--jobs of enlisted soldiers who have been in the Army for approximately 18 months subsequent to completing Advanced Individual Training (AIT).

Project A defined a five-factor model of the performance domain. As outlined in Chapter I of this report, a basic division exists within this model between "can do" (i.e., maximal performance indicators) and "will do" (i.e., more typical, motivational performance indicators) components. Given that the five-factor Project A model of the performance space showed a high degree of consistency across the jobs sampled, it is likely that this model can be extended to other existing MOS and to new MOS in the future. We believe that the motivational components of performance are essential to any definition of soldier effectiveness and that they therefore must be included in developing the present job component validity model. Thus, a requirement for a synthetic validation model for the Army's purposes is that the model contain components/elements that are consistent with the Army's current conceptualization of performance (Campbell, McHenry, & Wise, 1987).

Job Component Models

In this section, we describe and evaluate (based on the criteria discussed above) specific job component models in terms of their applicability for use in the Army. To facilitate this discussion, the models have been organized by the conceptual approaches they represent.

It is not our intent to describe all models that have appeared in the literature, as more thorough coverage of these models can be found elsewhere (e.g., see Fleishman & Quaintance, 1984). We concentrate on "major" models, which are those which have received much attention or wide application relative to other models of the same general type. For each major model we describe: (a) the rationale (or anticipated application) for its development, (b) how it was developed, (c) its descriptors, and (d) its applications and/or methods. The models are evaluated in terms of the evaluation criteria outlined above. We also briefly describe a number of "minor" models. These are models that are not well developed or tested but contain relevant components and/or procedures. For example, some of the "minor" models are really just sets of job components developed in single-job studies or in occupation/job grouping studies. We included them in our review if they met one of the following criteria: (a) motivational components were included, (b) some of the components were judged relevant for a subset of MOS, based on previous work with these MOS in Project A, or (c) components were developed for jobs within military settings. The reader may refer to Appendixes B, C, D, and E for brief descriptions of these "minor" models.

Literature Review: Procedure and Format

We used a structured format for recording information from all articles reviewed for this report. In particular, the following information was recorded: (a) whether empirical results, a review of the literature, and/or methods and procedures were given; (b) the type of job component model (e.g., behavior description); (c) whether motivational components were included; (d) any applications of the model in the article (e.g., selection/classification, job evaluation, training, job description/classification); (e) mention of components-job performance linkages; (f) inclusion of evaluation criteria; and (g) whether comparisons were made between models. We recorded any other relevant information in a comment section.

If a job component model was presented, discussed, or applied in the article, reviewers indicated: (a) number of components; (b) list of any motivational components included in the model; (c) list of any "can do" components included in the model; (d) any of the model's components not included in the Project A job performance measures (e.g., physical requirements were not included in Project A); (e) a rating, on a scale of 0% to 100% of the completeness of the model for describing enlisted job performance in the Army (and if less than 100%, a description of the types of components that were missing); and (g) a rating from "very difficult" to "very easy" of how difficult it would be for Army job experts to use the components to describe Army enlisted jobs and/or rate Army enlisted job performance requirements. Where applicable and informative, we provide responses to these structured questions for the specific models discussed in the following sections.

Task Characteristics Models

"Task Taxonomy" from Manpower and Personnel Requirements Determination Methodologies (MANPERS) Project (Knerr et al.). Two models were created as part of a larger Army project to determine manpower requirements for operator and maintenance MOS (Knerr et al., 1982). An MOS structure model (taxonomy) was developed for determining the MOS for a new equipment system. MOS titles were sorted on the basis of lines of equipment, Career Management Fields (CMF), and organizational structure of the Army. A task structure model (taxonomy) was developed to (a) confirm the MOS clustered in the MOS structure model, and (b) generate organized task lists for those MOS, based on behavior requirements.

The task structure model was derived in the following manner. Descriptions of entry level duties for a wide variety of MOS were obtained. The descriptions were sorted into clusters on the basis of similarities in performance requirements and the clusters were sorted into broad function categories. Most of the descriptions within clusters included both an action verb and a direct object designating equipment. Modifiers applied to the verbs were of three types: (a) enabling techniques, (b) constraints, and (c) performance level. Equipment was specified as (a) generic, (b) components, or (c) whole systems. These steps were followed for two broad groupings of MOS (operator and maintainer); the resulting categories were combined into the "Task Structure Taxonomy." Six major sections and 20 subsections are contained in the final taxonomy (shown in Table 1).

The task structure model was presented in an interim report of the MANPERS project. Reliability and validity evidence was not included. The model is both comprehensive and developed on the basis of a subset of Army MOS. It is therefore useful for describing Army enlisted MOS, except for two components concerned with formal supervision, e.g., "Administration of....", and "Functions in other units." The motivational component within this model is "Personal Leadership Roles." This model was given a moderate rating for comprehensiveness (even though it was developed on the basis of Army MOS), specifically because it was missing types of tasks for non-maintenance/non-operations MOS. In addition, motivational components for enlisted MOS (meaning non-supervisory functions, tasks, etc.) were missing. The model is not appropriate for our intended purposes of description across all entry-level MOS, and SMEs could not rate all MOS using these limited components. However, the inclusion of military systems operations and tasks common to all soldiers would make the model more acceptable and logical to Army raters. Expert rater judgments could be used to link components within the MANPERS taxonomy to Project A predictors.

Task Characteristics Approach (Farina and Wheaton). The "Task Characteristics Approach" was developed by Farina and Wheaton (1971) for purposes of task classification. A task is defined as a "set of conditions that elicits performance" (Fleishman & Quaintance, 1984, p. 169). The general components of task characteristics are explicit goal, procedures, input stimuli, responses, and stimulus-response

Table 1

Manpower and Personnel Requirements Determination Methodologies: Project (MANPERS) Task Taxonomy[a]

A. Common Soldiers' Tasks
   1. Construction (e.g., Constructs individual fighting positions)
   2. Other Unit Functions (Performs outpost and security guard duties)
   3. Patrol (Performs proper dismounted movement techniques as part of a dismount team)
   4. Battlefield Survival (Protects self, weapons, and equipment from chemical and other contaminants)

B. Operation of Systems - Primary Functions
   5. Vehicle Operation (Provides a steady platform for stabilized weapons fire)
   6. Preparing System for Operation (weapons or communications) (Prepares range cards)
   7. System Operation (Fires missiles)

C. Operation of Systems - Subordinate Functions
   8. Power Generation
   9. Safety, First Aid, and Fire Prevention (in operating systems)
   10. Computer Operations (when not used with a combat system)

D. Administration, Job Aids, and Constraints
   11. Administrative Tasks
   12. Understanding, Following Job Guidance
   13. Constraints

E. Maintenance
   14. General
   15. Maintenance Actions
   16. Tools and Methods Used on Job

F. Supervisory Functions in Maintenance
   17. Direct Contributions to Unit Maintenance
   18. Administration of Unit Maintenance
   19. Personal Leadership Roles
   20. Functions in Other Units

---

[a]Adapted from Knerr, C. M., Miller, E. E., Hill, G. P., Nadler, L. B., Deppner, F. O., Dowell, S. K., & Somers, R. L. (1982). Interim report for manpower and personnel requirements determination methodologies (MANPERS). McLean, VA: General Research Corporation.

relationships. Nineteen descriptors--called task characteristics--(given in Table 2) were developed for these general components. Tasks were rated in terms of each characteristic on a seven-point rating scale. Each rating scale was anchored at the high, low, and middle points by task examples. The task characteristics approach was evaluated in terms of reliability and validity criteria.

A core of most reliable task characteristics indices was identified across three studies. Interrater reliabilities (Intraclass correlations) ranged from .78 to .98 (with one exception of .58--Amount of Muscular Effort Involved); indices with interrater reliabilities above .90 are indicated in Table 2. The authors recommended averaging three or more raters' ratings to assign task characteristic values.

Fleishman and Quaintance (1984) reviewed this work as a task characteristics approach to job description. However, it is actually a task description system, rather than a job component model. As such, to be useful for describing jobs, all tasks in jobs would have to be described on the set of characteristics that are discussed and listed below. Besides being a cumbersome procedure, decomposing tasks in this manner could obscure what those tasks are.

The degree to which task characteristics predicted learning or performance was one criterion for evaluating the approach. The technique called "postdiction," using performance measures (as criteria) abstracted from studies reported in the literature, was used; tasks used in these studies were rated on the task characteristics. Results of multiple regression analyses, using the six most reliable indices (see Table 2 for these six), indicated that the task characteristics were correlates of performance, and that it was possible to relate differences among tasks to variations in performance (for the two studies, Multiple R's were .82 and .73 after correction for small sample bias). Mean performance scores were used, by averaging across individuals, to minimize effects of individual differences.

The model is very complete as a task-based model. However, because the conceptual basis for the model is the task itself (as for all task-based models), then the model logically lacks motivational components. Army SMEs would have difficulty describing jobs with the task characteristics model; components of the model are too abstract for rating jobs and performance across MOS. Establishing links between these components and Project A predictors--at any of the hierarchical levels of the predictor space--(constructs, clusters, and factors) would be difficult.

Task Characteristics Approach: Minor Models. Additional task characteristics models were reviewed, but were found to be of limited usefulness for our purposes. The reader is referred to Appendix B for brief descriptions of these models and evaluative comments. Several are abstract in nature, one of which was developed specifically for training purposes (Cotterman 1959). Classification of skilled tasks was the focus

Table 2

Farina and Wheaton: Task Characteristics Approach[a]

Task Components		Task Characteristics
Explicit Goal	*	Number of output units
	*	Duration for which an output unit is maintained
	**	Number of elements per output unit
	**	Work load imposed by task goal
		Difficulty of goal attainment
Procedures	*	Number of procedural steps
		Dependency among procedural steps
		Adherence to procedures
		Procedural complexity
Input Stimuli	*	Variability of stimulus location
	*	Stimulus or stimulus-complex duration
		Regularity of stimulus occurrence
Responses	**	Precision
	**	Rate
	*	Simultaneity of responses
		Amount of muscular effort involved
Stimulus-Response Relationship		Degree of operator control
		Reaction time/feedback lag relationship
		Decision making

[a]Adapted from Farina, A. J., & Wheaton, G. R. (1971). Development of a taxonomy of human performance: The task characteristics approach to performance prediction. Washington, DC: American Institutes for Research.

* These six task characteristics indexes were most reliable, and were used in regression analyses (postdiction studies, n=28 judges for 15 tasks).

** These task characteristics indexes had reliabilities above .90, but were not used in postdiction studies.

for two models developed by Fitts (1962). Finally, the Job Diagnostic Survey (JDS) (Hackman & Oldham, 1975) contains scales that assess the degree to which jobs possess five core job characteristics. Because the JDS instrument was developed for use in the context of job redesign, rather than for job description or classification, its components do not apply here.

Behavior Description Models

Hierarchical Models--Behaviors, Activities, and Processes (Berliner, Angell, and Shearer). This three-level hierarchical model focused on performance measurement for military jobs (Berliner, Angell, & Shearer, 1964). At the most basic level are 47 specific behaviors in the form of verbs (e.g., detects, inspects, observes). At progressively higher levels, six activity types and then four behavioral processes encompass the specific behaviors (see Fleishman & Quaintance, 1984). The model, illustrated in Table 3, was developed by collecting judges' (n=8) independent groupings of specific behaviors into process categories; behaviors grouped together by a majority of the judges (6 of the 8) were retained. The hierarchical nature of the model was planned, due to: (a) the finding of high interrater reliability for both extremes--numerous specific categories and few general categories, and (b) the combination of both high intercategory differentiation (for many) and avoidance of sparse, unique categories. Specific behaviors in the scheme can be quantified by measures of (a) times, (b) errors, (c) frequency, (d) workload, and (e) motion dynamics. With the inclusion of instruments for collecting these measurements, a three-dimensional Behaviors X Measures X Instruments matrix can be constructed.

Major advantages of the model include its applicability at different levels of specificity and its simplicity. Problems with this model are (a) no reliability assessment using "real world" tasks, and terms with nebulous definitions; (b) lack of validity evidence; and (c) lack of motivational components.

SMEs would probably have difficulty using this model to describe Army enlisted MOS and to rate job performance requirements. At the specific behaviors level, the level appropriate for making these judgments, the components are verbs rather than the more common nouns or verb-noun statements. However, the list of verbs does seem to be comprehensive for covering a wide variety of actions performed across MOS. An exception is that specific behaviors within the motor processes category may not adequately tap motor components of Army jobs (i.e., core technical area tasks). The components and the judgments required were developed specifically in the context of military jobs. The potential exists to establish links between the specific behaviors and Army predictors.

Table 3

Hierarchical Model -- Behaviors, Activities, and Processes[a]

Processes	Activities	Specific Behaviors
Perceptual Processes	Searching for and receiving information	- Detects
		- Inspects
		- Observes
		- Reads
		- Receives
		- Scans
		- Surveys
	Identifying objects, actions, events	- Discriminates
		- Identifies
		- Locates
Mediational Processes	Information processing	- Categorizes
		- Calculates
		- Codes
		- Computes
		- Interpolates
		- Itemizes
		- Tabulates
		- Translates
	Problem solving and decision making	- Analyzes
		- Calculates
		- Chooses
		- Compares
		- Computes
		- Estimates
		- Plans
Communication Processes		- Advises
		- Answers
		- Communicates
		- Directs
		- Indicates
		- Informs
		- Instructs
		- Requests
		- Transmits

Table 3 (cont.)

Hierarchical Model -- Behaviors, Activities, and Processes[a]

Processes	Activities	Specific Behaviors
Motor Processes	Complex-continuous	- Adjusts - Aligns - Regulates - Synchronizes - Tracks
	Simple-discrete	- Activates - Closes - Connects - Disconnects - Joins - Moves - Presses - Sets

[a]*Reproduced from* Berliner, D. C., Angell, D., & Shearer, J. W. (1964). Behaviors, measures, and instruments for performance evaluation in simulated environments. Paper presented at a symposium and workshop on the quantification of human performance, Albuquerque, New Mexico.

Position Analysis Questionnaire (McCormick, Jeanneret, and Mecham). The Position Analysis Questionnaire (PAQ) is a structured job analysis questionnaire that also serves as a classification system for job behaviors. Development of the PAQ was based on an information-input, mediation process, work-output (S-O-R) model of human behavior. Job elements are tied to either one specific stage of this model, or to one of three aspects of work--interpersonal activities, work situation and job context, and miscellaneous (McCormick, et al., 1972). The elements are not phrased to contain job-specific or occupation-specific activities. Instead, job elements characterize human behaviors in general terms, e.g., work characteristics, activities, or conditions. Thus, descriptors are essentially a mixture of behavior descriptions and task characteristics.

Another set of PAQ job dimensions was also derived, based on attribute profiles of job elements. Job dimensions based on attribute profile data represent an ability requirements job component model; we discuss this model later on in the section on ability requirements models.

Job analysis data from a broad variety of jobs in a range of organizations have been collected with different versions (System I, System II) of the PAQ. Scales used to obtain ratings vary according to content of the element, e.g., Extent of Use, Amount of Time Spent, and Importance. A series of factor analyses has been conducted to determine the structure of work behavior (McCormick et al., 1972; Harvey, Friedman, Hakel, & Cornelius, 1987). Principal components analyses based on all elements yielded a set of 13 overall dimensions of work performance. In addition, specific job components within the six defined divisions of the PAQ (three parts of the model of human behavior--information input, mental processes, and work output--and three aspects of work--relationships with other persons, job context, and other job characteristics) were defined on the basis of principal components analyses. Principal components results for each section of the current widely-used form, System II, are listed in Table 4.

