

Capturing Utility Judgments Across Jobs:
Toward Understanding and Generalization

by

Laura F. Donnelly

dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
Psychology

APPROVED:

Philip Bobko, Chairperson

Eugene F. Stone

Shanta P. Kerkar

Robert C. Williges

Frederick S. Hills

September, 1985
Blacksburg, Virginia

CAPTURING UTILITY JUDGMENTS ACROSS JOBS:
TOWARD UNDERSTANDING AND GENERALIZATION

by

Laura F. Donnelly

(ABSTRACT)

The recent increase in utility research has provided improved methods for estimating the standard deviation of performance in dollars. Subjective estimates of an individual's overall worth to the organization allow the utility of various organizational interventions to be evaluated. However, this research does little to illuminate the dimensions underlying supervisory judgments of utility.

The present research explored the predictability of utility judgments from job-level information. A policy capturing approach was used to select the best subset of predictors, from a set of nine variables hypothesized to be related to judged utility. Results of regression analyses across five hypothetical performance levels indicated that six of the nine variables were most important for predicting utility values. These variables included both organizationally specific measures and subjective ratings of hypothetical constructs.

The policies underlying judgments of overall worth were captured to a substantial degree, with cross-validated R^2

values ranging from .46 to .69. A unit weighting scheme was applied to the six predictors, resulting in $\underline{r^2}$ values that exceeded the cross-validated $\underline{R^2}$ derived from regression analyses. This substantial predictability of utility judgments provided the capacity to generalize utility information from a sample of jobs to the population of interest.

Analyses comparing validity-based and utility-based clustering schemes explored the degree of convergence between the two approaches to classifying jobs. These analyses indicated that there was some overlap, with validity information being useful in establishing broad categories of jobs associated with similar utility-relevant attributes. At the same time, these analyses demonstrated that the two approaches were not equivalent.

Implications of this research are discussed, and several possible directions for future research are noted. It is suggested that such policy capturing procedures can enhance our understanding of judgments of overall worth, and expand the knowledge base upon which organizational decisions are made.

mother, , and brother, , who have provided continual love and support. In addition, my friends , , and helped in many ways. Most importantly, I suppose, they made me laugh when I needed it most. A final note of thanks goes to my husband, , who has been a wonderful source of strength throughout my years in graduate school. Without his advice, assurance and understanding, this dream could never have been realized.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
 <u>Chapter</u>	
	<u>page</u>
INTRODUCTION	1
Overview	1
Understanding Utility Estimates	6
Research Setting	8
An Overview of Policy Capturing Research	11
Relevant Empirical Research	13
Potential Predictors of Utility	15
Traditional Job Characteristics Model	16
Perrow's Model of Process Technology	18
Job Evaluation Research	19
Overview of Validity Generalization Research	21
Overview of the Present Research	23
METHOD	25
MOS Population and Sample	25
Criterion	26
Predictors	28
Variables obtained through Army sources	28
Additional variables not available through Army sources	34
Comparison of Utility-Based and Validity-Based Clusters	41
RESULTS	43
Predictors	43
Correlations with utility	45
Predictor Selection and Policy Capturing	48
Unit-Weighted Analysis	52
Factor Analysis	53
Cluster Comparisons	57
DISCUSSION	61
The Predictor Domain	61
Generalizability	66
Cluster Comparisons	67

Implications for Industrial/Organizational Psychology	68
Limitations of this Research and Future Directions	74
FOOTNOTES	79
REFERENCES	81

Appendix

	<u>page</u>
A. MATERIALS PROVIDED TO FIELD GRADE OFFICERS FOR EVALUATING THE UTILITY OF 19 MOS	91
B. SAMPLE JOB DESCRIPTIONS PROVIDED TO GRADUATE STUDENTS WHEN RATING JOB CHARACTERISTICS	96
VITA	99

LIST OF TABLES

<u>Table</u>	<u>page</u>
1. Mean Utility Values for 19 MOS	29
2. Means and Standard Deviations of Predictors in the Sample and the Population	44
3. Intercorrelation Matrix for Predictors within the Sample of 19 MOS	46
4. Average Correlations between each Predictor and Utility ^{1, 2}	47
5. Results of Regression Analyses by Performance Level (n = 19)	49
6. Results of Unit-Weighted Analysis using Six Predictors	54
7. Factor Loading Matrix for Principal Components Factor Analysis	56
8. Contingency Table of Frequencies for Comparison of Nine Validity and Nine Utility Clusters	59

INTRODUCTION

OVERVIEW

In recent years, there has been a resurgence of interest in evaluating the utility of human resource programs. Utility analysis allows organizations to determine the effectiveness, or value, of any intervention (e.g., an improved selection or classification system). The criterion for "effectiveness/value" may be expressed in metrics relevant to any desired goal, including increased productivity (quantity and quality of performance) or organizational profit (dollars saved). Two factors have jointly spurred the trend toward evaluating utility: 1) an increased awareness of the notion that effective utilization of employee resources has substantial impact on organizational functioning (cf. Cascio & Silbey, 1979); and 2) an increase in the availability of methods for estimating the parameters necessary for utility analyses (cf. Schmidt, Hunter, McKenzie & Muldrow, 1979).

Prior to the development of utility analysis methodology, traditional cost-accounting techniques were used to determine the cost effectiveness of organizational decisions. These methods, however, were seldom applied because they can be very costly and time-consuming. As an

alternative to cost-accounting techniques, Brogden (1949) and Cronbach and Gleser (1965) introduced equations for evaluating the utility of selection devices. These equations describe utility as a direct, multiplicative function of the selection test validity (\underline{r}), selection ratio, and the standard deviation of employee performance (\underline{SD}_y) in the metric of interest. This resulting utility is then compared to the cost of the selection procedure (C). Thus, the gain in utility at the individual level of analysis is:

$$\Delta U = (\underline{r} \times \underline{SD}_y \times \bar{Z}_x) - C \quad (\text{Equation 1})$$

where ΔU refers to the change in utility, as compared to random selection, and \bar{Z}_x is the mean standardized predictor score for selected applicants (which is directly related to the selection ratio). At the organizational level, the gain for the entire organization, across persons and time, may be expressed as:

$$\Delta U = N \times T \times (\underline{rSD}_y \bar{Z}_x) - C \quad (\text{Equation 2})$$

where N = number of employees selected with the new procedure, and T = average tenure of selectees.

Whereas the Brogden (1949) and Cronbach-Gleser (1965) approach simplifies the algebraic determination of utility, the parameter \underline{SD}_y is still quite difficult to estimate. Recently, Schmidt, Hunter, McKenzie and Muldrow (1979)

proposed a simple, inexpensive technique by which supervisors indirectly estimate the standard deviation of performance worth. With this procedure, supervisors are asked to estimate the dollar-value of employee performance at the 15th, 50th, and 85th percentiles. Assuming that performance worth follows a normal distribution, the differences between the 50th and 15th and between the 85th and 50th percentiles provide two estimates of \underline{SD}_y . The mean value of these estimates (averaged across supervisors) serves as the estimate of \underline{SD}_y in the subsequent utility analysis. Bobko, Karren and Parkington (1983) evaluated the accuracy of this procedure. Their study examined supervisory estimates of total yearly sales for insurance counselors. They found that supervisors using this technique provided very accurate estimates for the "sales" component of performance, as compared to actual, archival data. It should be noted that this procedure is only one of several methods for estimating \underline{SD}_y (Bobko, in press).

Although substantial progress has been made in developing methods for estimating the value of performance, little is known concerning the dimensions underlying the judgments of such values. That is, what are the correlates of people's estimates of overall utility? Recent studies in utility analysis have demonstrated that dollar value of the

standard deviation of performance is related to employees' annual wage (Hunter & Schmidt, 1982). Hunter and Schmidt (1983) suggested that 40% of mean salary may provide an adequate approximation of the standard deviation of performance in dollars.

The finding that salary paid to employees may be an index of the value of performance simply suggests that salary may be a suitable surrogate for the concept of performance goals. This need not always be the case, however, particularly for non-profit and service-oriented organizations with primary goals other than profit-maximization. In fact, Eaton, Wing and Mitchell (1985) have argued that traditional salary-based performance estimates may be inappropriate in many contexts. With this in mind, they developed two alternative methods for estimating the standard deviation of performance which may be more appropriate in situations where supervisors are accustomed to thinking in terms of the relative productivity of employees or where employees operate very complex, expensive equipment and are focal to the productivity of a costly system. Their "Superior Equivalents Technique" focuses on supervisory evaluations of the relative performance of employees. It involves obtaining estimates of the standard deviation of performance in terms of the number of superior

performers needed to produce the output of a fixed number of average performers. These "performance units" can then be converted to dollars if such metrics are desired. Their second alternative is the "System Effectiveness Technique" which is based upon changes in aggregate system performance. Total aggregate performance of a system is determined by the number of units comprising the system (e.g., number of employees or machines) as well as the performance level of these units. In this approach, the value of improved individual unit performance in achieving higher system performance is the cost of the increased number of units needed to achieve the higher aggregate performance.

Eaton et al.'s (1985) work highlights two important points. First, it suggests that an additional step may be required to convert SD_y information into monetary units. Utility estimates provided by job experts may not always be best expressed in terms of dollars. In particular, use of salary as an index of utility involves the assumption that the linkage between salary and criticality of jobs to goal attainment is substantial. The assumption of such a relationship may not always be justified. Further, even if there were a one-to-one correspondence between salary and true worth to an organization, our understanding of why various jobs are associated with different relative

utilities is not advanced by such knowledge. Eaton et al. (1985) argued that supervisors may develop and use informal algorithms as a basis for providing utility information. Thus, we are left with the question of what factors are considered by judges when such estimates are made.