Use of the PAQ instrument under a variety of circumstances with a number of different rater types (e.g., trained job analysts, untrained raters, SMEs) has yielded reliabilities ranging from .59 to .89 (McCormick et al., 1972; Smith & Hakel, 1979; Cornelius, et al., 1984; Friedman & Harvey, 1986; Harvey & Hayes, 1986). In terms of validity criteria, the PAQ has been applied in employee selection and classification and job description and classification. It also has been used in synthetic validation and job evaluation. Gutenberg, Arvey, Osburn, & Jeanneret (1983) presented evidence that two PAQ dimensions, Information processing and Decision making, could identify tests to use/not use for selection into jobs having different characteristics. In a job evaluation study, compensation was compared among different jobs that had behaviorally similar job characteristics (McCormick et al., 19,2). Although the job titles differed, the similar job characteristics required similar attributes of incumbents. Roughly equivalent wage/salary rates were found for different jobs with behaviorally similar job characteristics.

The model was given a high rating for completeness for describing job performance (90%). The model's comprehensiveness comes as no surprise, as the instrument was developed in order to describe a broad range of jobs. Other than the lack of motivational components, the model could be used for describing Army enlisted MOS. However, Army SMEs would have difficulty rating performance across MOS and describing jobs using PAQ dimensions, as the instrument is aimed at a post-college reading level and is intended for use by trained job analysts. Linkages with Project A predictors could be established using PAQ dimensions.

Table 4

Position Analysis Questionnaire: Divisions and Job Dimensions[a]

Division 1: Information Input

1. Perceptual interpretation
2. Input from representational sources
3. Visual input from devices or materials
4. Evaluating - judging sensory input
5. Environmental awareness
6. Use of various senses

Division 2: Mental Processes

7. Decision making
8. Information processing

Division 3: Work Output

9. Using machines, tools, or equipment
10. General body vs. sedentary
11. Control and related physical coordination
12. Skilled or technical activities
13. Controlled manual or related activities
14. Use of miscellaneous equipment or devices
15. Handling, manipulating, and related activities
16. Physical coordination

Division 4: Relationships with other persons

17. Interchange of judgmental and related information
18. General personal contact
19. Supervisory, coordination, and related activities
20. Job-related communications
21. Public-related personal contacts

Division 5: Job context

22. Potentially stressful or unpleasant environment
23. Personally demanding situations
24. Potentially hazardous job situations

Division 6: Other job characteristics

25. Nontypical vs. typical day work schedule
26. Businesslike situations
27. Optional vs. specified apparel
28. Variable vs. salary compensation
29. Regular vs. irregular work schedule
30. Job demanding responsibilities
31. Structured vs. unstructured job activities
32. Vigilant or discriminating work activities

Table 4 (cont.)

Position Analysis Questionnaire: Divisions and Job Dimensions[a]

Overall Dimensions	
33.	Decision, communication, and general responsibilities
34.	Machine and equipment operation
35.	Clerical and related activities
36.	Technical and related activities
37.	Service and related activities
38.	Regular day schedule vs. other work schedules
39.	Routine and repetitive work activities
40.	Environmental awareness
41.	General physical activities
42.	Supervising or coordinating other personnel
43.	Public, customer, and related contact activities
44.	Unpleasant, hazardous, or demanding environment
45.	(Unnamed)

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

Job Element Inventory (Cornelius and Hakel). The Job Element Inventory (JEI) was developed (Cornelius & Hakel, 1978) to overcome the major limitation of the PAQ -- its reliance on job analysts with high verbal ability (above college level). The JEI instrument is written at the 10th grade reading level, and thus can be completed by incumbents, rather than job analysts. Unlike the PAQ, it has only one response scale per item--"Relative Time Spent." Factor analyses were conducted with all 153 items to determine overall dimensions. Separate factor analyses were conducted for each of five a priori (as based on the PAQ) divisions. Overall and division dimensions are given in Table 5.

Harvey and his associates compared factor structures of the PAQ and JEI with a sample of 85 government jobs (Harvey et al., 1987). The authors concluded that both instruments measured the same underlying dimensions of work behavior. Based on their results: (a) subjective similarities in factor structure were found; comparison among factors showed parallel structure, supporting the assumption of a relatively stable underlying structure of work; and (b) quantitative similarities in factor structure were found, as correlations between JEI and PAQ factors were in the .80's and .90's. Average interrater reliability across 85 government positions was .90 (n was at least two for each position). In the investigation of its utility for job evaluation (with the Coast Guard), JEI factors predicted PAQ compensation points; multiple correlations ranged from .73 for the five overall factors to .86 for the 31 overall and division factors.

The model is very complete for describing enlisted performance, and was given a high rating for completeness for describing job performance (90%). All 26 division dimensions would be useful for describing first-tour Army jobs, except for the formal supervision dimensions. Among the ten dimensions that seem to have some motivational aspects (shown in Table 5) are two that would apply to the Army project: Working under demanding job situations, and Engaging in personally demanding situations.

These dimensions were rated as "Very Easy" for SMEs to provide ratings on due to both the lower reading level and the comprehensiveness of the model. Army predictors and JEI job dimensions could be linked for synthetic validation purposes.

Table 5

Job Element Inventory Dimensions[a]

Overall Dimensions

1. Decision making/General responsibility
2. Skilled job activities
3. Information processing activities
4. Physical activities/related environmental conditions
5. Using equipment/providing service

Division Dimensions

1. Interpreting what is sensed
2. Using various sources of information
3. Being aware of environmental conditions
4. Using various senses
5. Decision making/People
6. Decision making/Things
7. Decision making/Quantitative
8. Decision making/Medical
9. Performing skilled/technical activities
10. Performing handling/related manual activities
11. Using miscellaneous equipment/devices
12. General physical coordination
13. Controlling machines/processes

Table 5 (cont.)

Job Element Inventory Dimensions[a]

*	14.	Supervision/coaching and related activities
	15.	Public/related personal contacts
*	16.	Supervision/judgment/coordination
*	17.	Exchanging job-related information
	18.	Engaging in general personal contacts
*	19.	Being in a stressful/unpleasant environment
*	20.	Being in hazardous job situations
*	21.	Engaging in personally demanding situations
*	22.	Being alert to changing situations
*	23.	Performing repetitive activities
*	24.	Interpersonal responsibility
*	25.	Working under demanding job situations
	26.	Performing unstructured vs. structured work

[a]Adapted from Harvey, R. J., Friedman, L., Hakel, M. D., & Cornelius, E. T., III. (1987, April). Dimensionality of the Job Element Inventory (JEI), a simplified worker-oriented job analysis questionnaire. Paper presented at the Second Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta, GA.

* Motivational component

Functional Job Analysis (Fine). A job classification system incorporated in the Dictionary of Occupational Titles (DOT) coding structure was originally developed by Fine (1962, 1963, 1974). Functional Job Analysis (FJA) is both a uniform and consistent way to describe what workers are expected to do and a method of measuring levels of worker activity. A standardized language is used in the system, and a distinction is made between actions of the worker and outcomes of his/her activity: the actions are modified by the means to accomplish the end results. Actions occur in relation to one of three independent and ordinal hierarchical worker function scales; these are listed in Table 6.

A worker's relationship to each People, Data, and Things scale is specified by the highest appropriate function in that hierarchy, ruling out more complex functions and including all simpler functions (those below the targeted appropriate function). Weights ranging from 1 to 8 are applied to indicate the degree of involvement for each scale; this specifies importance of Things, Data, and People to the job.

Table 6

Functional Job Analysis Function Scales[a]

Things	*3a.	Precision Working
	b.	Setting Up
	c.	Operating-Controlling II
	2a.	Manipulating
	b.	Operating-Controlling I
	c.	Driving-Controlling
	d.	Starting Up
	1a.	Handling
	b.	Feeding-Offbearing
	c.	Tending
Data	** 6.	Synthesizing
	**5a.	Innovating
	b.	Coordinating
	4.	Analyzing
	3a.	Computing
	b.	Compiling
	2.	Copying
	1.	Comparing
People	**7.	Mentoring
	**6.	Negotiating
	**5.	Supervising
	4a.	Consulting
	b.	Instructing
	c.	Treating
	3a.	Coaching
	b.	Persuading
	c.	Diverting
	2.	Exchanging Information
	1a.	Taking Instructions-Helping
	b.	Serving

[a]Adapted from Fine, S. A. (1962). Functional job analysis as a method of indirect validation: A study in synthetic validation. Unpublished doctoral dissertation, George Washington University, Washington, DC.
* Fine listed components hierarchically within scales to reflect the degree of complexity of those components.
** Motivational component

Later developments with the FJA approach included the addition of four more scales, each having five to eight levels (Fleishman & Quaintance, 1984). The four scales are: (a) Worker Instructions, (b) Reasoning Development, (c) Mathematical Development, and (d) Language. The approach can be used to categorize jobs as wholes, or to compare and measure requirements of individual tasks and synthesize information across tasks. Fleishman and Quaintance (1984) concluded that (a) FJA has high utility in the areas of performance appraisal, training, and performance test development, and (b) information regarding both reliability and external validity is sparse at this time.

Four raters independently rated 100 job definitions to investigate reliability of worker function ratings. Average proportions of exact agreements among rater pairs was .59 for the Things scale, .66 for the Data scale, and .88 for the People scale (Fine, 1962). An investigation of the validity of FJA compared expert judges' estimates of importance of the components for successful job performance with empirical prediction, using the GATB. Judges' estimates were equivalent to empirical prediction of required aptitude levels or aptitude patterns for 63 of the 85 jobs studied (Fine, 1962).

All of the model's scales would be useful for describing Army first-tour MOS except the Supervision scale. Three of the "People" scales--Mentoring, Negotiating, and Consulting--are of an informal supervisory nature, and would therefore be appropriate to use for first-tour soldiers. Motivational components are included in the model (shown in Table 6). The model received a high rating for completeness for describing Army job performance (90%). However, Army job experts might have difficulty applying this system to describe jobs or to rate performance across MOS, as the hierarchical nature and the separation of the three scales would be problematic for them. Direct empirical links between Army predictors and FJA components could be investigated.

Occupation Analysis Inventory (Cunningham, Boese, Neeb, and Pass). The Occupation Analysis Inventory (OAI) was developed by Cunningham and his associates (Cunningham, et al., 1983) to describe and classify occupations based on their educational requirements. The inventory is based on the information-processing entities of Information Received, Mental Activities, Work Behavior, Work Goals, and Work Context. Although classified as a behavior description model, task characteristic elements are mixed in with behavior description elements. The OAI contains 617 job elements; these are rated on one of four rating scales: (a) significance to the job, (b) extent of occurrence, (c) applicability, or (d) an element-specific scale. A set of 215 jobs was rated by rater pairs, and a much larger sample of 1414 jobs was rated by a single rater (Cunningham et al., 1983). Additional ratings of 103 human attributes on these elements were collected, and an ability requirements model was derived (similar to the PAQ work in this area). Again, the ability requirements model will be discussed with others of its type.

With its 617 job elements, the OAI is able to differentiate a wide range of jobs rather specifically. Factor analyses of the job ratings yielded 24 easily-interpreted and stable higher-order factors (see Table 7 for a listing of these higher-order factors derived from job rating data). Mean interrater reliability for job element ratings of the 215 jobs was .53. Multiple correlations showed definite relationships between OAI factor scores and the GATB scores (ranging from .57 to .83), although substantial shrinkage occurred with cross-validation (corresponding average cross-validated correlations shrank to .15 through .67) (Cunningham et al., 1983).

The model is comprehensive and would be useful for rating Army jobs, with the addition of motivational components. The rating scales and components have been successfully used by military SMEs (Air Force); Army SMEs would have little or no difficulty using the OAI system. Linkages of OAI components to Army predictors would be possible as for several of the other structured job analysis instruments (e.g., PAQ, JEI).

Table 7

Occupation Analysis Inventory (OAI): Work Dimensions[a]

1. Human development, assistance, and conflict resolution
2. Sales, service, and public relations
3. Routine semantic and symbolic activities: clerical activities
4. Biological or health-related activities
5. Mechanical repair, maintenance, and operation
6. Activities related to visual aesthetics
7. Utilization and processing of numerical data
8. Botanical activities
9. Activities related to physical science and technology
10. Electrical or electronic repair, maintenance, and operation
11. Building or repairing structures
12. Use of technical-scientific devices
13. Working with animals
14. Improving or monitoring the physical performance, capability, or adjustment of others
15. Food preparation or processing
16. Technical planning and drawing
17. Assembly and fabrication activities
18. Environmental maintenance and planning
19. Performing arts activities
20. (Uninterpretable)
21. (Uninterpretable)
22. Vehicle and mechanized equipment operation
23. Organizing and supervising the work of others

Table 7 (cont.)

Occupation Analysis Inventory (OAI): Work Dimensions[a]

24. (Uninterpretable)
25. Instructing
26. Material handling or arrangement
27. (Uninterpretable)
28. Verbal communication

[a]Adapted from Cunningham, J. W., Boese, R. R., Neeb, R. W., & Pass, J. J. (1983). Systematically derived work dimensions: Factor analyses of the Occupation Analysis Inventory. Journal of Applied Psychology, 68, 232-252.

Behavior Description Approach: Minor Models. There are additional "minor" behavior description models; the interested reader should refer to Appendix C for more information. Willis' (1961) model was developed to investigate implications of learning principles and behavior classifications for training devices. Four broad task types were defined in a semantic approach by Bennett (1971) with worker-oriented verbs as the descriptors. Peterson and his colleagues developed models for specific jobs which include motivational components (Peterson et al., 1977; Peterson & Houston, 1980) and additional components useful for the Army (Peterson et al., 1982). A final minor model reviewed was the British adaptation of the PAQ, called the Job Structure Profile (JSP) (Patrick & Moore, 1985).

Behavior Requirements Models

Job Element Method (Primoff). Primoff originally developed his Job Element Method to determine selection criteria for trade and industry jobs. The procedure also qualifies as a form of synthetic validation (see discussion of J-coefficient on page 9 of Chapter 2). This is a structured method whereby SMEs provide ratings of job elements--knowledge, skills, abilities, interests, and/or personal characteristics--characteristics of the worker that influence his or her success on the job. Job elements can be used for (a) identifying selection tests and (b) developing tailor-made tests for specific jobs.

In order to identify selection tests, SMEs generate job elements and then rate the extent to which each element is present in a focal job. Ratings are summed over raters and then used to calculate--via a mathematical formula--a validity estimate (the "J-coefficient") of a standardized selection test(s). Job elements identified by SMEs are also used for purposes of test development. However, ratings are obtained on four alternative scales that assess the importance and relevance of the elements for the job.

Primoff (1972) reported a study in which six elements were identified for clerical jobs: (a) Spelling ability, (b) Reading ability, (c) Grammar ability, (d) Word Meaning ability, (e) Filing ability, and (f) Education ability. These six components could be useful for describing Army MOS that contain clerical activities; Project A job performance measures overlap somewhat with these elements (e.g., Record and File Information, Prepare Technical Forms and Documents).

Examples of job element lists for other job categories are given by McCormick (1979); these are:

For trades and labor occupations -

Operation of motor vehicles
Knowledge of riveting
Knowledge of welding
Theory of electronics
Ability to use electrical drawings
Trouble-shooting (mechanical)

For office positions -

Ability to proofread by self
Ability to help people find things in files
Ability to meet short deadlines
Ability to do editorial checking for spelling
Accurate and rapid typing
Memory for names and faces of people
Knowledge of secretarial practices.