UNDERSTANDING UTILITY ESTIMATES

A major goal of the present research is to investigate the degree to which utility estimates across jobs can be assessed (with a policy capturing approach) using job-level data as predictors. To the extent that the policy capturing attempt is successful, this research should help to illuminate the types of dimensions which underlie judgments of utility. A second major goal of this research builds upon information obtained in relation to the first goal. It involves investigating the question, "Can the captured policy derived from a sample of jobs be successfully applied to other jobs in a given population?" Consideration of this question quickly gives rise to a new concept: utility generalization.

Utility generalization as a concept is analogous to that of validity generalization (cf. Schmidt & Hunter, 1977; Schmidt, Hunter, Pearlman & Shane, 1979). Application of validity generalization techniques allegedly reduces the

need for empirical criterion-related validity studies, by taking advantage of results from relevant studies already conducted. The concept of validity generalization (discussed in more detail below) is based on the assumption that some jobs have certain test-related characteristics/requirements in common. Then, similar jobs would be associated with similar predictor-criterion relationships. Similarly, emergence of the concept of "utility generalization" depends on the notion that there are certain characteristics of jobs, related to their organizational value, which can discriminate among jobs. The "similar" jobs could be grouped together on the basis of utility, or overall worth, to the organization. In turn, knowledge of the types of jobs which are most valuable can allow organizations to focus their limited resources on improving performance in these jobs through, for example, enhanced training and development efforts.

No conceptual framework currently exists which can provide a basis for generalizing utility information. One temptation could be to proceed by using job clusters which are based on validity generalization principles, such as similarity of required abilities. However, validity and utility are conceptually distinct (Cronbach & Gleser, 1965), implying that the factors which make jobs similar in terms

of validity need not be the same factors which make jobs similar in terms of utility. This distinction becomes clear upon examination of Equations 1 and 2 (page 2). That is, the relationship between predictor and criterion (r) is included as only one component of utility. Thus, tests with equivalent validity in two organizations may be associated with very different utilities if the standard deviation of performance and/or the selection ratio (and therefore, the mean standardized predictor score) varies across these organizations.

RESEARCH SETTING

The U. S. Army is currently conducting a \$22.8 million, 9-year study, an ultimate goal of which is to develop more valid and cost-effective selection and classification measures. To the extent that this goal is achieved, the new system will enable an improved "matching" of applicants to jobs (these job types are known as Military Occupational Specialties, or MOS). The net result of this system will be more effective utilization of the personnel available to the Army. Because such an undertaking is enormously complex and time-consuming, the Army has chosen to develop the system through intensive study of a sample of 19 (out of approximately 250) entry-level MOS. However, the system is

intended to be Army-wide, requiring that the Army be able to generalize findings from the 19 MOS to the much larger population of 250 job types. This setting presents an excellent opportunity to investigate the decision processes underlying estimates of utility. Further, knowledge gained about such judgments can then be applied toward development of techniques which will allow generalization of utility information.

One crucial aspect of the Army project involves an assessment of how valuable varying levels of performance, both within and across MOS, actually are to the Army. For example, they are interested in comparing the worth of a below-average infantryman to that of an average or above-average infantryman (within-job) as well as comparing an average infantryman to an average clerk (across-job). This assessment will allow Army decision makers to evaluate critical tradeoffs of varying personnel classification schemes. For example, if the Army finds that superior performance in one MOS is less valuable than even poor performance in another MOS, their efforts will best be spent in predicting and maximizing performance in the more valuable MOS. For purposes of manpower allocation, it is important to know the worth of MOS performance so that assignments can be made which will increase the probability

of obtaining some important goal (e.g., potential lives saved). Therefore, the Army has collected information on the utility of varying performance levels within MOS.

The process of capturing and generalizing utility information in this research is made more difficult because the units in which utility is measured in the Army project are different from those which are typically used. As noted above, most recent work on utility analysis has involved obtaining judgments in monetary metrics (with the notable exception of Eaton et al., 1985). The Army, however, is not a profit-making organization and its wage system may not accurately reflect market values. Thus, "worth to the Army" is not simply a financial concept, but one that suggests the capacity to help achieve the Army's non-profit-oriented goals. The present research, unlike most prior research, should be of relevance to the many organizations having primary goals other than profit-maximization (cf. Schneider, 1980, for an extended discussion concerning the inappropriateness of strictly financial considerations for evaluating human effectiveness in service organizations).

The use of policy capturing within the above context represents an attempt to apply decision-making research to a current organizational problem. This approach will allow a systematic, reliable and valid means of extending judges'

decision policies beyond the MOS which were actually evaluated. In fact, in the research plan for the major Army project, a need for research-based procedures was emphasized as follows:

...present selection and classification procedures could be improved by taking advantage of recent technological advances and developments in decision theory. There is a need for developing a formal decision-making procedure that is aimed at maximizing the overall utility of the classification outcomes to the Army (Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, and Army Research Institute, 1983a, p. 1/5).

In sum, this research builds upon several theoretical foundations. It represents an interconnection of research on policy capturing, utility analysis, and validity generalization. We now turn to overviews of these areas, and their relationships to the current research.

AN OVERVIEW OF POLICY CAPTURING RESEARCH

Policy capturing is a strategy which analyzes the statistical relationship between information cues and actual decisions, ultimately providing mathematical descriptions of the judgmental policies that are used. The attempt in this research will be to use a policy capturing approach to obtain what Hoffman (1960) referred to as a "paramorphic" representation of the judgment process. This term suggests

that the psychological processes underlying a decision can be simulated by weighting predictors (also called input variables or cues) in a regression model.

The power of simple linear models to predict judgments is a consistent theme in the literature. Judges, however, often report that they combine information in complex ways. Hoffman, Slovic and Rorer (1968) used an Analysis of Variance approach, which uncovered a number of statistically significant interactive (configural) effects. This result would seem to support introspective reports from decision makers indicating that they use complex combinatorial rules in making judgments. However, Hoffman et al. (1968), as well as other researchers (e.g., Goldberg, 1968; Slovic & Lichtenstein, 1971), note that the increment in predictive power provided by these configural effects is extremely slight. It should be noted that even though interaction effects may yield very small contributions to predictive accuracy, the practical significance of these small gains may warrant their inclusion in some situations (Hoffman et al., 1968).

Use of the linear model in the present research will guide predictor selection efforts. That is, the primary focus of using the policy capturing paradigm in this research is to identify those job-level predictors which

consistently relate to utility estimates. The policy capturing approach is well-suited to the task of selecting a subset of variables which best predicts human judgments. Dawes (1979) noted that the human's strength in information processing lies in creatively selecting and coding relevant information, rather than in integrating different cues. Policy capturing is therefore an excellent means by which to select from a set of potential variables those which have the greatest influence on the criterion (cf. Christal, 1968).

Relevant Empirical Research

In a field study, Taylor and Wilsted (1974) captured the performance appraisal judgment policies of officers rating U.S. Air Force Academy cadets. Although raters evaluated ten individual performance factors, their overall ratings were adequately described as a linear combination of three cues. Taylor and Wilsted (1974) noted that there was high multicollinearity among the ten cues, and suggested that the remaining seven variables were essentially redundant with the chosen three. An additional finding of interest was that use of a more complex nonlinear model (including interaction terms) did not increase predictive ability.

Another policy capturing application of relevance to the present study is Robinson, Wahlstrom and Mecham's (1974) use of a multiple regression strategy to predict compensation rates. As discussed previously, compensation is often considered to be an index of utility. These researchers selected and weighted dimensions obtained from job analyses, using the Position Analysis Questionnaire, of 131 jobs in a utility company. This approach resulted in a sample multiple correlation of .90 ($R^2 = .81$). The company adopted this predictive model as a less costly method of job evaluation which, at the same time, perpetuated the philosophy of the job evaluation committee.

Finally, Brady and Rappoport (1973) employed a three-step research strategy to capture the underlying policies from both industry and Atomic Energy Commission (AEC) judgments about appropriate solutions to the problem of nuclear safeguards. Briefly, their strategy involved: a) development of an interview guide which probed various specific elements, indicated by experts as important to an effective safeguards system; b) cluster analysis of interview responses, resulting in groupings of specific elements into more comprehensive cue dimensions; and c) linear regression analyses which weighted and combined cue variables, yielding a policy equation describing industry

judgments and an equation describing AEC judgments. The nature of their study resembles the present research since, in both cases, cue dimensions have to be determined, with little a priori knowledge of which variables are considered by decision makers. That is, decision makers are not presented with a controlled set of stimuli and asked to make their judgments from these data alone. In such situations, the researcher must, by some means, "construct" a set of cues which potentially influence judgments.

POTENTIAL PREDICTORS OF UTILITY

Since the dimensions upon which judgments were based in the present project were not specified a priori, an important research step involved identification of influential information cues.

As discussed above, the recent increase in utility research has not focused on development of predictive models of utility. Improved methods for estimating the standard deviation of performance in dollars have been developed, but this research does little to illuminate the dimensions underlying supervisory judgments of utility (cf. Burke & Frederick, 1984; Karren & Bobko, 1983). As noted previously, the Army's primary focus on increasing utility is not profit-oriented. Therefore, the search for

predictors in this research should extend beyond strictly monetary factors. Interviews with Army Research Institute personnel indicated that job characteristics such as "probability of engaging in combat" or "minimal required level of general aptitude" may be important predictors of judged utility. The Army has conducted considerable research on such factors, establishing specific scales and guidelines applicable to the entire range of MOS. Thus, Army records and documentation represent a valuable source of information in this research. Overviews of literatures relevant to other potential predictors are presented below.

Traditional Job Characteristics Model

The non-monetary nature of the variables being considered pointed to the "job characteristics" model as a potentially useful framework for predictor identification. The traditional model of job characteristics suggests that five core dimensions of jobs--level of skill variety, task identity, task significance, autonomy, and feedback--constitute "job scope" and are related to certain psychological states experienced by the worker. In turn, these psychological states are hypothesized to lead to desirable work outcomes (Hackman & Oldham, 1975, 1976).