Although Primoff (1972) did not present reliability results he provided some guidance for measuring the "reliability" of a J-coefficient. In rating importance of elements rather than rating job proficiency of people, the rater does not get caught up in personal factors related to the ratee. In addition, it is more practical to collect ratings of job elements from multiple raters than to collect ratings of worker job proficiency from multiple raters (Primoff, 1957). Reliability is measured not as the lack of dispersion in ratings for a group of judges, but rather as the agreement between final results of two groups of raters for an identical job (Primoff, 1972).

Since the J-coefficient is an estimate of the validity coefficient, calculated J-coefficients have been compared with obtained validity coefficients. J-coefficients consistently approximate obtained validities across a variety of jobs and tests (Primoff, 1959, 1972). Primoff (1959) reported a J-coefficient of .48 vs. an obtained validity of .51 for a predictor test for industrial planning jobs.

The Job Element Method is a procedure for deriving a unique set of elements for a job or set of jobs, rather than an actual job component model. Thus, it is not evaluated against our set of evaluation criteria. The technique could be employed by the Army, with elements generated by SMEs, but the resulting sets of elements for each MOS would not serve the needs of the current project (one model covering all entry-level MOS). However, the J-coefficient formulas are useful for the Army's synthetic validation work, as described in Chapter 2.

Job Components Inventory (Banks and Miller). Banks and Miller (1984) extended the application of the Job Components Inventory (JCI) from low skill youth jobs (the population it was developed for) to letter carrier jobs as part of an investigation of the reliability and convergent validity of the JCI. Five of the seven main sections of the JCI contain items of a behavioral requirements nature. These sections are: (a) Tools and Equipment, (b) Physical and Perceptual Requirements, (c) Mathematical Requirements, (d) Communication Requirements, and (e) Decision-making and Responsibility.

Banks and Miller (1984) obtained adequate reliability using the JCI, in terms of both inter-rater agreement (overall mean correlation = .75) and supervisor-incumbent agreement (mean = .72). Validity evidence was not reported.

The model received a moderate rating for completeness for rating Army enlisted job performance. Although the five behavioral requirements components would be useful for Army enlisted MOS, they are very general and are not at the most appropriate level for synthetic validation purposes. One component included in the model taps motivation--Decision-making and Responsibility--while the remaining four are "can do" components. Compared with the Project A job performance model, it does not sufficiently address the motivational aspects of performance (with the single Decision-making component). Further, Army SMEs would probably have difficulty providing ratings, finding the components too general. However, the judgments required (e.g., either Yes-No or four-point frequency responses) are very straightforward. There is a potential to link Army predictors and JCI job components.

Behavior Requirements Approach: Minor Models. The reader is referred to Appendix D for additional behavior requirements models. Those classified as "minor" relative to the Army's specific needs are:

- Gagne's (1974) model of learning process categories

- Folley's (1964) training theory-based model to aid development of military training devices and programs

- E.E. Miller's (1969) classification of both tasks and training strategies

- Alluisi's (1967) performance function categories for assessing complex performance

- Posner's (1962) information-processing tasks model

- Sternberg's (1977) information-processing model, applied to mental abilities

- Fleishman and Quaintance's (1984) Criterion Measures Approach

- Levine and Teichner's (1973) Systems Language (Information-Theoretic) Approach

- Drewes' (1961) Elemental Motions model

- R. B. Miller's (1962) Task Categories Approach.

Ability Requirements Models

Position Analysis Questionnaire (McCormick, Jeanneret, and Mecham). Development and use of the PAQ was discussed in the behavior description models section. Judges (psychologists) rated the relevance of 68 selected attributes to each PAQ job element; median values for these attributes were used to reflect the profile for the elements. Factor analyses yielded 21 dimensions for the six PAQ divisions (McCormick et al., 1972). Attribute dimensions are listed in Table 8.

Average reliability for various pairs of individuals using the PAQ was .79 for the overall dimensions; the average item reliability was .80. Compared to the dimensions based on job analysis data, attribute profile dimensions explained more variance on PAQ divisions. Also, attribute dimensions were easier to interpret and label (having a more consistent internal structure) than job analysis dimensions (McCormick et al., 1972).

McCormick et al. (1972) investigated the validity of the attribute model by testing the assumptions that (a) jobs differ in the level of a given aptitude required for successful performance, and (b) individuals tend to migrate into jobs they have the requisite abilities for. The investigators derived job dimension scores on the 21 attribute dimensions for 179 positions; these positions were matched against 90 positions for which criterion data were available. Dimension scores were used to predict two sets of criteria--mean test scores on the (nine) GATB tests and validity coefficients. Cross-validated regression analyses yielded a median $\underline{R}$ (across the nine tests) of .70 for the mean test score-attribute profile correlations and a median $\underline{R}$ (across the nine tests) of .44 for the validity coefficient-attribute profile correlations.

Table 8

Position Analysis Questionnaire: Job Dimensions Based on Attribute Dimensions[a]

Dimensions of information input
1. Visual input from devices/materials
2. Perceptual input from processes/events
3. Evaluation of visual input
4. Nonvisual input
5. Physical/environmental awareness
6. Verbal/auditory input/interpretation
Dimensions of mediation processes
7. Use of job-related knowledge
8. Information processing
Dimensions of work output
9. Manual control/coordination activities
10. Control/equipment operation
11. General body/handling activities
Dimensions of interpersonal activities
12. Interpersonal communications
13. Serving/entertaining
14. Signal/code communications
Dimensions of the work situation and job context
15. Unpleasant/hazardous physical environment
16. Personally demanding situations
Dimensions of miscellaneous aspects
17. Attentive/discriminating work demands
18. Unstructured/responsible work activities
19. Paced/regular work activities
20. Businesslike work situations
21. Merit income

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

Based on the results, McCormick et al. (1972) concluded that compared to validity coefficients, mean test scores served as superior criteria in representing the importance of tests to jobs. Although predictor and criterion data were not obtained for the same jobs, the authors concluded there was support for synthetically determining aptitude requirements with the PAQ.

Even though it doesn't include motivational components, the model is very complete, and would prove useful for the Army's purposes. The PAQ is typically administered by highly-trained job analysts. If Army job experts, rather than job analysts, were to use the PAQ, they would probably have difficulty with the high reading level. Linkages with predictors could be investigated by obtaining expert judgments of validities or importance of each component for successful performance.

Occupation Analysis Inventory (Cunningham, Boese, Neeb, and Pass). The OAI was described in the behavior description approaches section (above). Briefly, five general divisions of the instrument (and 617 job elements) were derived from an information-processing framework (Cunningham et al., 1983). Trained psychology graduate students and job analysts rated elements on their requirements for 102 defined human attributes. These attributes formed six categories: (a) general vocational capabilities, (b) cognitive abilities, (c) psychomotor abilities, (d) sensory capacities, (e) interests, and (f) needs. The authors conducted "sectional analyses," or factor analyses of six sections of the OAI, defined as: (a) information content, (b) physical work behavior, (c) representational work behavior, (d) interpersonal work behavior, (e) work goals, and (f) work context. Following the sectional analyses, a higher-order factor analysis was conducted that treated the six sets of sectional factors as one general set of work dimensions. The resulting 21 factors represent an ability requirements model. These 21 dimensions are listed in Table 9. Based on factor loadings, three to six sectional factors were identified for each dimension.

The mean reliability coefficient for the attribute weights (mean ratings across 9 or 10 raters) was .80. Results of correlational analyses with the eight GATB tests showed that in addition to characterizing jobs in job-oriented terms, OAI factors characterize jobs in human terms. Multiple correlations between sectional attribute rating factors and GATB tests ranged from .52 to .87 (although correlations showed considerable shrinkage in cross validation).

Motivational components are presently under-represented in the model. One sectional factor--Responsibility--is motivational in nature, and loads on several dimensions (e.g., Planning and innovation, Health-related Interaction and responsibility). If additional motivational components were included, the model would be very complete for describing Army jobs. As previously stated, the OAI has been used in military settings; rating scales and components are acceptable to military SMEs.

Table 9

Occupation Analysis Inventory: Attribute Requirements Dimensions[a]

Dimension Title
1. Machine operation, maintenance, and repair
2. Development and supervision of others
3. Mathematical/symbolic activities
4. Health/biological activities
5. Representation and production of figural arrangements and relationships
6. Activities related to the aesthetic appearance of others
7. Agricultural/botanical activities
8. Clerical activities
9. Verbal communication: Writing and speaking
10. Performing and visual/decorative activities
11. Material processing and modification
12. Business/sales activities
13. Activities requiring coordination and balance
14. Health-related interaction and responsibility
15. Construction and assembly activities
16. Planning and innovation
17. Direct interpersonal communication
18. Electrical/electronic maintenance, repair, and operation
19. Measuring, testing, and inspecting activities
20. General tool usage
21. General physical labor

[a]Adapted from Cunningham, J. W., Boese, R. R., Neeb, R. W., & Pass, J. J. (1983). Systematically derived work dimensions: Factor analyses of the Occupation Analysis Inventory. Journal of Applied Psychology, 68, 232-252.

Linkages between OAI attribute dimensions and Army predictors could be easily investigated by obtaining expert judgments of validities of attributes for job components.

Attribute Assessment: Initial Testing of Scales for Determining Human Requirements for Military Jobs (Smith and Rossmeissl). Smith and Rossmeissl (1986) followed an approach advocated by Fleishman (Fleishman & Quaintance, 1984) to obtain rational estimates of human attributes required for successful job performance. Estimates of these required attributes are obtained from qualified judges using 7-point Behaviorally-Anchored Rating Scales (BARS). Ratings are summarized to represent "profiles" of required attributes. For selection and classification purposes, attributes of an individual can be matched to these derived profiles.

The Attribute Assessment Scale (AAS) is a set of 22 BARS, designed to be used by supervisory personnel, not psychologists. The AAS was empirically derived from Project A data, specifically, from estimates of the validity of 53 predictors against 72 criterion constructs. The model's 22 components very closely approximate the 21 attributes (cognitive, perceptual, psychomotor, biographical, vocational interest, and temperament) identified in the Project A taxonomy (Wing et al., 1984). Components are categorized into three groups: Cognitive/perceptual, Physical/psychomotor, and Noncognitive (Temperament and Interest); individual components are listed in Table 10.

Non-commissioned Officers (NCOs) used the AAS to estimate profiles of human attributes required for three different performance levels for two different MOS (Smith & Rossmeissl, 1986). Interrater reliability for the AAS was very low (at least 30 raters would be needed to achieve reliability of .60). In the follow-up study using only one performance level and generic, rather than behaviorally-anchored scale anchors, interrater reliabilities for the mean ratings were .73, .84, and .43 for three MOS (Motor Transport Operator, Administrative Specialist, and Ammunition Specialist, respectively).

Tryout of the attribute scales yielded only reliability data; validity was not investigated. Components added at a more specific level would probably yield greater differences in attribute requirements among MOS; a drawback with the AAS study was the uniformity of attribute profiles across (a limited number of) MOS (Smith & Rossmeissl, 1986). Eight motivational components are included; these are the noncognitive scales, marked in Table 10. The model contains a comprehensive set of components for describing enlisted performance in the Army. In addition, it would be easy for Army job experts to use the scales to describe jobs and/or rate performance. The model has the capability for linking job components to predictors; expert judges or psychologists could estimate validities of various tests for attribute requirements.

Psychomotor and Physical Proficiency Abilities (Fleishman and Colleagues). Fleishman and his colleagues (Theologus, Romashko, and Fleishman, 1970; Fleishman, 1975) set out to develop a taxonomy of human performance to apply to a range of theoretical and applied areas. Their approach was to conduct extensive factor and correlational analyses to define abilities common across performance on a range of tasks. This was done by designing tasks to tap hypothesized ability categories and administering tasks to subjects. Then task scores were correlated and

Table 10

Attribute Assessment Scale[a]

Attributes Included in the Attribute Assessment Scale

Cognitive/Perceptual Attributes

1. Verbal Ability
2. Memory
3. Reasoning Ability
4. Number Facility
5. Mechanical Comprehension
6. Information Processing
7. Closure
8. Visualization
9. Perceptual Speed and Accuracy

Physical/Psychomotor Attributes

10. Physical Strength
11. Stamina
12. Multilimb Coordination
13. Dexterity
14. Steadiness/Precision

Noncognitive Attributes

* 15. Social Interaction
* 16. Stress Tolerance
* 17. Conscientiousness
* 18. Work Orientation
* 19. Self Esteem/Leadership
* 20. Athletic Ability/Energy
* 21. Realistic Interests
* 22. Investigative Interests

[a]Adapted from Smith, E. P., & Rossmeissl, P. G. (1986). Attribute assessment: Initial test of scales for determining human requirements of military jobs. Alexandria, VA: Army Research Institute, Selection and Classification Technical Area. (Draft)

* Motivational factor

factor-analyzed. Eleven perceptual-motor factors and nine physical proficiency factors were identified; these are listed in Table 11.

Judges' estimates of abilities were used to predict aspects of task performance and various learning measures (Fleishman, 1975). Expert raters estimated ability requirements for tasks, using behaviorally-anchored scales and empirically determined scale values for the tasks. Interrater reliabilities with scales of this type ranged from .55 to .65.

In terms of validity evidence, some summary statements were made by Fleishman (1975). Both experimental and correlational studies have investigated interactions between characteristics of tasks and ability requirements. Based on comparisons and correlations with ability requirements for "reference" tasks, differences in ability requirements were related to variations in task dimensions.

The model has the capability of incorporating new data and new components, but as it is so complete, it is not likely that much would have to be added to accommodate new MOS. However, motivational components are missing. Raters would be likely to accept and be able to use the model's components; as long as rating procedures were clarified (i.e., the concept of "validity" described in detail). Linkages could be established with Project A predictors by collecting SME validity judgments.

Ability Requirements Approach: Minor Models. Additional ability requirements models are briefly described in Appendix E, with listings of their components. Those discussed include two studies to develop job-specific taxonomies, for the Correctional Officer job (Peterson et al., 1977), and for government higher-level clerical jobs (Kintop & Mussio, 1974). The dimensions and categories of Guilford's (1971) Structure of Intellect model are also presented.

Table 11

Fleishman and colleagues: Ability Requirements Taxonomy[a]

Cognitive, Perceptual, Sensory, Physical, and Psychomotor Components
1. Verbal Comprehension
2. Verbal Expression
3. Ideational Fluency
4. Originality
5. Memorization
6. Problem Sensitivity
7. Mathematical Reasoning
8. Number Facility
9. Deductive Reasoning
10. Inductive Reasoning
11. Information Ordering
12. Category Flexibility
13. Spatial Orientation
14. Visualization
15. Speed of Closure
16. Flexibility of Closure
17. Selective Attention
18. Timesharing
19. Perceptual Speed
20. Static Strength
21. Explosive Strength
22. Dynamic Strength
23. Stamina
24. Extent Flexibility
25. Dynamic Flexibility
26. Gross Body Equilibrium
27. Choice Reaction Time
28. Reaction Time
29. Speed of Limb Movement
30. Wrist-finger Speed
31. Gross Body Coordination
32. Multilimb Coordination
33. Finger Dexterity
34. Manual Dexterity
35. Arm-hand Steadiness
36. Rate Control
37. Control Precision

[a]Adapted from Theologus, G. C., Romashko, T., & Fleishman, E. F. (1970). Development of a taxonomy of human performance: A feasibility study of ability dimensions for classifying human tasks. Washington, DC: American Institutes for Research.

Summary and Conclusions

There are four major types of job component models: the task characteristics, behavior description, behavior requirements, and attribute requirements approaches. These approaches differ primarily in the type of descriptor used to define job/task categories. We have reviewed the "major" models of each type. Additional variations of these models are discussed briefly in Appendices B-E. Several evaluation criteria were described, and models were evaluated against these criteria. Four of these criteria related to the use of particular job component models in describing Army jobs. The fifth criterion concerned linkage to potential selection measures and the sixth addressed conformance with a model of Army job performance.

The twelve "major" models reviewed in this chapter are summarized in terms of the evaluation criteria in Table 12. We would like to emphasize several points and then draw some conclusions from the table. First, the term "interrater reliability" has been used many times throughout the discussion of specific job component models. However, the reader is advised that these reliabilities are not directly comparable across studies. These interrater reliabilities were calculated on the basis of various scales (e.g., frequency, time, importance), various computational methods (e.g., intra-class, Pearson r's), and various entities (e.g., single elements, means of single elements, profiles). The reader should be cautioned that treating all interrater reliabilities reported here as directly comparable estimates would be mixing apples and oranges.

A second point is that cell entries in several columns of the table are based wholly or in part on subjective judgments by the authors. Other reviewers might have interpreted the evidence differently. Most subjective in nature is the Acceptance/Ease of Use criterion.

Several conclusions can be drawn from the table. First, no single approach seems uniformly best. Second, most models lack motivational components. Third, no one model in its current state is entirely appropriate for our purposes. Fourth, the method(s) associated with the application of any given model may be useful to the synthetic validation project, apart from the job components that comprise it.

As previously mentioned, several authors (e.g., Prien, 1977; Ash, 1982) have recommended a multi-methodological approach to development and use of job component models. Gathering task-based and behavior-based job analytic information has definite advantages. Several of the most widely-used job component models, for example the PAQ and the OAI, contain items representing a mix of descriptor bases. Developing a mixed model that incorporates task-based, behavior-based, and attribute-based elements and methods for describing jobs appears to be a viable option for the Army.

The second main conclusion drawn from the review is that a majority of models do not include motivational factors. The expectation that

Table 12. Summary of Job Component Models

	CRITERIA*				
	Use in Job Description				
MODEL (by author(s))	(1) Reliability	(2) Validity	(3) Comprehensiveness	(4) Acceptance/Ease of Use	(5) Ability to Provide Linkages with Selection Tests
Task Characteristics;					
(1) Knerr, et al. (1982) "MANPERS" Project Task Taxonomy	Not Available	Not Available	Lacks some technical components and some motivational components	Moderate; components are limited but phrased in familiar (military) terms	Moderate
(2) Farina & Wheaton (1971)	Moderate-High; Intraclass correlation coefficients ranged from .58 - .98	Moderate; R between job components and performance = .82 and .73	Lacks motivational components	Low; components are abstract and detailed	Moderate
Behavior Description					
(3) Berliner et al. (1984)	Not Available	Not Available	Lacks motivational components	Low; component form-verbs-is restrictive, but set of components is comprehensive for military jobs (originated in military context)	Moderate
(4) McCormick et al. (1972) Position Analysis Questionnaire (PAQ)	Moderate; .59 - .89	High; similar job characteristics across jobs are correlated with wage/salary rates	Lacks motivational components	Moderate; components are comprehensive but reading level difficulty is high for SMEs	Moderate

*Criteria: Key
(1) Reliability or consistency across judges in component assignments
(2) Validity of resulting job descriptions
(3) Comprehensiveness of coverage of important aspects of job performance
(4) Acceptance/Ease of Use by raters
(5) Ability to Provide Linkages with Selection Tests -- Project A predictors

Table 12. Summary of Job Component Models

MODEL (by author(s))	CRITERIA* Use in Job Description: (1) Reliability	(2) Validity	(3) Comprehensiveness	(4) Acceptance/Ease of Use	(5) Ability to Provide Linkages with Selection Tests
(5) Cornelius & Hakel (1978) Job Element Inventory (JEI)	High; average inter-rater reliability across jobs = .90	High; multiple *R*'s obtained *for* predicting PAQ compensation points from JEI factors ranged from .73 (first-order overall factors) to .86 (first-order overall plus division factors)	Includes some motivational components but lacks others	High; components are comprehensive and at appropriate reading level	Moderate
(6) Fine (1962, 1974) *Functional* Job Analysis (FJA)	Moderate; Average *proportions of exact* rater pair agreements, by scale: .59 Things, .66 Data, .88 People	Moderate; analysts' *results* equaled empirical prediction of required aptitude levels/patterns for 63 jobs	High	Low; hierarchical nature of model and scales are difficult to relate to Army jobs	Low-Moderate
(7) Cunningham, et al. (1983) Occupation Analysis Inventory (OAI)	Low; Mean interrater reliability = .53	High; multiple *R*'s between job components and GATB scores ranged from .57 - .83	Lacks technical components	High; components are comprehensive and model has been used successfully with military jobs	Moderate
(8) Banks & Miller (1984) Job Components Inventory (JCI)	Moderate; Overall interrater mean correlation = .75; supervisor-incumbent mean agreement = .72	Not Available	Lacks motivational components and some technical components	Low; components are very general	Low

*Criteria: Key
(1) Reliability or consistency across judges in component assignments
(2) Validity of resulting job descriptions
(3) Comprehensiveness of coverage of important aspects of job performance
(4) Acceptance/Ease of Use by raters
(5) Ability to Provide Linkages with Selection Tests -- Project A predictors

Table 12. Summary of Job Component Models

	CRITERIA*				
	Use in Job Description				
MODEL (by author(s))	(1) Reliability	(2) Validity	(3) Comprehensiveness	(4) Acceptance/Ease of Use	(5) Ability to Provide Linkages with Selection Tests
Ability Requirements:					
(9) McCormick, et al. (1972) Position Analysis Questionnaire (PAQ Attribute Dimensions)	Moderate; .79 for overall dimensions; .80 for average item reliability	High; mean aptitude (GATB) test scores were predicted more accurately by attribute profiles than were validity coefficients (.70 and .44, respectively) for "matched" jobs	Lacks motivational components	Moderate; components are comprehensive but reading level is inappropriately high for SMEs	High
(10) Cunningham, et al. (1983) Occupation Analysis Inventory (OAI Attribute Requirements Dimensions)	Moderate; .80	High; multiple R's between sectional attribute rating factors and GATB tests ranged from .52 - .87 (before shrinkage)	Lacks motivational components	High; components are comprehensive and model has been used successfully with military jobs	High
(11) Smith & Rossmeissl (1985) Attribute Assessment Scale (AAS)	Moderate; Interrater reliabilities of .43, .73, and .84 for three different MOS	Not Available	High	High; used successfully within military context	High

*Criteria: Key
(1) Reliability or consistency across judges in component assignments
(2) Validity of resulting job descriptions
(3) Comprehensiveness of coverage of important aspects of job performance
(4) Acceptance/Ease of Use by raters
(5) Ability to Provide Linkages with Selection Tests -- Project A predictors

Table 12. Summary of Job Component Models

	CRITERIA*				
	Use in Job Description				
MODEL (by author(s))	(1) Reliability	(2) Validity	(3) Comprehensiveness	(4) Acceptance/Ease of Use	(5) Ability to Provide Linkages with Selection Tests
(12)Fleishman, et al. (1969, 1970, 1975) Cognitive, Perceptual, Psychomotor, & Physical Proficiency Abilities	Moderate; .57 - .65	Moderate; variations in task dimensions have been related to differences in ability requirements	Lacks motivational components	Moderate; components are comprehensive but some are too abstract for easy interpretation	High

*Criteria: Key

(1) Reliability or consistency across judges in component assignments

(2) Validity of resulting job descriptions

(3) Comprehensiveness of coverage of important aspects of job performance

(4) Acceptance/Ease of Use by raters

(5) Ability to Provide Linkages with Selection Tests -- Project A predictors

three of the four basic job component approaches (the behavior description, behavior requirements, and attribute requirements approaches) might contain motivational dimensions/factors was not supported by the literature. Components categorized as motivational in our review of the literature came mostly from proprietary technical reports, suggesting that these components have been included only where their importance was stressed.

Based on an informal content analysis of the models that did contain motivational components, five basic categories emerge. Motivational components seem to qualify as one of the following: (a) social habits, presentation, or interaction; (b) work habits; (c) personality or personal preferences, and (d) interests. These areas are consistent with the view of the structure of Army job performance developed in Project A.

Regarding the third conclusion, many of the models were developed for a specific purpose or situation. Therefore, they are not comprehensive enough for a synthetic validation model, even though their components are at an appropriate level of specificity. On the other hand, models driven by theory development have more abstract components that would be rather difficult to apply to Army tasks or activities. However, all models could be expanded to include more components and thus provide coverage of all entry-level Army jobs at an appropriate level of specificity.

The fourth conclusion drawn from Table 12 concerns methodologies suggested in the literature for utilizing these models. As the Army has already collected predictor data and concurrent validity data for a set of representative MOS, the choice of a methodology such as the J-Coefficient Approach makes sense. Linkages between predictors and elements can be calculated and compared to obtained validities for the subset of MOS studied intensively in Project A.

The criteria we used in evaluating models throughout this review must be applied to the development and evaluation of the Army's final model, regardless of which major type of approach it most closely approximates. Knerr et al. (1982) supplied criteria for evaluating component models for facilitating Army manpower and personnel requirements decisions. Among these were several practical criteria that also apply to job component model use. The final model(s) should be: (a) descriptive of MOS, (b) able to discriminate among MOS, and (c) consistent with Army regulations and practices. In addition, the model(s) should: (d) not require direct observation for determining categories, (e) use terms familiar to SMEs, and (f) facilitate decisions.

One final conclusion can be drawn on the basis of our review and the summary table (Table 12). We clearly must take an eclectic approach in the development of a job component model for the Army's use. We will be able to adopt components from some of the models discussed. However, we also will need to adapt the model and the components themselves as necessary to fit the specific Army synthetic validation context.

EXPERT JUDGMENTS

Participation of expert judges will be required in the Synthetic Validation Project as the basis for establishing two of the linkages in the synthetic validation model. First, judges will provide estimates of criticality of job components for job performance. Second, judges will provide validity estimates that link individual job components to established predictor measures. These are two very different types of judgments, and demand judges with different expertise. It is likely that Army job SMEs will participate in job description. SMEs in the area of individual differences and test validation will likely provide validity judgments.

In this chapter, we first address job description judgment issues, specifically: (a) reliability and validity of the judgments, (b) relationships among individual difference variables and judgments, and (c) rater training. We then address the linkage between components and predictors and consider (a) the quality of judged validities and (b) qualifications of judges.

Job Description Judgments

Reliability and Validity of Judgments

A central issue of the literature concerning job description reliability is the effect of role of the judge on reliability and validity of the judgments obtained. Typically, investigations compare patterns of job description judgments and rating means among three judge groups: incumbents, supervisors, and job analysts.

Smith and Hakel (1979) found contradictory conclusions concerning degree of agreement among incumbent, supervisor, and job analyst descriptions of the worker's job. They cited studies reporting large discrepancies in allocation of time to major job areas (Lawshe, 1953) and in determination of a foreman's responsibilities (Meyer, 1959). However, Hazel, Madden, and Christal (1964) reported that supervisors and incumbents had better agreement on general duties than on specific tasks.

Our review of more recent research indicates greater support for agreement, rather than disagreement, among supervisors' and incumbents' job descriptions. In analyzing a correctional officer's job, Peterson, Houston, Bosshardt, and Dunnette (1977) found high agreement between supervisors' and incumbents' importance ratings on 106 task statements. Correlation of the average profiles between supervisors and incumbents was .82. The correlation between supervisors' and incumbents' rating profiles of 48 worker characteristics was .72. Effectiveness rating profiles of critical incidents were also highly correlated (.75-.98) among supervisors, administrators, incumbents, and inmates.

Traditional reliability analyses and generalizability analyses were conducted to assess agreement among incumbent, manager, and job analyst judges in developing the general management Position Information Inventory for the J. C. Penney Co. (Toquam, McHenry, Hanson, McGue, & Peterson, 1986). Results of both interrater reliability and generalizability analyses indicated that incumbents, managers, and job analysts reliably and consistently identified responsibilities associated with incumbents' positions. Results from the generalizability studies indicated that on average, job analysts gave lower ratings than managers and incumbents, and there was no difference in the ratings given by managers and incumbents.

In their research, Smith and Hakel (1979) found that PAQ job descriptions from incumbents, supervisors, job analysts, and college students intercorrelated in the .90s. According to Smith and Hakel, the judge sources yielded judgments that were practically comparable. When they examined the mean ratings among judge types, they found that, compared to job analysts, supervisors and incumbents tended to "inflate ratings." Findings that incumbents and managers provide higher ratings than job analysts have been interpreted in several ways. Toquam et al. (1986) interpreted the higher agreement between incumbents and managers as support for their ratings. Smith and Hakel interpreted similar results to support reliance on job analysts, due to their lack of response bias.

Cornelius, Denisi, and Blencoe (1984) argued that Smith and Hakel's (1979) approach to computing agreement among judges overestimates the level of agreement. One problem was that Smith and Hakel computed validities by first obtaining mean values for PAQ elements for each judge type across all jobs, and then correlating these mean values for each judge type. According to Cornelius et al., the preferred method is to first compute all pairwise correlations among judges <u>within</u> a job, and then compute the mean for these correlations across jobs. The latter procedure produces considerably different results. In a partial replication of the Smith and Hakel investigation with the preferred calculation method, a mean convergent validity of .58 was obtained (between college students and job experts), as opposed to the .90s found in the earlier investigation (Cornelius et al., 1984). Unfortunately, the replication did not include comparisons of supervisor and incumbent ratings.

A second methodological issue raised was the treatment of the "Does Not Apply" (DNA) responses in calculating correlations. Including DNA responses artifactually introduces more variance, therefore elevating correlations (DeNisi, Cornelius, & Blencoe, 1987). Average correlations for college student-job expert ratings dropped from the .90s to .41 after excluding DNA responses and correcting for measurement error (Cornelius et al., 1984).

When considered together, these results indicate that incumbents, supervisors, and job analysts provide reasonably reliable job description ratings. There also is evidence that job analysts provide lower mean ratings than incumbents and supervisors. The question of what these differences mean remains. Are incumbents and supervisors providing judgments that are too lenient, or are job analysts' judgments too strict? Perhaps some attempt to minimize leniency effects on supervisor or incumbent judgments is warranted.

The validity of job description is less well documented. Often validity is assumed rather than demonstrated--in the absence of external criteria for job descriptions, reasonably reliable job descriptions are typically accepted as valid descriptions.

Validity of job description is sometimes assessed by comparing ratings from alternate rater groups with those of job analysts. As noted above, there was moderate agreement between ratings from incumbents and supervisors with ratings from job analysts. In addition to using job analysts' ratings as benchmarks for "validity" of job descriptions, job descriptions have been used to predict salaries. Job descriptions are frequently used to predict salary levels in job evaluation, and the predicted salary levels typically correlate well (.80s to .90s) with actual salaries (McCormick, 1979). However, in both of these types of validity investigations, the external criterion is not independently obtained. Any error (such as method bias) in the job description simply perpetuates itself.

Individual Difference Variables

There is little research on the effects of individual difference variables on job description. However, there has been considerable research about rater differencs in performance evaluation. While the focus of performance evaluation is a specific individual, the focus of job description is the typical performance of individuals. We speculate that the findings may be relevant for job description. Cascio (1987) summarized the literature on individual difference variables as moderators of judgment outcomes. Listed below are the variables likely to be relevant to job description judgments:

Age	No consistent effects
Education level	Significant but extremely weak effect
Interests, social insight, intelligence	No consistent effect
Cognitive complexity	Unrelated to accuracy, halo, acceptability, or confidence in ratings
Job experience	Significant but weak positive effect on quality of ratings
Performance level	Effective performers tend to produce more reliable and valid ratings.