Stone (in press) presents a meta-analytic review of research on job scope-job satisfaction and job scope-job performance relationships. He concludes that in both laboratory- and field-based studies, job scope variables are strongly related to individuals' affective responses about their work and their jobs in general. However, the relationship between job scope variables and job performance criteria is inconsistent across laboratory and field studies and is considerably weaker than the relationship for satisfaction outcomes. It therefore appears that the model's link between job characteristics and personal psychological states is strongly supported, whereas the hypothesized relationship between job characteristics and performance is not. Unfortunately, the lack of correspondence between job characteristics and job performance criteria implies that analysis in terms of these characteristics will also not be very useful for identifying predictors of job utility.

It should be noted, however, that an extension of the job characteristics model may be useful in the present research. Stone and Gueutal (1985) used a multidimensional scaling procedure to identify three dimensions underlying perceptions of jobs. These three dimensions were labeled "job complexity" (representing a summary of the components

evaluated in the traditional job characteristics model), "serves the public", and "physical demands". The "job complexity" dimension is similar to a variable used in this research, referred to as "variety" (which is discussed in the next section). Also, the "physical demands" characteristic seemed especially relevant in the Army context, and was included as a potential predictor of judged utility. In fact, this dimension was measured through two means, both in terms of a strength requirements measure and a broader measure of the physical demands of jobs.

Perrow's Model of Process Technology

Perrow (1970) developed a "degree of nonroutineness" typology which is used as a model for measuring the concept of process technology. Two dimensions of work activities which are relevant to organizational structure and process comprise this typology. The first dimension is the number of exceptions which are encountered in the work. As such, it indexes task variety, or the frequency of unexpected, novel events which occur in performance of job duties. The second dimension involves the analyzability of work activities, with processes which are well-understood and standardized being highly analyzable. Conversely, processes which are not well-understood, and in which techniques

and/or procedures for handling problems are not available, are low on analyzability.

Typically, Perrow's framework is used to conceptualize and measure the level of process technology within a given department of an organization. However, the nature of work performed in different jobs (within as well as across departments) may vary on these dimensions just as the primary task of departmental units may vary. Thus, the use of Perrow's framework in the present research represents an extension from the measurement of technology within departmental units to applying these dimensions at the level of single jobs.

Job Evaluation Research

The objective of job evaluation is to establish a job structure using systematic and consistent rules which appraise each job's relative value as compared to other jobs in the organization. As such, it attempts to develop a system which is viewed by employees as fair from an "internal equity" point of view. The two quantitative methods of job evaluation, the point-factor and factor comparison approaches, are most relevant to the present research because they evaluate job worth by breaking jobs down into component parts. In each of these approaches,

evaluators determine a set of specific factors in terms of which jobs should be analyzed.

Examination of the factors considered in several job evaluation systems reveals substantial convergence across specific plans. This set of "core" factors includes: physical demands and/or working conditions, supervisory controls, complexity of the work, and skills required (e.g., education, experience, initiative, mechanical ability). An additional factor emerged from consideration of traditional job evaluation practice which represented a fascinating and potentially powerful concept. In the present research, this concept is referred to as "cost or consequence of error" in the performance of job duties. Possibly expressed in terms of "accountability" (e.g., in the Hay method, Bellak, 1984), "responsibility for safety of oneself (or others)" (cf. Morris, 1973; Patten, 1977), or "impact of error" (Karr & Fischbach, 1984), some form of this dimension is frequently included in factor comparison and point-factor job evaluation systems. The "cost of error" variable is thus conceptualized in the present research as measuring the seriousness of consequences which are likely to result from poor job performance.

OVERVIEW OF VALIDITY GENERALIZATION RESEARCH

Conceptually, the nature of this research problem is similar to that underlying the development of validity generalization techniques. Heightened legal pressure on organizations (e.g., from the Equal Employment Opportunity Commission) to justify their use of specific selection aids was a major impetus for the recent increased emphasis on determining the validity of selection tests. However, along with the increased desire to assess the validity of tests came a desire for the capability to transport or generalize empirically-derived test validities from a given job or job type to other jobs. This motivation has led to a great deal of research, under the rubric of "validity generalization" (e.g., Schmidt & Hunter, 1977; Schmidt, Hunter, Pearlman & Shane, 1979).

Research on validity generalization has seriously questioned the assumption that employment test validities are situation-specific. As evidence against situational specificity, Schmidt and Hunter (1977) identified seven potential sources of artifactual, between-study variance in observed validity coefficients. Further, they demonstrated that three of these seven sources of variance accounted for about 50% of the observed variance in four distributions of validity coefficients. Schmidt, Hunter, Pearlman and

Shane's (1979) study strengthened these conclusions, suggesting that aptitude and ability test validities can be generalized beyond the specific test-job combinations that are empirically evaluated.

Various researchers have proposed modified validity generalization procedures (e. g., Callender & Osburn, 1980; Raju & Burke, 1983), all of which provide results that support the viability of the validity generalization concept (Burke, 1984). Schmidt, Hunter and Pearlman (1981) demonstrated that task differences among jobs do not moderate the validity of aptitude tests in selection. They suggested application of general job analysis methods which permit grouping jobs in terms of broad content structure or inferred underlying abilities. Such job-similarity groupings result in the identification of job clusters, within which validity can be generalized.

Similarly, the present investigation of "utility generalization" represents an attempt to generalize utility information from a sample of jobs to the relevant population. As with validity generalization, the first step involves identifying those characteristics which underlie judgments of utility. Only after predictors of utility have been identified is it possible to investigate the generalizability of utility results. Thus, the viability of

a concept of "utility generalization", analogous to that of validity generalization, will be explored. Further, although validity and utility are related, Cronbach and Gleser (1965) note that they are distinct concepts. Based on this distinction, patterns of job clusterings resulting from a validity generalization approach should be different from those based on application of a utility generalization methodology.

OVERVIEW OF THE PRESENT RESEARCH

In sum, this research is conducted with two objectives in mind. The major goal is to investigate the degree to which subjective estimates of job utility can be captured, using a multiple regression strategy with job-level data as predictors. Thus, a policy capturing approach is used to determine the subset of job characteristics, from a larger pool of potential variables, which best predicts utility judgments. To the extent that utility values can be accurately predicted, examination of the predictive model should improve our understanding of the types of factors which underlie estimates of utility. Secondly, provided that predictability of utility is high, the resultant model can be used to generalize utility information to a larger set of jobs. Further, the degree of convergence between

clusters of jobs established for validity generalization purposes and those arising from consideration of utility information is explored.

METHOD

MOS POPULATION AND SAMPLE

The population considered for the present research is a representative subset of the Army's 238 entry-level jobs (MOS). Rosse, Borman, Campbell and Osborn (1983) clustered a set of 111 MOS on the basis of job-content similarity. These 111 MOS included all large MOS (annual minimum of 300 new job incumbents for the fiscal year 1983-84 period), and smaller MOS, such that each Career Management Field was proportionately represented by at least one-third of its MOS. For the present research, nine of the 111 MOS were eliminated due to missing data on one or more of the crucial predictor variables. In fact, the Army has since rescinded or deleted seven of these nine MOS. Important predictor information was not available for the other two MOS. Thus, the relevant population in this study consists of 102 entry-level MOS.

The sample of 19 MOS chosen for in-depth study was selected to be representative of the entry-level MOS population. The sample was chosen on the basis of several criteria, including (a) sufficient sample size within MOS; (b) representative coverage of the Armed Services Vocational Aptitude Battery (ASVAB) aptitude areas; (c) representation

of high priority MOS (those essential in a national emergency); and (d) representative coverage of the Army's Career Management Fields (Campbell, 1983).

Analysis of the sample reveals that 19 of the Army's 30 Career Management Fields, and all nine aptitude composite clusters were represented. Although the number of MOS seems small, these 19 MOS account for approximately 45% of all Army accessions -- or, approximately 68,000 recruits per year (Eaton, Goer, & Zook, 1984). Further, 44% of 1981 women recruits were represented in the sample; 44% of 1981 black recruits were represented; and 43% of 1981 Hispanic recruits were represented. The absolute level of female and minority representation in the sample paralleled that of the population: 15% female, 27% black, and 5% Hispanic (Human Resources Research Organization et al., 1983b). Thus, MOS included in the study sample are fairly representative of the MOS population.

CRITERION

Utility value judgments made by 12 field grade officers (i.e., majors, lieutenant colonels, and colonels) served as the criterion in this study. This information was collected for the ongoing Army project described earlier. Field grade officers were chosen as expert judges for two reasons: a)

the study thereby incorporated the goals/values of senior military leaders; and b) field grade officers, unlike higher-ranking officers, had sufficient knowledge of the day-to-day activities involved in various MOS.

The estimation procedure used to obtain criterion data was substantially pilot-tested in seven workshops conducted in both the United States and Germany over a six-month period. Officers estimated utility for five hypothetical levels of performance (performance at the 10th, 30th, 50th, 70th and 90th percentiles) within each of the 19 MOS being studied. Specifically, they evaluated each of the 95 (19 MOS X 5 levels) MOS/performance level combinations by means of allocating points, using a 90th percentile infantryman with a standardized worth of 100 points as an anchor. The standard, therefore, served simply as a reference point, and no other restrictions were imposed on the rating scale. This standard was chosen because previously-conducted workshops revealed that officers found the infantryman/90th percentile to be a common referent--and hence, an easy standard from which to anchor their ratings.

For the rating task, officers were presented with a wartime scenario¹ and a brief job description for each MOS. They were asked to assign points indicative of MOS/performance level worth, relative to the 100-point

standard, in that wartime context. The scenario, instruction set, and sample descriptions which were used are provided in Appendix A. The median interrater reliability was .80. The Army subsequently averaged these point estimates across the 12 raters, yielding 95 mean criterion values. The reliability of this mean criterion vector was over .95 (Sadacca & Campbell, 1985).

The scale values provided by field grade officers were transformed, such that the value of an average (50th percentile) infantryman (MOS 11B) was set equal to a scale value of 1.0. The resultant utility values are presented in Table 1. Because these estimates were made on a ratio scale, straightforward interpretations of the relative values indicated by these numbers are possible. For example, Table 1 shows that an "Armor Crewman" performing at the 90th percentile (utility = 1.60) is twice as valuable as a 90th percentile "Carpentry and Masonry Specialist" (utility = .80).

PREDICTORS

Variables obtained through Army sources

A crucial step in this research involved identification of potential predictors of utility. Although published utility literature highlights the strong relationship

TABLE 1

Mean Utility Values for 19 MOS by Performance Level
(n = 12 Officers)

MOS	Performance Level (Percentile)				
	10	30	50	70	90
Administrative Specialist (71L)	-.07	.29	.47	.74	.86
Ammunition Specialist (55B)	.12	.46	.69	.90	1.13
Cannon Crewman (13B)	.30	.69	.93	1.24	1.49
Carpentry & Masonry Specialist (51B)	.00	.09	.37	.61	.80
Chemical Operations Specialist (54E)	.20	.53	.86	1.16	1.38
Combat Engineer (12B)	.20	.65	.96	1.22	1.52
Food Service Specialist (94B)	.09	.33	.59	.83	1.04
Infantryman (11B)	.29	.71	1.00	1.30	1.58
Light Wheel Veh./Pwr. Gen. Mech. (63B)	.17	.51	.68	1.02	1.24
Armor Crewman (19F)	.40	.68	1.03	1.26	1.60
Manpads Crewman (16S)	.19	.57	.83	1.09	1.38
Medical Specialist (91B)	.17	.48	.79	1.07	1.37
Military Police (95B)	.15	.47	.71	.97	1.20
Motor Trans. Operator (64C)	.06	.39	.59	.83	.97
Petrol. Supply Specialist (76W)	.16	.51	.72	.82	1.11
Radio teletype Operator (05C)	.13	.54	.77	1.09	1.30
TOW/Dragon Rep. (27F)	.10	.53	.74	.99	1.33
Unit Supply Specialist (76Y)	.08	.40	.60	.91	1.07
Util. Heli. Rep. (67N)	.15	.49	.82	1.06	1.32

between salary and utility indices (e.g., Bobko et al., 1983; Hunter & Schmidt, 1982), discussions with Army Research Institute (ARI) personnel pointed to non-monetary characteristics as possible predictors. Based on these discussions, reviews of Army documentation, and review of the general research area of "job characteristics" (e.g., Hackman & Oldham, 1975, 1976; Stone, in press; Stone & Gueutal, 1985), a preliminary set of possible variables was determined. Availability of this information was investigated, and ultimately, data for the MOS population were obtained through Army sources on seven variables. These data were at the job level of analysis, such that all five performance levels within an MOS were associated with the same value on a given predictor variable. These seven predictors included three composite variables, formed a priori by unit weighting the standardized component variables. The variables for which data were collected are described below (variable abbreviations are shown in parentheses), accompanied by a brief rationale for their inclusion in this study.

- COMBAT PROBABILITY (TCOMBAT) -- the a priori standardized sum of a dichotomous variable, open/closed (CLOSED) to women (Department of the Army, 1984) and a three-level Army classification indicating the relationship of each

job to actual combat involvement (COMBAT). The point-biserial correlation between these two indices, CLOSED and COMBAT, was .82. Rationale: It was reasoned that the perceived value of Army jobs would increase as their functions became more central to combat operations, since interviews with field grade officers indicate that this is one of the Army's primary missions.

- TRAINING (TTRNG) -- the a priori standardized sum of training length (TRLGTH), measured in weeks, and an estimate of the average training cost, per incumbent, in a given MOS (TRCOST) (U.S. Army Finance & Accounting Center, 1983). The correlation between these two indices, TRLGTH and TRCOST, was .85. Rationale: This variable was considered because it includes factors such as cost of equipment used by job incumbents and amount of preparation required for performance of job duties. To some degree, it represents the Army's "investment" in soldiers. It was expected that the most valuable MOS would be those associated with larger training investments.
- ENLISTMENT BONUS (TBONUS) -- the a priori standardized sum of a dichotomous variable indicating MOS eligibility for the Army College Fund (ACF), and the amount (in dollars) offered for enlisting in the MOS (BONUS) (U. S.

Army Recruiting Command, 1984). The point-biserial correlation between these two indices, ACF and BONUS, was .64. Rationale: This composite represented the inducements offered individuals at the time of enlistment. It was expected that MOS which are judged to be more valuable to the Army would be those for which more desirable enlistment bonus packages were offered.

- STRENGTH REQUIREMENTS (STRENGTH) -- an Army classification scheme which indexes the physical strength requirements of MOS, with five levels ranging from light to very heavy (Department of the Army, 1984). This variable relates to the "physical demand" job dimension reported by Stone and Gueutal (1985). Rationale: Given that MOS included in this study are entry-level jobs, it was reasoned that MOS associated with high utility values would, in general, be those requiring substantial physical effort and strength. Note that for higher-level jobs, the relationship between physical demands and utility would be expected to be weaker or actually become negative.
- SELECTIVE REENLISTMENT BONUS (SRB) -- the average bonus paid to soldiers within an MOS upon reenlisting (based on Fiscal Year 1983 figures) (Department of the Army, 1982). Rationale: This variable was included because it

represents what the Army is willing to pay to retain soldiers in various MOS. It reflects the degree to which the Army has difficulty recruiting qualified individuals as well as the desirability of trained, experienced job incumbents both to the Army and in the private sector.

- APTITUDE -- the cutoff score on one of the nine Armed Services Vocational Aptitude Battery (ASVAB) composites used as a requisite for entry into an MOS. These cutoffs vary from 80 to 110 (McLaughlin, Rossmeissl, Wise, Brandt & Wang, 1984). Rationale: It was hypothesized that, in general, MOS having more stringent ASVAB entrance requirements would tend to be those viewed as more valuable to the Army.
- PROMOTABILITY (PROMOT) -- the mean length of time (in years) taken for soldiers in a given MOS to reach the E-4 level; based on data for fiscal year 1981 (Department of the Army, 1983). The E-4 level was chosen because it represents the transition from private to specialist or noncommissioned officer status, and is thus a significant promotion for first-tour enlistees. Rationale: It was expected that promotions would come sooner for those MOS associated with high utility values, with promotions representing one means by which the Army may retain individuals in these jobs.

Each of the above predictors was proposed because it reflected a variable of interest which was available from Army sources. Although each predictor was discussed separately, it is not being suggested that these dimensions are orthogonal to one another. Even a cursory examination will reveal that there may be substantial covariance between certain variables. For example, since the combat probability concept is so fundamental to Army operations, substantial correlations between TCOMBAT and other predictors should not be surprising. Quite simply, the Army aggressively recruits (TBONUS), attempts to train (TTRNG) in a quick and cost-effective manner, and tries to retain (SRB) individuals in combat-related jobs.

As indicated above, data for these seven predictors came from information sources within the Army system (e.g., Technical Reports and Army manuals). The value associated with each variable for a given MOS is determined by Army standards and/or practice.

Additional variables not available through Army sources

In order to enhance the predictor domain, a qualitative review of the 19 job descriptions was undertaken. The goal of this review was to identify additional factors which varied across jobs, and might be related to judged utility.

Ratings of these dimensions for each MOS could then serve as predictor data. This analysis led to the identification of several conceptually-defined dimensions. Framing these dimensions in terms of constructs which have received research attention suggested that four rating dimensions could be worthy of study in this research.

The first two variables identified in this manner were denoted as "variety" and "analyzability", the two dimensions considered in Perrow's (1970) "degree of non-routineness" typology (see prior discussion). "Variety" was considered as a potential predictor on the premise that MOS involving substantial problem resolution, with job demands varying across situations, would tend to be judged as more valuable to the Army. The concept of "analyzability" appeared useful in that jobs requiring exercise of one's judgment and application of experience-based knowledge in problem resolution would seem to be those of highest value to the Army.

The third variable, "cost/consequence of errors", was derived from the literature on job evaluation. It seemed particularly relevant in the Army context because there is tremendous variability in the consequences of error across Army MOS. For example, the probable consequence of poor performance in the "Food Service Specialist" position (e.g.,

soldiers missing a meal) is substantially less severe than that for the position of "Cannon Crewman" (e.g., serious injury or loss of life). Thus, it was hypothesized that MOS associated with higher utility values would be those for which the consequences of poor performance would have the greatest, most severe impact.

The final variable, physical demands, was adapted from Stone and Gueutal's (1985) multidimensional scaling research. This dimension was included because it encompasses several aspects of the physical component of jobs. As such, it could be used as a complement to the strength requirements information provided by Army sources (see previous section).

Seven-point rating scales were developed which probed the degree to which these dimensions were present in each of the Army jobs. For example, the rating scale for "cost/consequence of errors" ranged from 1 ("Extremely low cost of errors; consequences are minimal") to 7 ("Extremely high cost of errors; consequences are extremely serious"). Descriptions of these variables are provided below, exactly as they were presented in the rating task. Again, variable abbreviations are shown in parentheses.

- TASK VARIETY (VARIETY) -- "This dimension refers to the frequency with which unexpected exceptions or novel

events are encountered in performance of job duties and tasks. It is an index of task predictability. Thus, when workers encounter a large number of unexpected situations, with frequent problems, task variety is high. However, when there are few problems, and when day-to-day requirements are repetitious, variety is low."