It appears that only rater performance level has clear implications for the quality of ratings. Other individual factors--job experience and educational level--have been related to significant but very weak effects on improving judgment quality. In one investigation of individual differences as moderators of responses on job description questionnaires, Silverman, Wexley, and Johnson (1984) found that experience and age did not significantly affect ratings of clerical work activities.

In sum, research suggests that performance level, and perhaps education level and job experience might lead to more accurate job performance ratings, but their relevance for job description must be demonstrated.

Rater Training

Several of the studies described thus far (Smith & Hakel, 1979; Toquam et al., 1987) have shown that job incumbents and supervisors tend to provide higher mean job description judgments than job analysts. If the Army intends to use incumbents and supervisors for job description, rater training may help to reduce rater errors. Most rater training attempts to reduce "halo" and "leniency" effects on performance evaluations, rather than on job description. Although training for performance evaulation is not designed for job description per se, ways in which training can improve judgment reliability and validity are suggested by this area of literature.

Much of the research has demonstrated that training judges to avoid various types of rating errors is successful. Ivancevich (1979), for example, found that extensive discussion of potential rating errors reduced leniency and halo effects. Similarly, Latham, Wexley, and Pursell (1975) found that training managers to be sensitive to the problems related to rating errors was very effective. Ratings provided by the managers six months after training were found to be free of contrast, halo, similarity, and first impressions errors. Although there is much support for training judges, Bernardin and Pence (1980) cautioned that training may improve reliability of the judgments (i.e., reduce the rating errors) at the expense of accuracy.

Providing training to naive judges for the purpose of job descriptions has yielded dismal results. Friedman and Harvey (1986) showed that with practice and extensive job information, naive raters could not provide PAQ judgments of the same level of reliability and validity as experienced job analysts. Extensive job information and practice for novices will not substitute for real job content knowledge.

Summary and Conclusions of Job Description Judgment Literature

The literature on the use of expert judges supports the following decisions regarding the Army project's job description requirements: (a) job supervisors, incumbents, and job analysts provide reliable and accurate ratings; (b) some discussion of "leniency effects" should be provided in the research protocol and/or judgment procedures should attempt to reduce these errors; and (c) some attempt should be made to select judges that are "effective performers" in their own jobs.

Validity Judgments

An important aspect of the Army's synthetic validation project is the linking of known predictors to job components that come out of the job description research. If the linkage is to be established judgmentally, the accuracy of such estimates and the qualification of judges have to be addressed.

A number of researchers have employed expert judges to estimate test validities for job performance and have examined accuracy of the estimates. Peterson, et al. (1982) had 14 judges estimate the validity of 14 cognitive tests for 46 performance areas in clerical, technical and professional jobs. The validity judgments for the tests were used to obtain a predicted composite score for each performance area. Similarly, as test data were available, a composite score was obtained with tests serving as predictors of each performance area. The two methods yielded almost identical results; the average difference between the 46 pairs of predicted scores was only .0165.

Schmidt, Hunter, Croll, and McKenzie (1983) also explored the extent to which expert judgments could substitute for criterion-related validities. The jobs studied were steward, communications technician, hospital corpsman, communications technician, radioman, firecontrol technician, aviation machinist's mate, aviation ordnanceman, and aviation electrician's mate. The tests were the six subtests of the Naval Basic Test Battery (which included tests of cognitive aptitudes such as arithmetic and verbal reasoning, and mechanical, shop, and electronics tests). The criterion measure was overall success in training, measured primarily through paper-and-pencil methods. Validity judgments by 30 personnel psychologists were compared with validities obtained from large validation studies with sample sizes exceeding 3,000. Schmidt et al. (1983) found that although judges tended to underestimate the true validity of a test, pooled judgments of 30 judges were as accurate as a criterion-related study with a sample size of 1,164. Therefore, "given highly trained and experienced judges, expert judgment may provide more accurate estimates of validity for cognitive tests than do local criterion-related validation studies" (p. 590).

These two investigations indicate that expert judgments yield accurate estimates of validities for tests of general aptitudes. Using similar procedures, Wing, Peterson, and Hoffman (1984) extended both the predictor and criterion space beyond general aptitudes and overall success. Wing et al. had judges estimate true validities for a 53 by 72 predictor-criterion matrix. The predictors included measures of cognitive abilities, spatial-psychomotor abilities, information processing, mechanical ability, social skills, vigor, and motivation.

The criteria were grouped into five types of skills: technical, clerical, combat, interpersonal, and commitment.

Reliability analyses of judges' estimates yielded satisfactory results. The mean reliability for the vector of predictors and the vector of criteria was .97 and .99, respectively. The reliability of cell means (i.e., validity of a given predictor for a given criterion) across all raters was .96. Single-rater reliabilities were .52 for predictors and .41 for criteria. Further analyses were conducted on the cell means, i.e., the judged validity of each predictor for each criterion.

The accuracy of every test-criterion estimate could not be ascertained because prior empirical research used different criteria. Where corresponding empirical validities were available, judgmental validities were comparable with one important difference. Similar to the Schmidt et al. (1983) results, validity estimates were lower than the empirical estimates. In the near future, when empirical validity data gathered for the concurrent validation phase of Project A are compared against these judgments, a clearer picture of the accuracy of these judgmental validities will be possible.

In summary, researchers have employed the same basic procedure for estimating validities across military as well as non-military jobs. Validity judgments have been obtained for a variety of tests (e.g., general aptitudes, interests, motivation) and a variety of criteria (e.g., training, hands-on measures, ratings). The results are compelling; the procedure yields highly reliable and perhaps somewhat conservative estimates of validity.

Judge Qualifications

Hirsh, et al. (1986) replicated the Schmidt et al. (1983) study with 28 less experienced judges (recent Ph.D.'s in industrial-organizational psychology). Results indicated that, compared to the experienced judges, less experienced judges provided inferior judgments. The systematic error of the less experienced judges, although small, was almost four times that of the experienced judges (i.e., .0732 vs. .019). The authors suggested that "the accuracy of a single expert is equal to the accuracy of the pooled judgments of ten less experienced judges" (p.344). Furthermore, while experienced judges tended to slightly understate empirical validities, the less experienced judges tended to overestimate empirical validities. However, the validity estimates of several less experienced judges are as accurate as that of a small sample empirical study.

Summary of Validity Judgment

Reliance on validity estimates from expert judges appears to be a viable complement to obtaining empirical validities. Judged validities are accurate for a variety of tests and criteria in numerous jobs. Judgments from highly experienced psychologists will yield very precise estimates. Less experienced psychologists can give useful judgments as well. However, experienced psychologists underestimated empirical validities, and no formulas currently exist to correct for such underestimates. One way to circumvent the problem might be to obtain relative validity values rather than absolute validities.

SUMMARY AND FUTURE DIRECTIONS

Synthetic validation has been described as "a technique of promise" (Mossholder & Arvey, 1984, p. 331), that could move personnel psychology "toward a scientific basis for the art of decision" (Guion, 1976, p. 821). To date, however, relatively few attempts have been reported. Factors restricting the number of synthetic validation studies to date include the traditional view that the validity of a test will not generalize across situations or time, and the focus on whole jobs, rather than job components, in most validation studies.

Several recent developments have begun to expand these limited approaches to investigating validity. The work of Schmidt and Hunter has promoted the application of validity generalizability. Fleishman and his colleagues have stressed the need for a taxonomic approach to job performance. Finally, case law has recently drawn attention to validation of all personnel decisions (e.g., hiring, training, and promotion). Procedures for conducting validation studies have diversified in response to these developments. However, the effort involved in taking a synthetic validation approach and the perceived complexity of associated statistical procedures still restrict its use. The Army's Synthetic Validation Project provides a unique opportunity for a large-scale, comprehensive application and evaluation of synthetic validation.

Each chapter of this review has focused on linkages in the synthetic validation process. The second chapter examined the methodology and procedures involved in establishing the linkages required for synthetic validation. The review of job component models provided criteria and alternative models from which to develop a job component model for the Synthetic Validation Project. The expert judgment chapter identified factors that affect the reliability and accuracy of judgments. Our review of these areas of literature revealed that past synthetic validation efforts have combined alternate choices of (a) job component model, (b) procedures for establishing linkages, and (c) judges for different purposes. Issues surrounding these three areas were discussed separately, in order to stress the importance of evaluating and identifying an optimum choice for each. However, the selections for appropriate job component model, synthetic validation methods, and expert judges are not completely independent. For example, the choice of job component model affects the selection of SMEs for providing criticality judgments and the statistical techniques employed to link job components and predictors.

This review of synthetic validation procedures has also identified several issues that warrant further research. One opportunity for further research concerns the methods used to establish prediction equations for individual job components. The reliability and validity of judgment-based estimates is not well-established across the range of possible types of job components. Even with the massive base of data collected in Project A, empirical validity estimates for each job component are not possible. The Project A data do, however, provide an excellent opportunity for assessing the validity of judgment-based approaches. Issues to be investigated include the effectiveness of different types of judgment paradigms, different judges, and different training procedures as well as the validity estimates obtained via the various job component models.

A second major opportunity for further research concerns the choice of the job component model that is used for job description and also for linkage to selection tests. A wide range of possible models was reviewed and several evaluation criteria for these models were discussed. No one approach appears uniformly best. There is a necessary tradeoff between ease of use for job description and ease of linkage to selection tests. At one extreme, task-based job components lead to reliable job descriptions--it is relatively easy to tell whether a task category is or is not a part of the job. The primary concern with a task-based approach is that it is relatively difficult to develop prediction equations for each job component since each may include a range of skill requirements. At the other extreme, ability requirement models provide components that are easily related to selection tests, but relatively difficult to relate to specific jobs. Behavior requirement and job behavior models fall between these two extremes. Further research will be required to assess the tradeoff between ease of use in job description and ease of linkage, in general and in the specific context of Army enlisted jobs.

A third opportunity for research concerns the development of job descriptions using alternative job component models. Different indicators of the relevance of each component (e.g., importance ratings, frequency of performance), as well as the choice of judges and the effects of alternative approaches to training judges need to be investigated for each type of job component model.

A final issue concerns the model used to develop an overall prediction composite and to estimate its validity. Of the possible alternatives, the J-coefficient model appears to be a reasonable approach. However, the tradeoffs between statistically esoteric J-coefficient derivations and formulas and simpler, straightforward unit-weighting procedures will need to be evaluated in terms of prediction. An additional concern is the degree of heterogeneity in the population of 250-plus entry-level jobs, and what effects heterogeneity might have on the formation of (e.g., number of predictors)--and the accuracy of--the composites. Overall job performance prediction equations developed by Project A provide a basis for evaluating different approaches.

REFERENCES

Alluisi, E. A. (1967). Methodology in the use of synthetic tasks to assess complex performance. Human Factors, 9, 375-384.

Ash, R. A. (1982). Job elements for task clusters: Arguments for using multi-methodological approaches to job analysis and a demonstration of their utility. Public Personnel Management Journal, 11, 80-89.

Balma, M. J. (1959). The concept of synthetic validity. Personnel Psychology, 12, 395-396.

Banks, M. H., & Miller, R. L. (1984). Reliability and convergent validity of the Job Components Inventory. Journal of Occupational Psychology, 57, 181-184.

Bennett, C. A. (1971). Toward empirical, practical, comprehensive task taxonomy. Human Factors, 13, 229-235.

Berliner, D. C., Angell, D., & Shearer, J.W. (1964). Behaviors, measures, and instruments for performance evaluation in simulated environments. Paper presented at a symposium and workshop on the quantification of human performance, Albuquerque, New Mexico.

Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.

Campbell, J. P. (1986). Validation analysis for new predictors. Unpublished manuscript.

Campbell, J. P., Hanser, L. M., & Wise, L. L. (1986). The development of a model of the Project A criterion space. Paper presented at the 28th Annual Conference of the Military Testing Association, New London, CT.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1987). Analysis of criterion measures: The modeling of performance. Paper presented at the Second Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta.

Cascio, W. F. (1987). Applied Psychology in personnel management. New Jersey: Prentice-Hall.

Cornelius, E. T., DeNisi, A. S., & Blencoe, A. G. (1984). Expert and naive raters using the PAQ: Does it matter? Personnel Psychology, 37, 453-464.

Cornelius, E. T., & Hakel, M. D. (1978). A study to develop an improved enlisted performance evaluation system for the U.S. Coast Guard. Washington, DC: Department of Transportation, United States Coast Guard.

Cotterman, T. E. (1959). Task classification: An approach to partially ordering information on human learning. (WADC TN 58-374). Wright Patterson Air Force Base, OH: Wright Patterson Air Development Center.

Cunningham, J. W., Boese, R. R., Neeb, R. W., & Pass, J. J. (1983). Systematically derived work dimensions: Factor analyses of the Occupation Analysis Inventory. Journal of Applied Psychology, 68, 232-252.

DeNisi, A. S., Cornelius, E. T., & Blencoe, A. G. (1987). Further investigation of common knowledge effects on job analysis ratings. Journal of Applied Psychology, 72, 262-268.

Dickinson, T. L., & Wijting, J. P. (1976). Policy-capturing as a procedure for synthetic validation. Paper presented at the meeting of the Rocky Mountain Psychological Association.

Drewes, D. W. (1961). Development and validation of synthetic dexterity test based on elemental motion analysis. Journal of Applied Psychology, 45, 179-185.

Farina, A. J., & Wheaton, G. R. (1971). Development of a taxonomy of human performance: The task characteristics approach to performance prediction. Washington, DC: American Institutes for Research.

Fine, S. A. (1962). Functional job analysis as a method of indirect validation: A study in synthetic validation. Unpublished doctoral dissertation, George Washington University, Washington, DC.

Fine, S. A. (1963). Functional Job Analysis (FJA) as a method of indirect validation. American Psychologist, 18, 438-445.

Fine, S. A. (1974). Functional job analysis: An approach to a technology for manpower planning. Personnel Journal, 53, 813-818.

Fitts, P. M. (1962). Factors in complex skill training. In R. Glaser (Ed.), Training research and education. Pittsburgh: University of Pittsburgh Press.

Fleishman, E. A. (1975). Toward a taxonomy of human performance. American Psychologist, 30, 1127-1149.

Fleishman, E. A., & Quaintance, M. K. (1984). Taxonomies of human performance: The description of human tasks. Orlando: Academic Press, Inc.

Folley, J. D., Jr. (1964). Development of an improved method of task analysis and beginnings of a theory of training. (NAVTRADEVCEN 1218-2). Port Washington, NY: USNTDC.

Friedman, L., & Harvey, R. J. (1986). Can raters with reduced job descriptive information provide accurate Position Analysis Questionnaire (PAQ) ratings? Personnel Psychology, 39, 779-789.

Gagne, R. M. (1974). The conditions of learning (3rd ed.). New York: Holt, Rinehart & Winston.

Guilford, J. P., & Hoepfner, R. (1971). The analysis of intelligence. New York: McGraw-Hill.

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

Guion, R. M. (1965b). Personnel testing. New York: McGraw-Hill.

Guion, R. M. (1976). Recruiting, selection, and job placement. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology (pp. 777-828). Chicago: Rand-McNally.

Gutenberg, R. L., Arvey, R. D., Osburn, H. G., & Jeanneret, P. R. (1983). The moderating effects of decision making/information processing job dimensions on test validities. Journal of Applied Psychology, 68, 602- 608.

Hakel, M. D. (1986). Personnel selection and placement. In Rosenzweig, M. R. & Porter, L. W. (Eds.), Annual review of psychology, 37, (pp. 351-389). Palo Alto, CA: Annual Reviews Inc.

Hackman, J. R., & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. Journal of Applied Psychology, 60, 159-170.

Hamilton, J. W., & Dickinson, T. L. (1987). Comparison of several procedures for generating J-Coefficients. Journal of Applied Psychology, 72, 49- 54.