- ANALYZABILITY OF WORK ACTIVITIES AND PROBLEMS (ANALYZ) -- "This dimension refers to the degree of difficulty involved in identifying correct solutions to problems which are encountered in the work. It indexes the extent to which cause-effect relationships underlying problems are understood. For example, if the work has been reduced to a number of standardized steps, and established guidelines can be used to identify solutions to various problems in the work, analyzability is high. However, if problems encountered are unique and appropriate courses of action for problem resolution are difficult to determine, analyzability is low."
- COST/CONSEQUENCE OF ERRORS (ERRCOST) -- "This dimension refers to the severity of consequences resulting from poor job performance. It can be measured in terms of monetary losses, disruption of organizational operations, safety of oneself or others, or in any other relevant terms. It involves not only immediate consequences but

also the long-term consequences of errors and the degree to which such errors can or cannot be remedied."

- PHYSICAL DEMANDS (PHYSDEM) -- "This dimension indexes the degree to which a job requires physical strength, activity and movement. It also encompasses factors such as cleanliness of the job, level of incumbent health hazards, and responsibility for equipment. Thus, a physically non-taxing, clean and safe job would be "low" on this dimension. On the other hand, a strenuous, dirty and hazardous job would be considered high on physical demands."

Six raters (advanced graduate students in Industrial/Organizational Psychology) evaluated each of the 19 MOS in terms of the four dimensions described above. Raters were provided with the 19 job descriptions, the dimension descriptions (as presented above), and the four rating scales. Job descriptions came from Army Regulation 611-201 (Department of the Army, 1984), which provides descriptions of all Army MOS. Maximum length of these job descriptions was 750 words, with most between 300 and 500 words. Two sample descriptions are provided in Appendix B. The order of presentation of the job descriptions was randomized. Two orders of presentation for the rating dimensions were constructed, with three raters using one order and three

raters using the other. The four dimension ratings were made after reading each job description. As indicated previously, the physical demands variable was included in the rating procedure to supplement information on strength requirements obtained through the Army system. The correlation between these two indices, STRENGTH and PHYSDEM, was .56. Thus, a final a priori composite representing "working conditions" (WORKCON) was constructed, composed of physical requirements information from the Army classification and that obtained through the rating procedure.

The intraclass correlation coefficient estimates of interrater reliability (Shrout & Fleiss, 1979) for the four rating scales were: task variety--.57; analyzability--.15; cost/consequence of errors--.49; physical demands--.56. Due to the low reliability for analyzability, this variable was eliminated. Using the Spearman-Brown formula, the reliabilities for the mean rating profiles of the three remaining dimensions were: task variety--.89; cost/consequence of errors--.85; physical demands--.89.

The same six raters analyzed the remaining MOS (n=83) in the population according to the variety and cost of error dimensions using the procedure described above. Only these dimensions were included because regression analyses of the

19 MOS (see below) suggested that they were the only rating variables having unique explanatory variance. Further, use of two dimensions rather than the initial four made the rating task less onerous for raters. The intraclass correlation coefficients based on these 83 MOS indicated the following interrater reliabilities: task variety -- .26; cost/consequences of errors -- .48. This resulted in Spearman-Brown estimates of reliabilities for the mean ratings of: task variety -- .68; cost/consequence of errors -- .85.

In sum, the following nine predictor variables were considered in this study:

- TCOMBAT (CLOSED + COMBAT); a composite representing likelihood of engaging in combat
- TTRNG (TRLGTH + TRCOST); a composite of Army expenditures on soldier training
- TBONUS (ACF + BONUS); a composite representing inducements offered for enlistment
- WORKCON (STRENGTH + PHYSDEM); a composite of the physical demands or working conditions of MOS
- SRB; selective reenlistment bonus
- APTITUDE; the cutoff score on the Armed Services Vocational Aptitude Battery required for entry into an MOS

- PROMOT; the mean length of time taken for soldiers to reach the E-4 level
- VARIETY; ratings of the frequency of unexpected or novel events encountered in the work
- ERRCOST; ratings of the severity of consequences resulting from poor job performance

COMPARISON OF UTILITY-BASED AND VALIDITY-BASED CLUSTERS

The Army currently uses nine aptitude composite validity clusters to aid in selection and classification. These nine clusters were formed by differentially weighting the ten cognitive subtests of the Armed Services Vocational Aptitude Battery. Of the 102 MOS in the population of interest, 100 are associated with one of these nine composites (the specific composites appropriate for the remaining two MOS could not be identified). Thus, these groupings can be considered "validity clusters". Alternatively, MOS could be grouped together based on similarity of their judged (or predicted) utility.

Cronbach and Gleser (1965) point out that validity and utility are related, yet distinct concepts. Equations 1 and 2 (p. 2) make this relationship explicit, indicating that validity is only one component necessary to the evaluation of utility. Since information on both validity and utility

are available in this study, the degree of similarity between job clusters based on test validity information and those emerging from utility considerations can be investigated.

As a supplemental analysis, the major goal of this comparison was to examine the degree of convergence and highlight any salient differences between the validity and utility clustering schemes. By definition, validity and utility are not equivalent; this analysis sought to demonstrate their nonequivalence. A comprehensive and detailed examination of specific similarities and differences was beyond the scope of this research. Each MOS was coded according to its associated aptitude cluster, with these nine groupings representing the validity clustering approach. Utility clusters of MOS were formed on the basis of similarities in the pattern of values for the identified components of estimated utility. Such an approach to "utility generalization" was appealing because of its clear analog in validity generalization. That is, a validity generalization approach holds that certain jobs are "similar" due to the pattern of abilities required to perform them. Validity can be generalized within these clusters of similar jobs. Similarly, cluster analysis based on utility components should suggest groupings of jobs which are similar in terms of utility.

RESULTS

PREDICTORS

Table 2 presents means and standard deviations of the predictor variables (separate data are provided for variables entering into composites) for both the sample of 19 MOS and the total population of 102 MOS. It is clear from Table 2 that sample and population values for most predictor variables are comparable. The primary consideration in examining Table 2 concerns the degree to which the standard deviations of variables in the sample are comparable to those in the population, since variance differences may affect the magnitude of correlation coefficients. Comparisons of sample and population standard deviations for both variables entering into the training composite (TRLGTH and TRCOST) suggested that sample values for the training measures were range-restricted.

Specifically, the sample standard deviation for TRLGTH was 2.14, whereas it was 9.19 in the population. Similarly, the standard deviation for TRCOST was 8767 in the sample, but 16,844 in the population. Range restriction corrections to correlations involving these variables were considered.

It is crucial to note, however, that the correlations, per se, were not the parameters of interest. Rather, the

TABLE 2

Means and Standard Deviations of Predictors in the Sample
and the Population

Variable	<u>Sample Data</u>			<u>Population Data</u>		
	<u>n</u>	Mean	<u>SD</u>	<u>N</u>	Mean	<u>SD</u>
CLOSED	19	1.26	0.45	102	1.18	0.38
COMBAT	19	2.00	0.75	102	1.95	0.64
ACF	19	0.74	0.45	102	0.53	0.50
BONUS	19	2394.74	2306.87	102	1607.84	2152.17
STRENGTH	19	4.53	0.96	98	4.02	1.28
PHYSDEM	19	4.38	1.27	not collected		
TRLGTH	19	15.21	2.14	95	20.32	9.19
TRCOST	19	14585.58	8767.01	95	22474.24	16844.61
SRB	19	2.35	2.97	102	2.61	3.28
APTITUDE	19	89.74	5.65	102	94.07	7.37
PROMOT	19	1.80	0.07	100	1.78	0.11
VARIETY	19	3.47	1.23	102	3.01	0.91
ERRCOST	19	4.63	1.22	102	3.72	1.18

focus was on regression coefficients in the policy capturing analyses, not correlations. Furthermore, the policy capturing analyses were conducted using unstandardized regression analysis. This results in parameter estimates which are unaffected by range restriction² (Bobko & Schemmer, 1980).

Table 3 provides intercorrelations among the nine independent variables for the sample data.

CORRELATIONS WITH UTILITY

The average within-performance level correlations between each predictor and the criterion, judged utility, are presented in Table 4. These values were obtained by converting the five simple-order correlations (i.e., at each performance level) between a given predictor and utility to the corresponding Fisher's Z values, and then converting the average Fisher's Z value back to a correlation. Note that the correlations tend to be high, with seven of the nine correlation coefficients exceeding .40. In particular, the correlations for TCOMBAT and ERRCOST are quite high, being .720 and .718, respectively.

Each of the average correlations was tested for significance by treating the average as a mean of five normal variables (i.e., Fisher's Z 's). Note that each of

TABLE 3
Intercorrelation Matrix for Potential Predictors
Within the Sample of 19 MOS

	TCOMBAT	TBONUS	WORKCON	TTRNG	SRB	APTITUDE	PROMOT	VARIETY	ERRCOST
TCOMBAT	1.0000								
TBONUS	0.5706	1.0000							
WORKCON	0.4859	0.5369	1.0000						
TTRNG	-0.2228	-0.2958	-0.0906	1.0000					
SRB	0.4134	0.3615	0.4096	-0.1174	1.0000				
APTITUDE	-0.5117	-0.6877	-0.5239	0.4505	-0.3010	1.0000			
PROMOT	0.4357	0.5474	0.4535	-0.3912	0.2609	-0.4123	1.0000		
VARIETY	0.3908	0.0065	0.3455	-0.0731	0.3937	0.0456	-0.0464	1.0000	
ERRCOST	0.6360	0.2311	0.4799	0.1088	0.4589	-0.0956	0.4124	0.5633	1.0000

Note: The sample size is n=19 Military Occupational Specialties
The $p < .05$, two-sided critical value for each correlation is .456.

TABLE 4

Average Correlations between each Predictor and Utility^{1, 2}

Predictor	Range of Correlations	Average Correlation
TCOMBAT	.696 - .730	.720**
TBONUS	.410 - .587	.492**
WORKCON	.329 - .513	.404**
TTRNG	-.183 - .034	-.047
SRB	.548 - .649	.596**
APTITUDE	-.305 - -.113	-.163
PROMOT	.333 - .534	.409**
VARIETY	.452 - .602	.524**
ERRCOST	.666 - .776	.718**

¹ Each average correlation is based on five correlations (i.e., at 5 performance levels)

² Each of the five correlations has a sample size of n = 19 MOS

** p < .01

the seven positive correlations is significant at the $p < .01$ level, with the remaining two correlations being slightly negative. Neither of these two negative correlations is statistically significant ($p > .05$).

PREDICTOR SELECTION AND POLICY CAPTURING

Using the predictors described above, ordinary least squares regression analyses were performed to determine the subset of variables which maximally predicted utility for the sample of 19 MOS. Utility values were regressed onto the 9 independent variables at each performance level, resulting in five regression analyses. (The specific technique used was best subset regression, which has the flexibility of both adding and deleting variables at each step.) Results of these analyses are presented in Table 5.

The multiple R^2 and the cross-validated R^2 , estimated using the Lord-Nicholson formula (Darlington, 1968), obtained at each step were plotted together to determine the point at which increases in R^2 were associated with decreased values of the cross-validated R^2 (Wherry's test selection method; Wherry, 1975).³ This method, therefore, identifies the point at which an increase in the sample-based R^2 , achieved through the addition of another variable, is not large enough to offset the effect of another

TABLE 5

Results of Regression Analyses by Performance Level (n = 19)

<u>Performance Level 10</u>			
<u>R²</u>	<u>shrunken R²</u>	<u>Step</u>	<u>Variables in the Model</u>
.524	.438	1	TCOMBAT
.598	.461	2	+ SRB
		3	+ PROMOT
.649	.463	3	SRB replaced by VARIETY

		4	+ TBONUS
		4	PROMOT replaced by ERRCOST
.696	.462	4	VARIETY replaced by APTITUDE
.723	.429	5	+ SRB
		6	+ PROMOT
.743	.373	6	ERRCOST replaced by VARIETY
<u>Performance Level 30</u>			
<u>R²</u>	<u>shrunken R²</u>	<u>Step</u>	<u>Variables in the Model</u>
.485	.392	1	TCOMBAT
		2	+ ERRCOST
.614	.483	2	TCOMBAT replaced by TBONUS
		3	+ APTITUDE
.710	.556	3	ERRCOST replaced by TCOMBAT
.783	.616	4	+ SRB

.802	.592	5	+ ERRCOST
.817	.553	6	+ WORKCON
<u>Performance Level 50</u>			
<u>R²</u>	<u>shrunken R²</u>	<u>Step</u>	<u>Variables in the Model</u>
.602	.530	1	ERRCOST
.696	.593	2	+ TCOMBAT
.747	.613	3	+ SRB
		4	+ APTITUDE
.794	.635	4	SRB replaced by TBONUS
.847	.685	5	+ SRB

.853	.641	6	+ VARIETY

Table 5 (continued)

<u>Performance Level 70</u>			
shrunken			<u>Variables in the Model</u>
<u>R²</u>	<u>R²</u>	<u>Step</u>	
.529	.444	1	TCOMBAT
.675	.565	2	+ SRB
.796	.688	3	+ APTITUDE

.817	.676	4	+ TBONUS
.824	.637	5	+ VARIETY
.831	.588	6	+ PROMOT

<u>Performance Level = 90</u>			
shrunken			<u>Variables in the Model</u>
<u>R²</u>	<u>R²</u>	<u>Step</u>	
.600	.528	1	ERRCOST
.692	.587	2	+ TCOMBAT
		3	+ SRB
.770	.648	3	ERRCOST replaced by APTITUDE
.809	.662	4	+ ERRCOST
.838	.666	5	+ TBONUS

.853	.641	6	+ VARIETY

predictor on the cross-validated (shrunken) R^2 value. These comparisons indicated that 3- to 5-variable models were optimal. The point at which the shrunken R^2 decreased within each performance level is noted in Table 5 by a broken line. Examination of these models across performance levels indicated substantial convergence regarding which of the variables were retained. Six variables were found to be pervasive throughout these models: ERRCOST, TCOMBAT, TBONUS, SRB, APTITUDE, and VARIETY.⁴

Thus, these six variables were chosen as the common set of independent variables, across performance levels, which were most important for capturing utility judgments. It should be noted that Table 5 reveals differences in the predictability of utility across performance levels. Specifically, the R^2 s for the 10th and 30th percentile performance levels are lower than those for the remaining performance levels. Examination of the variance for the criterion variable, utility, at each performance level offered an explanation for these differences in R^2 s. The variances of utility values at the 10th, 30th, 50th, 70th and 90th percentile performance levels are 0.011, 0.023, 0.031, 0.036, and 0.054, respectively. Note that the variance of utility values at the 90th percentile performance level (0.054) is five times greater than that at

the 10th percentile (0.011). This reduced variance in the criterion at the lower performance levels served to limit its predictability (i.e., correlation with other variables).

An analysis of covariance, using performance level as the covariate, was performed to test for homogeneity of regression coefficients. The group of interaction terms (i.e., performance level by each of the six predictors) was significant, indicating that the homogeneity of regression assumption was not statistically justified. However, the increment in R^2 due to this group of interaction terms was quite small in magnitude (from $R^2 = .957$ before inclusion to $R^2 = .967$ after inclusion of interaction terms, producing a change in R^2 of .010). This minimal increment of .010 suggests that the slight differences in beta weights across performance levels are not of practical significance.

UNIT-WEIGHTED ANALYSIS

Although the R^2 s and shrunken R^2 s obtained from ordinary least squares regression were quite encouraging, an analysis using unit-weighted predictors was performed to determine if such an approach may be preferable to that using regression weights. One advantage of unit-weighting is that there is no validity shrinkage because regression coefficients are not estimated from sample data (cf. Einhorn & Hogarth, 1975; Wainer, 1976).

The correlations (and squared correlations) between the utility value predicted by summing the six standardized variables and the true criterion values are presented, by performance level, in Table 6. Note that the r^2 obtained in this manner exceeds the shrunken R^2 from ordinary least squares regression in every case. Thus, prediction of utility values for MOS which were not evaluated directly would be most accurate through the application of a unit weighting approach.

FACTOR ANALYSIS

A principal components factor analysis was conducted to investigate, in a more general sense, the dimensions which underlie utility estimates. All 13 original variables were included in the analysis, resulting in a p/N ratio of 13/19. Bobko and Schemmer (1984) note that although large p/N ratios are not particularly problematic in terms of eigenvalue shrinkage, factor loadings may not remain stable across samples under such conditions. Therefore, the pattern of factor loadings produced in this analysis should be interpreted with caution. Results were quite clear in this case, however, and are described below, with the above caveat in mind.

TABLE 6

Results of Unit-Weighted Analysis using Six Predictors

<u>Performance Percentile</u>	<u>r</u>	<u>r²</u>
10	.809**	.654
30	.844**	.713
50	.905**	.819
70	.874**	.764
90	.914**	.835

Note: The sample size is n = 19 MOS

** p < .01

Three factors having eigenvalues exceeding 1.0 were retained and subjected to varimax rotation. These three factors accounted for $(9.362/13) \times 100 = 72\%$ of the total variance. Table 7 displays the resultant factor loading matrix. The first factor accounted for 27% of the variance, with substantial loadings on CLOSED, COMBAT, SRB, VARIETY, ERRCOST and PHYSDEM. The second factor, accounting for 27% of the variance, had substantial loadings on BONUS, STRENGTH, APTITUDE, PROMOT, and ACF. The third factor accounted for 18% of the variance, with substantial loadings on TRLGTH and TRCOST.

The first factor seems to summarize the types of work done in the Army, or the "job requirements". For example, the Army is "in the business of" being combat-ready (CLOSED and COMBAT), with enlisted jobs which are highly physical and/or technical (PHYSDEM and SRB), and associated with substantial task variety and cost of errors (VARIETY and ERRCOST). Factor 2 seems to represent the means by which the Army meets its labor needs. That is, it selects on the basis of specific APTITUDE and STRENGTH requirements, offers bonus packages (BONUS and ACF), and promotions into higher-level jobs (PROMOT). Thus, this factor might be labeled as "person requirements". The final factor obviously represents the Army's investments in training, and might be

TABLE 7

Factor Loading Matrix for Principal Components Factor Analysis

VARIABLE	FACTOR 1	FACTOR 2	FACTOR 3
CLOSED	0.63094	0.44703	-0.35402
COMBAT	0.72398	0.42107	-0.05771
BONUS	0.25029	0.78315	-0.38119
STRENGTH	0.17094	0.83533	0.14629
TRLGTH	-0.13138	-0.14389	0.95326
TRCOST	0.10074	-0.08333	0.92350
SRB	0.59225	0.25888	-0.05122
APTITUDE	-0.12789	-0.73528	0.40504
PROMOT	0.21984	0.60659	-0.31912
VARIETY	0.83964	-0.28709	-0.07893
ERRCOST	0.86589	0.13239	0.13027
PHYSDEM	0.77008	0.26494	-0.03394
ACF	0.09148	0.79796	0.02137
Variance Explained:	Factor 1 3.513	Factor 2 3.498	Factor 3 2.350

labeled "training requirements". The implications of these factors are considered below.

CLUSTER COMPARISONS

As discussed previously, a supplemental analysis comparing validity-based and utility-based clusters was of interest. The distribution of MOS into validity clusters was established based on the Army's nine aptitude composite groupings. The distribution of MOS into utility clusters was determined by means of performing cluster analyses (using Ward's minimum variance hierarchical procedure), based on values for the six utility attributes identified earlier. That is, MOS were clustered according to profile similarity across TCOMBAT, ERRCOST, TBONUS, SRB, APTITUDE and VARIETY.