Harvey, R. J., & Hayes, T. L. (1986). Monte carlo baselines for interrater reliability correlations using the Position Analysis Questionnaire (PAQ). Personnel Psychology, 39, 345-357.

Harvey, R. J., Friedman, L., Hakel, M. D., & Cornelius, E. T., III. (1987, April). Dimensionality of the Job Element Inventory (JEI), a simplified worker-oriented job analysis questionnaire. Paper presented at the Second Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta, GA.

Hazel, J. T., Madden, J. M. & Christal, R. E. (1964). Agreement between worker-supervisor descriptions of the worker's job. Journal of Industrial Psychology, 2, 71-79.

Hirsh, H. R., Schmidt, F. L., & Hunter, J. E. (1986). Estimation of employment validities by less experienced judges. Personnel Psychology, 39, 337-344.

Ivancevich, J. M. (1979). Longitudinal study of the effects of rater training on psychometric error in ratings. Journal of Applied Psychology, 64, 502-508.

Kintop, C. L., & Mussio, S. J. (1974). The validation of selection devices for high level clerical classifications: A demonstration study in the use of synthetic validity and work sample tests. Division of Personnel Research, City of Minneapolis, Civil Service Commission.

Kirkland v. New York Department of Correctional Service. 7 FEP 700 (1974).

Knerr, C. M., Miller, E. E., Hill, G. P., Nadler, L. B., Deppner, F. O., Dowell, S. K., & Somers, R. L. (1982). Interim report for manpower and personnel requirements determination methodologies (MANPERS). McLean, VA: General Research Corporation.

Latham, G. P., Wexley, K. N., & Pursell, E. D. (1975). Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.

Lawshe, C. H. (1952). Employee selection. Personnel Psychology, 5, 31-34.

Lawshe, C. H. (1953). Psychology of industrial relations. New York, NY: McGraw Hill.

Lawshe, C. H., & Steinberg, M. (1955). Studies in synthetic validity: An exploratory investigation of clerical jobs. Personnel Psychology, 8, 281-301.

Levine, J. M., & Teichner, W. H. (1973). Development of a taxonomy of human performance: An information-theoretic approach. JSAS Catalog of Selected Documents in Psychology, 2, 28, (Ms. No. 325).

McCormick, E. J. (1959). Application of job analysis to indirect validity. Personnel Psychology, 12, 402-413.

McCormick, E. J. (1979). Job analysis: Methods and applications. New York: American Management Association, AMACOM.

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

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1987, April). Project A validity results: The relationship between predictor and criterion domains. Paper presented at the Second Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta, GA.

Meyer, H. H. (1959). Comparison of foreman and general foreman conceptions of the foreman's job responsibility. Personnel Psychology, 12, 445-452.

Miller, E. E. (1969). A taxonomy of response processes (HumRRO Tech. Rep. 69-16). Fort Knox, KY: Human Resources Research Organization.

Miller, R. B. (1962). Task description and analysis. In R.M. Gagne (Ed.), Psychological principles in system development. New York: Holt, Rinehart & Winston.

Mossholder, K. W., & Arvey, R. D. (1984). Synthetic validity: A conceptual and comparative review. Journal of Applied Psychology, 69, 322-333.

Patrick, J., & Moore, A. K. (1985). Development and reliability of a job analysis technique. Journal of Occupational Psychology, 58, 149-158.

Peques v. Mississippi State Employment Service. 488 F. Supp. 239, 22 FEP Cases 392 (D.C. Miss. 1980), aff'd, rev'd, and remanded in part, 699 F.2d 760, 31 FEP Cases 257 (5th Cir. 1983).

Peterson, N. G. (1980). The development of a generalizability model of job analysis and its application to a study of the state patrol officer job using a self-administered checklist. Unpublished Doctoral Dissertation, University of Minnesota.

Peterson, N. G., Hough, L. M., Dunnette, M. D., Rosse, R. L., Houston, J. S., & Toquam, J. L. (1987, April). Identification of predictor constructs and development of new selection/classification tests. Paper presented at the Second Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta, GA.

Peterson, N. G., & Houston, J. S. (1980). The prediction of correctional officer job performance: Construct validation in an employment setting. Personnel Decisions Research Institute, Minneapolis, MN.

Peterson, N. G., Houston, J. S., Bosshardt, M. J., & Dunnette, M. D. (1977). A study of the correctional officer job at the Marion Correctional Institution, Marion, Ohio: Development of selection procedures, training recommendations and an exit information program. Minneapolis, MN: Personnel Decisions Research Institute.

Peterson, N. G., Rosse, R. L., & Houston, J. S. (1982). The job effectiveness prediction system. Technical report #4: Validity analysis. Personnel Decisions Research Institute, Minneapolis, MN.

Posner, M. I. (1962). An informational approach to thinking (Technical Report, Office of Research Administration Project 02814). Ann Arbor, MI: Office of Research Administration.

Prien, E. P. (1977). The function of job analysis in content validation. Personnel Psychology, 30, 167-174.

Primoff, E. S. (1955). Test selection by job analysis: The J-Coefficient, what it is, how it works (Test Technical Series, No. 20). Washington, DC: U.S. Civil Service Commission, Standards Division.

Primoff, E. S. (1957). The J-Coefficient approach to jobs and tests. Personnel Administration, 20, 31-49.

Primoff, E. S. (1959). Empirical validations of the J-Coefficient. Personnel Psychology, 12, 413-418.

Primoff, E. S. (1972). Preliminary draft: The J-coefficient procedure. Washington DC: U.S. Civil Service Commission, Standards Division.

Pulakos, E. P., Wise, L. L., Arabian, J., Heon, S., & Delaplane, S. K. (1988). A review of procedures for setting job performance standards. Washington, DC: American Institutes for Research.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.

Schmidt, F. L., & Hunter, J. E. (1984). A within-setting empirical test of the situational specificity hypothesis in personnel selection. Personnel Psychology, 37, 317-326.

Schmidt, F. L., Hunter, J. E., Croll, P.R., & McKenzie, R. C. (1983). Estimation of employment test validities by expert judgment. Journal of Applied Psychology, 68, 590-601.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. Journal of Applied Psychology, 66, 166-185.

Silverman, S. B., Wexley, K. N., & Johnson, J. C. (1984). The effects of age and job experience on employee responses to a structured job analysis questionnaire. Public Personnel Management Journal, 13, 355-359.

Smith, J. E, & Hakel, M. D. (1979). Convergence among data sources, response bias, and reliability and validity of a structured job analysis questionnaire. Personnel Psychology, 32, 677-692.

Smith, E. P., & Rossmeissl, P. G. (1986). Attribute assessment: Initial test of scales for determining human requirements of military jobs. Alexandria, VA: Army Research Institute, Selection and Classification Technical Area. (Draft)

Sternberg, R. J. (1977). Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Erlbaum.

Theologus, G. C., Romashko, T., & Fleishman, E. F. (1970). Development of a taxonomy of human performance: A feasibility study of ability dimensions for classifying human tasks. Washington, DC: American Institutes for Research.

Toquam, L. J., McHenry, J. J., Hanson, M. A., McGue, M. K., & Peterson, N. G. (1986). Reliability analyses of the J. C. Penney General Management Position Information Inventory. Minneapolis: Personnel Decisions Research Institute.

Trattner, M. H. (1982). Synthetic validity and its application to the Uniform Guidelines validation requirements. Personnel Psychology, 35, 383-397.

Uniform Guidelines on Employee Selection Procedures. (1978). Federal Register, 43, 38290-38315.

Urry, V. W. (1978). Some variations on derivations by Primoff and their extensions. Washington, DC: U.S. Civil Service Commission, Personnel Research and Development Center. (TN-78-3)

Wernimont, P. F. (1962). Re-evaluation of a weighted application blank for office personnel. Journal of Applied Psychology, 46, 417-419.

Willis, M. P. (1961). Deriving training device implications from learning theory principles (NAVTRADEVCEN 784-1). Port Washington, NY: USNTDC.

Wing, H., Peterson, N. G., & Hoffman, R. G. (1984, August). Expert judgments of predictor-criterion validity relationships. Paper presented at the annual convention of the American Psychological Association, Toronto.

Wise, L. L., Campbell, J. P., McHenry, J. J., & Hanser, L. R. (1986, August). A latent structure model of job performance factors. Paper presented at the annual convention of the American Psychological Association Convention, Washington, DC.

Wise, L. L., Campbell, J. P., & Arabian, J. M. (1987). The Army synthetic validation project. Paper presented at the Linking Issues Workshop organized by the National Academy of Sciences Committee on the Performance of Military Personnel, Sante Fe, NM.

Wise, L. L., Campbell, J. P., & Peterson, N. P. (1987, April). Identifying optimal predictor composites and testing for generalizability across jobs and performance constructs. Paper presented at the Second Annual Conference of the Society of Industrial and Organizational Psychology, Atlanta.

APPENDIX A

Hamilton and Dickinson (1987) present the following two equivalent J-coefficient formulas:

$$J_{xy} = \frac{\underline{B}^*{}_x{}'\underline{R}_{ye}}{(\underline{R}_{xe}{}'\underline{R}_{ee}{}^{-1}\underline{R}_{xe})^{1/2}(\underline{R}_{ye}{}'\underline{R}_{ee}{}^{-1}\underline{R}_{ye})^{1/2}}$$

or

$$J_{xy} = \frac{\underline{B}^*{}_y{}'\underline{R}_{xe}}{(\underline{R}_{xe}{}'\underline{R}_{ee}{}^{-1}\underline{R}_{xe})^{1/2}(\underline{R}_{ye}{}'\underline{R}_{ee}{}^{-1}\underline{R}_{ye})^{1/2}}$$

where

$\underline{B}^*{}_x{}'$ = row matrix of standard score regression coefficients from the regression of test performance on the job elements

$\underline{B}^*{}_y{}'$ = row matrix of standard score regression coefficients from the regression of job performance on the job elements

$\underline{R}_{xe}$ = column matrix of correlations between test performance and the elements

$\underline{R}_{ye}$ = column matrix of correlations between job performance and the elements

$\underline{R}_{ee}{}^{-1}$ = inverse matrix of correlations among job elements.

$\underline{R}_{xe}{}'$ = row matrix of correlations between test performance and the elements

$\underline{R}_{ye}{}'$ = row matrix of correlations between job performance and the elements

APPENDIX B: TASK CHARACTERISTICS APPROACH--MINOR MODELS

Additional task characteristics models reviewed but judged to be of only limited usefulness are briefly described and evaluated here. Two of these models are abstract in nature. One was developed for training purposes (Cotterman, 1959), the other for skilled tasks (Fitts, 1962). The Job Diagnostic Survey instrument (Hackman & Oldham, 1975) was developed for limited application in the area of job redesign.

Task Classificatory Scheme to Generalize Principles of Human Learning (Cotterman)

Cotterman (1959) proposed various methods for developing and evaluating a classificatory system based on abstract task categories. He hypothesized that three sets of variables operate in a learning environment. "Basic" variables are essentially learning principles, such as knowledge of results and previous learning. "Task" variables in Cotterman's scheme are the same as task variables elsewhere, e.g., stimulus and response population sizes. "Subject" variables are simply individual difference variables.

Cotterman suggested sorting a representative and large sampling of human learning tasks on the basis of similarities and differences on these three variables. The resulting categories should then contain tasks for which similar principles of learning can be generalized/applied. Learning principles would not be expected to generalize similarly to tasks sorted into different categories. To categorize a task adequately, a general description of the task stimuli, responses, and transformations (input, processes, and output) would be required. Two examples of learning situation/task categories are "Tracking or Continuous Adjustment" and "Skilled Act (Single Criterion Response)" (Fleishman & Quaintance, 1984, p. 170).

Cotterman's illustrative task categorization scheme has not been refined or tested. More developmental work must be accomplished before the reliability, validity, and utility of the system can be evaluated.

Two Classificatory Systems for Skilled Tasks (Fitts)

Two different task characteristic classificatory systems were developed by Fitts (1962). Both were based on the notion that performance on skilled tasks requires stimuli, feedback processes, and responses. Previous experience or skill, learning rate, and performance level were three individual difference variables of interest.

Two descriptors were used in the first model: (a) the degree of gross body involvement and (b) the extent of external pacing activity. Three task characteristic categories derived from these are based on the combinations of (a) body rest/motion status at task initiation and (b) stability/motion of the environment at task initiation.

In the second model, categorization was based on task: (a) stimulus-response sequences, (b) coding of stimulus, (c) coding of response, (d) code transformations, (e) stimulus type and amount, (f) external feedback type and amount, (g) internal feedback type and amount, (h) physical system dynamics, (i) overall task complexity. This model represents an analytical view of performer, performer-machine interaction, or performer-environment interaction closed loop systems (Fleishman & Quaintance, 1984).

Neither of these models holds much promise as the basis for a synthetic validation model. The level of the components is too general. In addition, components were developed specifically to apply to skilled tasks; not all Army job tasks are of this type.

Five Core Job Dimensions Measured by the Job Diagnostic Survey (JDS) (Hackman and Oldham)

The Job Diagnostic Survey (JDS) (Hackman & Oldham, 1975) provides measures of (a) skill variety, (b) task identity, (c) task significance, (d) autonomy, and (e) feedback present in a job. Jobs may be described and classified on the basis of these five characteristics. However, the main intended uses of the JDS are to diagnose jobs as an input to planned job redesign and to evaluate effects of job redesign. The JDS is based on the premise that presence of the five core job characteristics creates three critical psychological states: (a) experienced meaningfulness of the work, (b) experienced responsibility for outcomes of the work, and (c) knowledge of the actual results of the work activities. The three psychological states must be present simultaneously for positive personal and work outcomes (e.g., high internal work motivation, high quality work performance, high satisfaction with the work, and low absenteeism and turnover) to materialize. As the JDS is not relevant to the purposes of this project, it has not been evaluated relative to our criteria.

APPENDIX C: BEHAVIOR DESCRIPTION APPROACH--MINOR MODELS

Fleishman and Quaintance (1984) reviewed two behavior description models that we have designated as "minor" models. Willis' (1961) model was developed to investigate implications of learning principles and behavior classifications for training devices. Four broad task types were defined in a semantic approach based on worker-oriented verb descriptors (Bennett, 1971). A taxonomy containing components useful for at least one related Army job--Military Police (95B)--was developed in a study of the State Patrol Officer Job (Peterson, 1980). A large consortium synthetic validation study (Peterson et al., 1982) yielded a taxonomy composed of 46 performance areas and six factors. A taxonomy developed for the Correctional Officer position by Peterson and his colleagues (Peterson et al., 1977)--as part of a synthetic validation study--is also described. The final minor model reviewed is the British adaptation of the PAQ, called the Job Structure Profile (JSP) (Patrick & Moore, 1985).

Input-Output Hierarchical Model (Willis)

Willis' (1961) hierarchical behavior classification model is based on an input-output scheme. The model was developed to investigate implications of learning principles and behavior classifications for training devices (Fleishman & Quaintance, 1984). Descriptor type varies both within and between levels. Some descriptors are unobservable functions, while others are overt behaviors. At the most basic level are the components (a) receptor activity (input), (b) CNS activity (black box), and (c) effector activity (output). At the next level are six groupings of behavior descriptors, such as Skilled Motor Acts, Overt Verbalization, and Discrimination-nonverbal cues. Training implications were developed for the "cells" of Willis' model. The model's disadvantages include the abstract nature of the behavioral descriptors, lack of definitions for the descriptors, lack of evidence that analysts can reliably assign tasks to categories, and lack of formal evidence supporting validity of training device decisions based on the model.