Three distinct utility clustering schemes were considered. The first approach requested nine clusters in the cluster analysis to match the dimensionality used within the Army's validity system. The latter two clustering schemes were chosen on the basis of results from an unrestricted cluster analysis. The R^2 values associated with solutions having from one to ten clusters were as follows: .00, .29, .47, .53, .59, .63, .66, .68, .71, .73. A plot of these values against the number of clusters

suggested possible breaks at three clusters ($\underline{R^2} = .47$) and at six clusters ($\underline{R^2} = .63$). Based on this consideration of the incremental variance explained by additional clusters, both the three-cluster solution and the six-cluster solution were justified. Thus, three comparative analyses were performed, determining the degree to which validity cluster membership aided the predictability of the three utility cluster distributions.

The index used to evaluate the correspondence between the validity clusters and the three utility clustering schemes was Goodman and Kruskal's (1954) λ_B , λ_B . λ_B is an index of predictive association, in this case determining the proportional reduction in the probability of error in predicting utility cluster membership which is afforded by specifying validity cluster membership. The value of λ_B ranges from 0 to 1. An exemplary contingency table (that for the nine utility cluster solution) used in calculating λ_B is displayed in Table 8. The value presented in each utility-validity cell is the frequency or number of MOS belonging to a given aptitude cluster which also fall into a given utility cluster.

The resultant values of λ_B from the 9x9 (i.e., 9 validity clusters by 9 utility clusters), 9x6, and 9x3 contingency tables are: $\lambda_B (9,9) = .27$; $\lambda_B (9,6) = .34$;

TABLE 8
Contingency Table of Frequencies
for Nine Validity and Nine Utility Clusters

VALIDITY CLUSTER	UTILITY CLUSTER								
	1	2	3	4	5	6	7	8	9
Combat	0	0	0	0	0	0	6	1	0
Clerical	1	4	0	0	2	5	0	0	0
Electronics	8	1	5	8	1	1	0	0	0
Field Artillery	0	0	0	0	0	0	1	0	1
General Maintenance	3	0	0	0	2	8	0	0	1
Mechanical Maintenance	6	0	0	0	0	4	0	0	3
Operators & Food	0	0	1	0	1	1	3	0	2
Surveillance and Communications	2	0	0	0	0	1	0	0	0
Skilled Technical	2	3	3	1	3	0	0	4	1

and $\lambda_B(9,3) = .42$. For example, knowledge of validity cluster membership provides a 27% increase in the predictability of utility cluster membership (when 9 clusters are retained) over that possible by only knowing the marginal frequencies for utility. Each of the above values for λ_B is statistically significant ($p < .01$).

Examination of Table 8 also provides an indication of the distribution of Army aptitude clusters across utility clusters. Whereas some aptitude clusters are concentrated into one or two utility clusters (e.g., "Combat" MOS in utility cluster 7), most aptitude clusters are represented in several of the utility clusters (e.g., the "Operators and Food" composite is spread over five utility clusters). Note that the 17 MOS in the Army's "Skilled Technical" composite are associated with seven distinct utility clusters.

DISCUSSION

THE PREDICTOR DOMAIN

The substantial predictability of utility estimates obtained from the regression analyses suggest that the preliminary set of variables selected as potential predictors was well-chosen. Six of the nine variables appeared particularly useful for capturing utility judgments. Of these six, four were Army-based measures, with the remaining strong predictors, ERRCOST and VARIETY, representing subjective ratings made by non-Army personnel based on information contained in MOS job descriptions. Use of these variables as predictors resulted in very high R^2 values, in the .70's and .80's. Thus, it appears that estimates of utility can be adequately captured using job-level information cues.

Perhaps the most surprising result concerns the ERRCOST variable, derived from job evaluation literature. ERRCOST was an extremely strong predictor of utility, about as good as even the strongest Army-based predictor, TCOMBAT (see Table 4). Although these two variables are substantially correlated, ($r = .636$), cross-validated R^2 values were increased by the inclusion of both variables in the model. Further, the ERRCOST measure was provided by graduate

students in industrial/organizational psychology who had minimal knowledge of specific Army policies, procedures, etc. The fact that such ratings proved valuable in the regression analyses indicates that Army jobs can be distinguished from one another on the basis of their associated consequences of error, even as judged by relatively naive raters. The usefulness of this variable may actually be enhanced further if judges who are more familiar with the Army were to provide such ratings. More fundamentally, however, this variable reflects the importance of the individual job incumbent in controlling negative consequences which may result from inadequate job performance. The degree to which incumbents are responsible for major outcomes such as large monetary losses or widespread injury and/or death seems to be a major influence on the perceived value of their jobs.

TCombat, as mentioned above, was also a very strong predictor. This is not surprising, however, as it represents the relationship of each job to one of the Army's primary missions -- combat. In a wartime setting, for which the Army must be constantly prepared, MOS associated with high "combat probability" values are those most likely to be directly involved. Thus, its major contribution to the prediction of utility judgments reflects its vital importance to the accomplishment of Army goals.

After ERRCOST and TCOMBAT are entered into the regressions, the next two important predictors are TBONUS and SRB. Both are bonus-related variables. TBONUS is the bonus package offered at enlistment and thus reflects the Army's difficulty in recruiting qualified personnel into jobs involving selected critical skills. The Army is willing to offer more desirable enlistment bonus packages to insure adequate recruitment in its most valuable jobs.

SRB is the selective reenlistment bonus which may be offered to incumbents when they sign up for a second tour of duty. Two primary factors are related to the selective reenlistment bonus which is offered for an MOS. First, MOS requiring substantial physical activity (such as work involving significant physical strain or hazardous job duties) tend to be eligible for an SRB. To a large degree, these are combat-related jobs, which explains the moderately high correlation ($r = .41$) between SRB and TCOMBAT. Additionally, MOS requiring skills which are in high demand in the private sector are eligible for selective reenlistment bonuses. The Army provides SRBs as incentives for these trained, experienced incumbents to remain in the Army. The combination of these two criteria results in the SRB variable being a very valuable predictor of judged utility by encompassing jobs which are highly valued because

of their physical skill components as well as those which are valuable due to the substantial technical skills and expertise involved.

The final two variables which were most useful in capturing utility judgments were APTITUDE and VARIETY. The nature of APTITUDE's contribution is not straightforward. As shown in Table 4, the simple correlation between APTITUDE and utility is negative, in contradiction to the original hypothesis for this variable. Thus, higher aptitude requirements are, in general, associated with lower utility values. Upon consideration, it seemed plausible that this relationship may have resulted because utility judgments for combat-related jobs tended to be high, and at the same time, aptitude requirements for such jobs are relatively low. APTITUDE is also negatively related to the other important predictors (see Table 3), with the exception of a weak positive relationship with VARIETY. In particular, correlations between APTITUDE and TCOMBAT and TBONUS are substantial, being $-.51$ and $-.69$, respectively. However, APTITUDE enters positively into the regression equations. This positive contribution suggests that when the other predictor variables are held constant (i.e., partialled out), APTITUDE has a positive correlation with utility.

To investigate the hypothesis that the negative zero-order correlation between utility and aptitude is due to low aptitude requirements for combat jobs, the semipartial correlation between utility and APTITUDE, with TCOMBAT partialled out, was computed. The resultant correlation was positive, indicating that APTITUDE's contribution to utility is positive after TCOMBAT's influence is removed. Thus, for MOS associated with comparable TCOMBAT values, those with higher aptitude requirements tend to be judged as more valuable.

VARIETY, a measure obtained by means of subjective ratings, was the final important predictor of utility. It should be noted that VARIETY's influence was less pervasive than that of the other five variables discussed, especially in terms of the models indicated by the Wherry selection method. However, VARIETY typically entered the regression analyses within one or two additional steps. In light of the exploratory nature of this research, as well as the theoretical significance of this variable, it seemed preferable to use more variables than necessary than to use too few. The contribution of VARIETY to the regression analyses suggested that something about this concept, in addition to what it shares with the other predictors, makes jobs involving frequent exceptions to a routine sequence of

activities more valuable to the Army. The VARIETY dimension indexes the degree to which jobs may be considered routine or nonroutine, as well as suggesting the level of complexity inherent in a job. Results indicated that, at least in the Army context, the perceived value of jobs increases as their component tasks become less routine and repetitive.

GENERALIZABILITY

The utility judgments provided by field grade officers can therefore be adequately described as a linear combination of these six job-level variables. Given that utility estimates could be captured, this study attempted to generalize the judgment policy based on 19 jobs to the larger population of jobs. Thus, the cross-validated R^2 values derived from sample data were of primary interest. Although these values were quite high due to the excellent original R^2 values (and despite the large p/n ratio), a simple unit-weighting scheme was found to be superior.

The improved predictability from unit weights was not surprising for two related reasons. First, research and literature on regression analysis has consistently confirmed the power of the linear model (cf. Dawes, 1979). Second, the stability of results from unit weighting has been found to be superior to least-squares weighting, particularly when

small to moderate sample sizes are used and when the predictors are intercorrelated (cf. Claudy, 1972; Dorans & Drasgow, 1978; Schmidt, 1971). Note that the independent variables in this study were intercorrelated, in some cases substantially. These factors, in combination with our interest in the shrunken R^2 values, resulted in the specific differential weighting scheme not being an important aspect of increased predictive accuracy.

CLUSTER COMPARISONS

Results of analyses comparing validity-based and utility-based MOS clusters indicated that there was some overlap between these two approaches to viewing jobs. Thus, knowledge of validity clusters does, to some degree, provide knowledge of the utility of jobs. In turn, this suggests that some of the test-related characteristics which make jobs similar for validity generalization purposes also make them similar in terms of utility. In particular, the convergence between the two clustering schemes increased as the number of utility clusters decreased. Thus, validity information is especially useful in establishing broad categories of jobs associated with similar utility-relevant attributes.