Semantic Classificatory Approach (Bennett)

Rather than follow either the rational (e.g., Miller, 1969; Gagne, 1974) or empirical (e.g., Fleishman, 1975) route to development of a taxonomy, Bennett (1971) took an alternative approach. His approach was first to locate worker-oriented verbs from task analysis work. Consensus judgments about applicability of a narrowed set of these verbs to tasks were collected, then verb pairs were intercorrelated. Factor analyses of these judgments yielded four dimensions: (a) Cognitive--verbs generally related to ideas (Decide, Judge, Plan, Adjust, Analyze, Compute, Synthesize, Read, Interact with Data, Think); (b) Social--verbs related to communicating with people (Ask, Answer, Interact with People, Talk, Listen, Persuade), (c) Procedural--verbs corresponding to technology or use of equipment, including some type of action (e.g., Operate, Follow procedures, Use equipment, Handle, Do, Read), and (d) Physical--vigorous activity verbs (Carry, Walk, and Handle).

Reliability and validity evidence was unavailable. The components in the model are very general, most likely not technically adequate for rating military performance requirements or for describing military jobs. None of the components are motivational in nature. Army SMEs would have difficulty describing jobs using the limited set of four components. Even providing ratings with the more specific (25 originally paired) verbs subsumed by the components would prove to be difficult.

State Patrol Officer Job Taxonomy (Peterson)

Peterson (1980) derived eight functional job categories from 70 State Patrol Officer job tasks in an application of a generalizability model of job analysis (see Table C-1). In a pretest, patrol officers rated job tasks on importance and time spent, and psychologists rated worker characteristics (any KSAOs necessary for successful job performance) on usefulness for successful performance and the extent to which each worker characteristic had been developed during training or on the job.

Table C-1

State Patrol Officer Job Taxonomy[a]

	Job Components
1.	Traffic law enforcement
2.	Traffic control
3.	Routine and emergency assistance to motoring public
4.	Equipment maintenance
5.	Accident investigation
6.	Testifying and other court-related activities
7.	Peace-keeping, security, and detention
8.	Miscellaneous

[a]Adapted from Peterson, N. G. (1980). The development of a generalizability model of job analysis and its application to a study of the state patrol officer job using a self-administered checklist. Unpublished Doctoral Dissertation, University of Minnesota.

No motivational components were included; all were "can do" components. The model does contain components that would be useful for one Army job. All components would apply to the military police job

(MOS 95B). Army SMEs would probably not be able to completely describe enlisted job performance with this limited set of components.

Job Effectiveness Prediction System Study (Peterson, Rosse, and Houston)

The Job Activity and Performance Description Questionnaire (JAPDQ) was used in a synthetic validity study (Peterson et al., 1982). Specific steps in the study were: (a) identification of job elements, (b) identification of predictors correlated with performance in each element, and (c) formation of a composite prediction equation for overall job performance. (See discussion of the JEPS study in the Synthetic Validation chapter of this review, pp. 11-13.) The JAPDQ contained 46 performance areas or job components. The report from the study is proprietary, therefore, performance area components are not listed.

Interrater agreement estimates for "Not Part of Job" ratings of the 46 components averaged .64 for two raters. Mean interrater agreement coefficients for two raters for Time Spent and Importance ratings were respectable (.53). A sample of 214 raters completed the JAPDQ a second time, after a four-to-eight week interval (mean interval between completions was 45 days). This feature of the design allowed the computation of test-retest reliabilities, which ranged from .74 to .85. Validity of the JAPDQ was supported through its use in a synthetic validation study; details can be found in the synthetic validation chapter (pp. 11-13).

All or most of the performance areas were thought to be relevant to Army enlisted jobs involving clerical and some technical activities. This may be due to the fact that there were several items for each of the areas (a) simple calculations, (b) basic record keeping, (c) written communication, (d) typing, and (e) administration. One item among the 46--"Suggests, develops, or designs new ways of doing things or modifies the way materials or tasks are organized"--might be classified as a motivational component. The remainder of the components are of a "can do" nature. Judgments required of raters (e.g., time spent, importance) are straightforward. However, the model would not be perceived as wholly appropriate for all Army jobs because the model incompletely describes enlisted job performance.

Correctional Officer Job Taxonomy (Peterson, Houston, Bosshardt, and Dunnette).

Peterson and his colleagues (Peterson et al., 1977; Peterson & Houston, 1980) conducted a validation study of selection procedures for correctional officers. Correctional officers and inmates generated behavioral incidents to describe excellent, average, and poor job performance of correctional officers (Peterson et al., 1977). The 162 incidents were content analyzed by the researchers; the sorting yielded nine performance dimensions. The performance dimensions were named and defined before conducting retranslation workshops with SMEs. Using a criterion of 51 percent agreement for the total group of judges (correctional officer incumbents, supervisors, administrators, and inmates; n = 31), 93 of the 162 critical incidents were classified into one of the nine dimensions (shown in Table C-2).

Reliability and validity evidence was not available. Ability Requirements dimensions were also developed in this study. (They will be discussed separately in Appendix E).

The model was given a low rating in terms of its completeness for describing Army job performance (10%), as it is missing all technical components. However, components judged to be motivational (shown in Table C-2) are included in the model. Of these, Handling Unusual Situations and Consistency would be both useful for describing Army jobs and applicable beyond a prison setting. Although the judgments involved (time spent, importance, etc.) are not demanding, Army raters would not be able to completely describe job performance across jobs with this limited set of components.

Table C-2

Correctional Officer Job Taxonomy Dimensions[a]

* 1. Handling of unusual situations and crises
  2. Communicating with inmates
  3. Supervision of lock area
* 4. Attitude toward inmates
* 5. Use of force
* 6. Working with other officers
  7. Control of contraband
* 8. Dealing with intimidation and harassment
* 9. Consistency

[a]Adapted from Peterson, N. G., Houston, J. S., Bosshardt, M. J., & Dunnette, M. D. (1977). A study of the correctional officer job at the Marion Correctional Institution, Marion, Ohio: Development of selection procedures, training recommendations and an exit information program. Minneapolis, MN: Personnel Decisions Research Institute.

* Motivational component

Job Structure Profile (Patrick and Moore)

The Job Structure Profile (JSP) is the British adaptation of the Position Analysis Questionnaire (PAQ). Modifications were made both in wording ("anglicization") and content (additions to improve discriminatory power in the intellectual and decision-making domains) (Patrick & Moore, 1985). The resulting instrument contained 248 items and 23 factors; in comparison, the PAQ has 194 items and 28 factors. Unlike the PAQ, JSP items are rated on a single "Extent of Use" scale. Like the PAQ, the JSP contains six divisions. These are listed in Table C-3, with their subheadings.

A study of the JSP's reliability was conducted with nine jobs: clerical, secretarial, and managerial positions in a construction firm, and ordnance surveyor and sales management jobs in a parcel delivery company. The median intraclass correlation for the reliability of the mean of eight raters across nine jobs was .95 (ranging between .93 and .96). For a single rater, the corresponding median reliability was .71, ranging from .63 to .75. Adequate test-retest reliability was also demonstrated (average reliability by item for instrument as a whole was .76, with division reliabilities ranging between .67 and .86; average reliability by rater was .88, ranging from .66 to .97).

The model has a mixture of descriptors; some components within the six divisions seem to describe task characteristics rather than human behaviors, e.g., #17 Job demands, #18 Job responsibilities, #19 Organization of work and job structure, and #21 Physical hazards. No motivational components are included in the model. Those whose labels imply motivational aspects are composed of subcomponents defining task characteristics (e.g., Amount of..., Type of..., The opportunity for...). The model was rated as 90% complete for describing enlisted job performance. Although the model comprehensively describes jobs and job performance, it lacks motivational components.

Table C-3

Job Structure Profile Divisions and Factors[a]

Job Divisions and Factors
Division 1: Information Input
1. Sources of job information
- Visual sources of job information
- Non-visual sources of job information
2. Estimation activities
Division 2: Information Processing
3. Decision making and planning
4. Reasoning and problem solving
5. Information Processing
6. Memorization
Division 3: Work output--psychomotor activities
7. Use of devices and equipment
8. Manual activities
9. Activities of the entire body
10. Level of physical exertion
11. Body positions/posture
12. Manipulation/coordination activities
Division 4: Work output--Interpersonal job activities and job characteristics
13. Communication
14. Interpersonal job activities
15. The personal and social aspects of work
16. Interpersonal job characteristics
Division 5: Job related demands and the organization of work
17. Job demands
18. Job responsibilities
19. Organization of work and job structure
Division 6: Other job characteristics
20. The work environment
21. Physical hazards
22. Dress worn
23. Miscellaneous job characteristics

[a]Adapted from Patrick, J., & Moore, A. K. (1985). Development and reliability of a job analysis technique. Journal of Occupational Psychology, 58, 149-158.

APPENDIX D: BEHAVIOR REQUIREMENTS APPROACH--MINOR MODELS

Ten behavior requirements models classified as "minor" relative to the Army's purposes are described here. The rationales for their development differ greatly, including, for example, training and information processing.

Learning Process Categories (Gagne)

Gagne's model (1974) is based on learning process categories and on the performance objectives and performance outcomes related to them. The model's primary application is classification for instructional design purposes. The five main learning categories of the model are Cognitive Strategy, Verbal Information, Attitude, Motor Skills, and Intellectual Skills. Intellectual Skills has five learning subcategories within it: Problem Solving, Rule, Defined Concept, Concrete Concept, and Discrimination. It is unlikely that Army jobs would be categorized in terms of learning processes for the purposes of synthetic validation.

Although this model is of only limited usefulness for the Army's synthetic validation project, several of these learning categories are of a motivational nature. Attitudes--choosing actions toward people, objects, events, etc.,--"are not learned by practice nor are they affected by a meaningful context. However, they may be modified by human role models" (Fleishman & Quaintance, 1984, p. 137). Cognitive Strategies, defined as "originate novel problems and solutions" (Fleishman & Quaintance, 1984, p. 136) are oriented toward the self-management of learning and thinking, are organized internally, and are subject to refinement as learning occurs. The focus here renders the model inappropriate for use by the Army.

Definition of Training Needs in Terms of Activities (Folley)

Folley's training theory spawned a model which aids the development of military training devices and programs (McCormick, 1979). Tasks are described in terms of the degree to which they require each of five types of behavior: (a) Procedure Following, (b) Continuous Perceptual Motor Activity, (c) Monitoring, (d) Communicating, and (e) Decision Making and Problem Solving. The only component from this set which might be considered motivational in nature is Decision Making and Problem Solving, described within the model as "piecing together facts, opinions, and other information and arriving at a conclusion about what action to take" (McCormick, 1979, p. 195). After a job analyst has indicated whether or not each activity is required in task performance, he/she is required to make two additional judgments about each activity. One judgment is proportion of total task time for the activity, the other is extent to which the activity requires the performer's attention. Following Folley's approach, the procedures for analyzing tasks are combined with both his training theory and information about the task (e.g., timing/relationships among activities, task performance conditions). All of this information is then applied to the development of training programs and devices. The components in this model are too general for the Army's use.

Perceptual-Motor Task Classification System (E. E. Miller)

Miller (1969) categorized both tasks and training strategies in order to assess which training methods were most appropriate for various task types. The four task types were: (a) Reactive-adjustive tasks, which require responding to series of cues (where "there is an underlying continuum for the stimulus and response, and the response directly alters the stimulus dimension. Tracking tasks fall into this category"--Fleishman & Quaintance, 1984, p. 131), e.g., driving a boat; (b) Reactive-choice tasks, which require choosing from a set of appropriate responses, e.g., typing; (c) Developmental-procedural tasks, which require performing a fixed order of a series of steps, e.g., following a recipe; and (d) Developmental-skilled performance tasks, e.g., catching a ball. Miller did not empirically evaluate the task classification system. The perceptual categories within this scheme are too general for categorizing Army tasks for synthetic validation purposes. The model's applicability is limited because the categories were developed specifically for training purposes. The four categories described would be useful for the perceptual-motor aspects of Army job tasks, but Miller's system ignores the cognitive, motivational, and interpersonal aspects of tasks.

Use of Synthetic Tasks to Assess Complex Performance (Alluisi)

Alluisi (1967) hypothesized that a set of seven "performance functions" was essential to typical job tasks, and attempted to identify measures (tests or tasks) of these functions. These functions represent a combination of observable behaviors, covert behaviors, and internal processes. Performance functions and the tests (tasks) chosen to represent them are: (a) Watch keeping: warning-lights monitoring, blinking-lights monitoring, and probability monitoring; (b) Memory: arithmetic computations; (c) Sensory-perceptual: visual target-identification; (d) Procedural: code-lock solving; (e) Communication: (no measure yet developed); (f) Intellectual: (no measure yet developed); (g) Perceptual-motor: (no measure yet developed). The components were synthesized into multiple-task performance situations, characterized as "synthetic work" (Alluisi, 1967). The criteria for choosing tasks for performance functions were: (a) predictive validity, (b) face validity, (c) sensitivity to relevant performance changes, (d) engineering feasibility, (e) reliability, (f) flexibility, (g) work-load variability, (h) trainability, and (i) control-data availability. Alluisi (1967) concluded on the basis of his analyses that test-retest reliability for the performance tasks was adequate, and that the tasks were orthogonal. The drawback of Alluisi's approach is that since it was not developed for describing behavior under task conditions, the descriptors cannot be used for determining which functions are required for real or new tasks, or for quantifying differences among the required functions (Fleishman & Quaintance, 1984).

Information-Processing Tasks Taxonomy (Posner)

Posner (1962) uses only one dimension for categorizing all information-processing tasks. This dimension, the relationship between the amounts of information input and information output, further distinguishes tasks into three categories. The first is Information Conservation, in which amounts of information input and output are equivalent. This task type requires that the performer retain all stimulus information in order to respond. The second category is Information Reduction. In this case, the amount of input information is reduced in some way before it becomes output information. The task response requires a loss of information. Information Creation is the third category; the amount of output information exceeds the input amount. A performer must convert input into multiple responses. Posner's research focused on Information Reduction tasks. His work offers some general principles for task learning related to task and information variables. The components are too general and abstract for the Army's synthetic validation model; however, these components could prove useful if incorporated into a hierarchical model. In addition, not all activities/tasks across Army jobs are information-processing tasks per se.

Information-Processing Theory of Mental Abilities (Sternberg)

Sternberg (1977) has attempted to develop a taxonomy of task performance based on knowledge of determinants of performance--underlying levels of components and processes that influence performance. He has developed a theoretically-based task hierarchy, but it has not been empirically tested.

Mental abilities are categorized into four hierarchical levels: composite tasks, subtasks, information-processing components, and information-processing metacomponents (Fleishman & Quaintance, 1984). Further, this information-processing theory represents a behavioral requirements approach; it focuses on organizing tasks by the processes, components, etc. that they require. From the performer's viewpoint, the composite task is the first level specified by the theory. Tasks that can be quantified in terms of response and error rates are used to investigate mental abilities and task performance. Examples of such tasks are: linear, categorical, and conditioned syllogisms; analogies; metaphorical completions and ratings; series completions; and classifications. In addition to meeting the quantifiability criterion, these tasks also were chosen because they were (a) reliable (reliability estimates across tasks and individuals are approximately .90 or above), (b) construct valid, and (c) empirically valid measures of the cognitive abilities they were developed to measure (Fleishman & Quaintance, 1984).

Tasks are further broken down into subtasks, the second level of the hierarchy. Subtasks involve a subset of the information-processing components involved at the task level. The third level of the task hierarchy is the component level; components are formed by decomposing subtasks. Each component is a primary information process that somehow translates sensory input to internal representation or internal representation to motor output, or transforms one internal representation into another. Components are of three types: (a) general--components required to perform all tasks for a certain domain, (b) class or group--components required to perform task classes within the domain, and (c) specific--components needed to perform only one task from the domain. The most important type is the class component category. Sternberg (1977) suggested the possibility of arranging hierarchically those tasks that are used to measure mental abilities. The class components used to perform each task determine the placement of those tasks within this hierarchy. Although class components used to perform tasks at a given level may vary, all tasks at that given level are of equal complexity. The levels of the hierarchy differ in complexity, with higher hierarchy levels being more complex. (This hierarchy directs task selection for research studies of the theory.) The fourth level in the theory is metacomponents, the control processes that influence decision-making, planning, and monitoring processes. These control processes are of a motivational nature according to the Army's view of motivational or "will do" components.

Although Sternberg's research may increase our understanding of the structure and content of mental abilities, this work is not yet advanced enough to offer useful components for the Army synthetic validity model. The information-processing theory base is perhaps not compatible with the Army job performance model. The range of tasks included in this type of approach is also too narrow to cover all Army jobs.

Criterion Measures Approach (Fleishman & Quaintance)

One of the models developed as part of the Taxonomy Project (Fleishman & Quaintance, 1984), the Criterion Measures Approach, is classified as a behavior requirements approach. However, the functions required for task performance are categorized in a behavior descriptive fashion. Based on transfer of information between interacting performer-machine units of a system, this model identifies four main task classes: machine-performer, performer-machine, machine-machine, and performer-performer. A task represents the information transmitted between these classes, and a process operates on information within a class. Four varieties of information transfer exist: (a) searching, (b) switching, (c) coding, and (d) tracking. Each type of information transfer applies to one of the four main task classes. Dependent measures, which the investigators also referred to as "descriptive measures," were used to operationally define the information transfer types. Task activities, with their "descriptive measure" (dependent measure) are given in Table D-1.

The Criterion Measures Approach was evaluated by categorizing one area of literature--learning and environmental effects--into the four task activity categories. A criterion suggested for evaluating work performance taxonomies was their ability to categorize and account for a body of literature (Fleishman & Quaintance, 1984). The Taxonomy Project

Table D-1

Criterion Measures Approach: Task Activities[a]

Criterion Measures
1) Searching: Receptors are seeking or orienting to signal sources at different positions/times. Examples - monitoring radar screen, seeking target Descriptive measure - Probability of detection
2) Switching: A discrete action changes the status of the succeeding component in a system. Example - turning light on and off Descriptive measure - Reaction time or latency
3) Coding: A detected signal is named/identified. (3 types: a. Simple - naming stimulus characteristics b. Group - grouping stimulus characteristics according to one classification c. Successive - using rules to translate or recode Descriptive measure - Percentage correct
4) Tracking: The alignment of a response with an input is maintained. Examples - steering car, aiming weapon Descriptive measure - time on target (or percentage decrement in time on target)

[a]Adapted from Fleishman, E. A., & Quaintance, M. K. (1984). Taxonomies of human performance: The description of human tasks. Orlando: Academic Press, Inc.

had developed a human performance database that contained evidence of associations between the same independent variables (effects of noise intensities, knowledge of results, and different practice schedules) and performance. Results indicated that these categories were useful for organizing literature in the practice schedule (distributed vs. massed) area, based on ease of applying the categories, and was found to hold for all areas the authors investigated. Fleishman and Quaintance (1984) further conclude, "A primary finding was that the system made it possible to organize the literature on distributed practice in terms of 1) functional relationships and 2) the effects of different functions for different task categories" (p. 238). Results for investigations with knowledge of results and noise effects were not as clear-cut as those for practice schedules. Overall, however, this approach has helped to summarize, focus, and direct diverse studies on conditions affecting task performance.

In terms of reliability of task assignment to categories, the Criterion Measures Approach has not yet been tested. Only certain task types have been investigated; the approach should be extended to use with tasks of a perceptual, motor, and cognitive nature. The model is not applicable to the Army's synthetic validation project, due to its intended purpose and focus.

Information-Theoretic Approach (Levine & Teichner)

The Information-Theoretic Approach, also known as the Systems Language Approach, was deductively derived from an information-processing model (Levine & Teichner, 1973). Fleishman and Quaintance (1984) summarized its development, terms, and applications, and evaluated its classification adequacy; the following summary reflects the main points.

A task is defined as an information transfer through an information channel (source to receiver). Processes are more general than tasks, and operate on information. Human tasks are not distinguished from machine tasks, and a systems language common to all tasks is used to break down any process into a series of more specific, intervening tasks. Tasks are categorized on the basis of four dimensions, as follows: (a) Nature of constraints--as a requirement of the theory, internal (how events are sampled) and external (what events are sampled) constraints that are imposed on the stimulus and response events must be specified. (b) Location of constraint--two types of constraints operate on input, while three types of constraints operate on output (e.g., imposed by task, by performer, or by situational characteristics). (c) Redundancy--constraints imposed on both the input and output create redundancies; these operate to enhance, degrade, or not affect information transmission. (d) Input-Output Information Relationships--after task completion, comparison between quantities of information input and output results in categorization as (1) information conservation, (2) information reduction, or (3) information creation. Input or output constraints are postulated to account for inequalities.

Within-same-category tasks have more in common than tasks which fall into different categories. The logic of the model is summarized as follows: "information transfer or task performance within a given category, defined by the nature and location of the constraints, is dependent on the amount of redundancy and the relationship between input-output information" (Fleishman & Quaintance, 1984, p. 253). A plan was formulated for evaluating the model in a dual (empirical and theoretical) process; steps outlined were: (a) conceptualization of constraint classes, (b) development of a laboratory task for testing input and output specifications of the model, (c) design of experiments for controlling constraints, (d) development of computer programs to simulate constraint classes and manipulate redundancies, and (e) matching of computer simulations and laboratory experimental results. This evaluation has not yet been conducted. Again, however, the information-processing model of task performance may not be compatible with the Army model of performance; it may be too specific to one type of task for synthetic validation purposes.

Elemental Motions (Drewes)

Drewes (1961) developed a series of dexterity tests (the Purdue Elemental Motions Tests) based on the Methods-Time Measurement (MTM) system of predetermined times. Four elemental motions were studied because they were assumed present in almost all assembly jobs, and also overlap with motion elements for putting pins into a board. These four motions are Reach, Grasp, Position, and Release of object. The other three components of the MTM system are Turn, Move, and Disengage. The model was rated as not at all useful for describing Army jobs, because it is specific to elemental motion tasks. In addition, it was judged to be potentially very difficult for Army SMEs to use in describing jobs or job performance requirements.

Task Strategies Approach (R. B. Miller)

Task Analysis, defined as a process that attempts to abstract behavioral implications from behavior description information, was the starting point for Miller's (1962) Task Strategies Approach. These behavioral requirements could be identified by analyzing intervening variables, functions, and processes in a Stimulus--Organismic Variables--Response--Feedback (S-O-R-F) model.

Tasks were defined as series of information-processing transactions occurring both between the operator and environment and within the operator. In this model, descriptors are functions/activities of an operator, performed to meet task demands. Miller named, without empirical evidence, subfunctions for two of the four system functions--reception of task information, retention of task information, interpretation and decision making, and motor response mechanisms (more generally: input reception, memory, processing, and output effectors).

Miller used four dimensions for categorizing tasks: (a) naming of task functions, (b) description of task content (subject matter), (c) identification of task environment (both internal--physiological and psychological--stressors on the operator and external conditions due to simultaneity of tasks), and (d) identification of level of learning dictated by the task, with respect to amount of practice. The approach was named for its focus on task strategies (for successful performance) related to the first dimension, task functions. Miller's work has a practical application; he developed a systems task vocabulary useful for systems analysis. The development of task functions to represent intervening processes was a basis for his identification of work strategies--behavior and task strategies--associated with more effective/efficient task performance.

Evaluation of this model has focused on the lack of precision and quantification of behavioral functions. In addition, low reliability was found in an attempt to assign tasks to categories by other researchers. These problems were addressed in later efforts to refine the model.

In reference to criteria for evaluation, the refined and expanded Task Strategies Approach still has some disadvantages. Its reliability has not been investigated, but is presumed improved over that of the earlier, less developed model. Categories within the system are not all mutually exclusive, as Miller attempted to fully describe human task performance transactions at the expense of redundancy. Some external validity evidence was shown by the association between task functions and work strategies. With its work strategies, the approach has potential utility, but has not been widely used.

APPENDIX E: ABILITY REQUIREMENTS APPROACH--MINOR MODELS

Three additional ability requirements models are reviewed here. Their relevance for the Synthetic Validation Project will be highlighted. A model developed on the basis of job analysis of the Correctional Officer job (Peterson et al., 1977) demonstrated concern for inclusion of motivational factors. A number of job components useful for describing jobs requiring broadly-defined abilities of a clerical nature are contained within a taxonomy developed for government jobs (Kintop and Mussio, 1974). A factorial research program in the cognitive ability area produced a comprehensive three-dimensional model of intellectual abilities (Guilford & Hoepfner, 1971).

Correctional Officer Job Taxonomy (Peterson, Houston, Bosshardt, and Dunnette)

Peterson and his colleagues conducted a four-phase job analysis as part of a Correctional Officer (CO) selection validation study (Peterson et al., 1977). The first phase consisted of semi-structured interviews with incumbent COs, CO supervisors, and prison staff to gather preliminary information about the CO job and worker characteristics. In Phase 2, structured interviews were conducted using a group seminar format to cover detailed aspects of job responsibilities, KSACs (C = personal characteristics), and unusual tasks, assignments, and problems. The third phase was the development and review of an inventory covering work context, job tasks, and worker characteristics. During Phase 4, the inventory containing 106 job tasks and 48 worker characteristics was administered to staff members (n=86). Four types of ratings were collected: (1) relative time spent on tasks, (2) importance of tasks for successful job performance, (3) usefulness of KSACs for successful job performance and (4) amount of each KSAC developed after starting the job. The researchers examined mean usefulness ratings for the 48 worker characteristics and combined those that appeared conceptually similar into the 18 factors listed in Table E-1.

Interrater agreement for 25 job experts' ratings for the 18 factors ranged from .72 through .96, indicating satisfactory reliability levels. Validity of the model was addressed via a construct validation approach. Psychologists (n = 10) rated the extent of measurement of 20 CO constructs by 51 tests; a set of optimum scales was identified. Job experts rated importance of CO constructs for successful job performance of CO job tasks. The psychologists' and job experts' ratings were then used to develop a numerical weighting scheme for using the instrument scores to predict a job applicant's likely standing on each CO construct and overall job performance. The authors concluded that both sets of judges were able to reach high levels of consensus concerning instrument-to-construct and construct-to-job task linkages. An inventory was developed (the Marion Correctional Officer Psychological Inventory--MCOPI) to be used as part of the applicant clinical screening process, with the benefit of providing information regarding likelihood of successful job performance.

The model was rated as relatively incomplete for rating performance across Army jobs, as it is missing technical skill components. However, 11 motivational components were included; these are indicated in Table E-1. Army job experts would be able to provide the required ratings. However, they would also find the model illogical for describing Army jobs, as all the necessary components are not included. The components of this model could be linked to Army predictors in the same manner as components and tests were linked in the actual study (see preceding paragraph).

Table E-1

Correctional Officer Job Taxonomy[a]

Critical Worker Characteristic Combinations or Factors
* 1. Knowledges and skills relevant for dealing with violent inmates (individuals or collective)
2. Basic "job routine" knowledge
3. Institution security and contraband detection knowledges
4. Basic communications skills
5. Equipment use skills
* 6. Physical condition
* 7. Maintaining co-worker relationships
* 8. Supervision and application of discipline
9. Dealing with emergencies or tense situations
10. Communicating, counseling and advising
*11. Tolerance of and appropriate reactions to ongoing stress
*12. Consistency and fairness
*13. Leadership
*14. Adaptability
*15. Dependability
*16. General attitude toward inmates
*17. Use of force
*18. Emotional stability

[a]Adapted from Peterson, N. G., Houston, J. S., Bosshardt, M. J., & Dunnette, M. D. (1977). A study of the correctional officer job at the Marion Correctional Institution, Marion, Ohio: Development of selection procedures, training recommendations and an exit information program. Minneapolis, MN: Personnel Decisions Research Institute.

* Motivational component.

High-level Clerical Job Taxonomy (Kintop & Mussio)

In a demonstration of the synthetic validation approach, Kintop and Mussio (1974) identified 29 knowledges and skills common to one or more of 12 high-level government clerical jobs. SMEs generated a listing of all possible tasks and behaviors in this set of related jobs, and job analysts identified these 29 skills and abilities as underlying performance of tasks and behaviors. The investigators' objective was to develop work-sample tests for the set of broad-based job components identified for these clerical jobs. Based on the decision that all of their "knowledge of" elements were inherent in the "ability to" elements, the authors limited the set of components to 15 abilities for which performance tests could be developed (see Table E-2).

Table E-2

Higher-Level Clerical Job Taxonomy[a]

Job Components (Ability to:)
1. Communicate (in written form)
2. Determine what and how much information is needed
3. Organize and process work in the most efficient way possible
4. Formulate, evaluate, and implement new or revised work methods
5. Plan, organize, and maintain a record-keeping system
6. Compare information
7. Code information
8. Copy information
9. Compute information
10. Compile information
11. Proofread, double-check, and/or correct information
12. Type
13. Gather, organize, and present data in a logical format
14. Follow written and oral instructions
15. Set up and keep financial records

[a]Adapted from Kintop, C. L., & Mussio, S. J. (1974). The validation of selection devices for high level clerical classifications: A demonstration study in the use of synthetic validity and work sample tests. Division of Personnel Research, City of Minneapolis, Civil Service Commission.

This model, as an ability requirements model, is based on abilities as descriptors. Thus, it contains components of a "can do," rather than a "will do" nature. These job components would be very useful for describing Army clerical jobs. In addition, some components would apply to higher-level jobs that require planning, organizing, etc. However, none of the general Army job components is tapped by this taxonomy. Army SMEs would not have difficulty using these components. Because these components are specific, they are more concrete than components commonly found in more general purpose ability requirements models.

Structure of Intellect Model (Guilford)

Guilford's model of intellectual abilities describes tasks in terms of an information-processing approach (Guilford & Hoepfner, 1971). The three broad dimensions of the model are (a) Contents, the information inputs or discriminable types of information, (b) Operations, the processes operating on information inputs, and (c) Products, the outputs or forms that information takes as a result of processing it. Each of these three broad dimensions incorporates several categories. The five Content categories, five Operations categories, and six Products categories are listed in Table E-3. Guilford identified 150 hypothetical abilities related to intelligence by plotting all combinations of the categories (e.g., as cells within a cube). Each cell then describes covert behaviors, and each behavior is identified by an operation performed on some information that yields a product. The existence of a majority of these hypothetical factors or abilities has been demonstrated in factor-analytic studies (Guilford & Hoepfner, 1971). Tests have been developed to measure the demonstrated and defined ability factors.

The model can be related to performance. Abilities required for performing a certain job can be identified by administering a battery of the reference tests identified for ability factors in the model. The model represents a very comprehensive system for describing intellectual functions. However, Army SMEs would find the system foreign, and have difficulty describing all tasks in terms of these intelligence-related abilities.

Table E-3

Guilford's Structure of Intellect Model[a]

Ability Factors
Dimensions/Categories:
Contents
1. Visual
2. Auditory
3. Symbolic
4. Semantic
5. Behavioral
Operations
1. Evaluation
2. Convergent production
3. Divergent production
4. Memory
5. Cognition
Products
1. Units
2. Classes
3. Relations
4. Systems
5. Transformations
6. Implications

[a]Adapted from *Guilford, J. P., & Hoepfner, R. (1971). The analysis of intelligence. New York: McGraw-Hill.