At the same time, the relatively low improvement in predictability overall, as well as the lower lambda values for larger numbers of utility clusters, both suggest that the two approaches are clearly not equivalent. There was substantial variability in validity cluster membership throughout the utility clusters examined. Further, any particular utility cluster may be composed of MOS from several aptitude composite clusters. Together, these results indicate that jobs which are sufficiently similar for purposes of validity generalization may indeed be quite different in terms of their perceived organizational value, and vice versa. Thus, an approach which channels organizational resources into jobs for which performance can best be predicted may be shortsighted. It is also necessary to consider the utility associated with such jobs, a concept which extends beyond the test-related factors that enhance validity.

IMPLICATIONS FOR INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

As previously noted, the fundamental conclusion from this research is that utility information can be captured. Such a demonstration enriches our knowledge of utility, providing initial evidence on the types of characteristics which may underlie utility estimates. With the recent

increased emphasis on evaluating utility, researchers have called for studies which could enhance our understanding of utility (cf. Burke & Frederick, 1984; Karren & Bobko, 1983). The present study represents a significant first step in this direction, which may ultimately lead to a clarification and broadening of the theoretical foundations upon which utility research is based.

The post-hoc factor analysis performed in this research provides some insight into the general dimensions which may be related to utility judgments. That is, whereas the individual variables selected through regression are quite meaningful in this context, it is not clear how relevant they would be in other contexts. For example, the concept of TCOMBAT is obviously integral to the Army, but has little meaning in most other settings. The factor analysis therefore allows examination of the more general, summary dimensions which the predictors represent. Three factors emerged, and were labeled "job requirements", "person requirements", and "training requirements".

The "job requirements" factor is the cornerstone of this three-factor system, with its meaning depending entirely on the nature of the work performed by an organization. In turn, both "person requirements" and "training requirements" will be cast in terms relevant to an

organization's job requirements. As an example, if the "job requirements" aspect involves provision of medical services, the "person requirements" factor may include educational level and a shift-work variable, and the "training requirements" factor may involve completion of a CPR class. Note that although the training variables measured in this study were not strongly related to utility, they may take on more importance in other organizational contexts. Even in the Army context, training may become a stronger predictor of utility judgments when overall worth of soldiers beyond the entry level (e.g., second tour soldiers) is evaluated. In sum, these three dimensions may be a fruitful area for future research into utility estimates, the specific definitions of which can be tailored to the dynamics of a given organization.

The present research is also unique in that utility was not measured in terms of individual incumbents, but at the level of jobs. Whereas prior utility research has examined the relative utilities of varying performance levels within a job, this research also addresses the differences in utilities across jobs. As such, this study is particularly relevant in situations where organizational changes or interventions are introduced that may affect all jobs in the organization. For example, potential gains in utility

resulting from administration of general cognitive tests to all applicants can be evaluated both within and across jobs.

Also, in the attempt to capture utility judgments for jobs, this study included predictors which were analogous to job characteristics variables (e.g., complexity). As discussed previously, job characteristics have been studied as correlates of employee satisfaction and/or employee performance (e.g., Hackman & Oldham, 1975, 1976). However, the present research demonstrates the usefulness of job characteristics as correlates of the perceived utility or organizational value of specific jobs.

Although most of the predictors used in this research were objective measures obtained through Army sources, results revealed that the predictor domain was substantially enhanced by introduction of additional, more theoretical considerations. Perrow's (1970) framework for measuring departmental technology was adapted and proved valuable in the present research. This framework involves the concepts of "frequency of unexpected exceptions in the work" and "analyzability of problems encountered". Although measurement of the latter concept, "analyzability", was unreliable, the former concept, referred to as VARIETY, was retained as a valuable predictor of utility. Similarly, the "cost/consequence of error" dimension used in this research proved very useful.

Use of such variables is noteworthy for two reasons. First, they represent theoretical concepts for which "objective" data collection may be unfeasible, but which are very meaningful and well worth the effort of subjective measurement. Perhaps more importantly, they point to the value of exploring research which has been conducted in related disciplines or fields of expertise. For example, Perrow's "variety" concept is rooted in the organizational theory literature, and the "cost of error" concept derives from the field of job evaluation. Both, however, have analogs in the industrial/ organizational literature. The "variety" concept seems to be similar to the notion of "job complexity" found in the traditional job characteristics literature, while the "cost of error" dimension is measured (as "Impact of Inadequate Performance") in the Professional and Managerial Position Questionnaire (PMPQ) used as a job analysis tool (Mitchell & McCormick, 1979). Thus, a multidisciplinary approach to problems can be rewarding in terms of both practical and theoretical considerations.

One key step in studying organizations, particularly when introducing interventions such as training programs, is the process of needs assessment (Wexley & Yukl, 1977). A threefold process, it consists of organization, job, and person analysis. The first, organization analysis, attempts

to identify the major goals and values of an organization and involves reading organizational statements of purpose, philosophies of operation, and other materials delineating organizational objectives. The importance of organization analysis is frequently discussed but it is rarely conducted thoroughly. Results of this study show that the relative values of organizational components reflected by such global statements can be established. The current study demonstrates that such macro statements can be captured empirically. Thus, researchers are encouraged to consider policy capturing of utilities as one empirical component in any organization analysis.

Finally, representing a result which is relevant to the Army's motivation, the predictability of utility judgments (as evidenced by the multiple correlations) is high. Such a result implies the capacity to generalize utility information from a sample of jobs to the relevant population of jobs within an organization. In turn, the benefits to be realized from evaluation of the relative utilities of jobs include enhanced selection and classification, placement, and training. In this regard, utility generalization is quite similar to validity generalization, in that it extends the opportunity for improved organizational functioning to include a larger range of jobs than may be practical otherwise.

LIMITATIONS OF THIS RESEARCH AND FUTURE DIRECTIONS

Several limitations of this research should be noted. First, the conclusions drawn depend on the assumption that the utility estimates provided by field grade officers were indeed valid estimates. Field grade officers seemed to be the most appropriate group of individuals for the rating procedure due to their sufficient knowledge of Army policies as well as the scope of duties involved in the enlisted jobs. Thus, they represented an optimal "balance" between these two considerations. Further, their estimates demonstrated high interrater reliability, and various methods of utility estimation which have been conducted by the Army (e.g., rank orders, paired comparisons, scaled estimates) all show substantial convergence (Sadacca & Campbell, 1985). However, the degree to which these estimates reflect "true" utility to the Army is unknown. Certainly, the validity of the obtained results would be questionable if the criterion values, utility judgments, were not the "best estimates" of true utility.

A second concern involves the small sample size used in this research. Although there were 95 distinct criterion values, they were all derived from 19 MOS. Therefore, n was often 19 in the analyses. Two points, however, diminish the seriousness of this concern. First, the sample of 19 MOS

was carefully selected to be representative of the population. Second, predictability of the criterion was quite high, even upon estimation of the cross-validated R^2 , despite the small sample size.

Related to the above sampling issue, generalizations of these results to other populations should be made with caution. One factor making the Army context unusual is that, as a government organization, it may not share certain values held by private sector organizations. For example, the Army does not have the profit-oriented goals integral to most organizations. Also, the military context in which the Army functions adds to its uniqueness. In terms of the specific results obtained here, these characteristics were clearly influential. The factor analysis, however, pointed to a generic set of three dimensions which are relevant in any context. Further, the methodology is applicable to a wide range of organizations.

The generality of these results may be further limited by the use of a wartime scenario for utility estimation. That is, judges made estimates of overall worth assuming a wartime context. It may be that comparisons of results using different scenarios would reveal meaningful scenario effects. Thus, it should not be assumed that the present results are applicable to other contexts.

A final concern from this research is that there may be other types of important predictors which were not identified here. Perhaps inclusion of some other variables would substantially improve the predictability of utility. It may also be that a more parsimonious framework could be found which subsumes, yet expands upon, the predictor set used here. In this regard, the use of job-level data in this research represents both a strength and a weakness. It is a strength in that the job characteristics which were identified allowed the relative utilities across jobs to be evaluated and understood. At the same time, however, potentially valuable information at the level of individual incumbents is lost.

This research suggests several directions for future research on utility analysis. In terms of criterion measures, it is essential that there be a justifiable rationale for the choice of judges who provide utility information. This should be a decision which is given considerable thought, and not one based on convenience or accessibility. As noted in this study, raters need to be familiar with the jobs being evaluated as well as the goals and values of the organization. Research can help to establish which organizational levels are best suited to this task. Similarly, the influence of scenario-related

factors on estimates of overall worth should be investigated. Study of possible scenario effects may increase our understanding of utility estimates, help indicate the most appropriate scenario in a given situation, and enhance our ability to generalize results.

Although results of this research indicate substantial predictability of utility, the approach demonstrated here should be replicated with sufficient samples from other populations of jobs. The methodology is potentially useful for any organization interested in evaluating utility, but seems especially relevant for service and non-profit oriented organizations where non-monetary values take on prime importance. In particular, future research may benefit from focusing on theoretical-based predictors, such as the "cost of error" and "variety" dimensions used in the present research. These variables seem worthy of further study since they enhanced the predictability of utility and were evaluated by relatively naive graduate students. Perhaps measurement efforts can improve upon the assessment of these constructs through the use of raters who are more knowledgeable about the organization being studied.

Finally, future research may reveal other important predictors of utility. The present study builds upon "job characteristics" research, although the dynamics involved

are quite different (i.e., no individual psychological states are posited) and the criterion of interest is utility rather than job satisfaction or performance. Rigorous study of utility estimation may ultimately provide a complete conceptual framework which encompasses both job-level and individual-level information.

In conclusion, this research demonstrates that utility judgments across jobs can be captured. Such information is useful for practical purposes (e.g., enhanced classification) as well as the improved theoretical understanding it provides of what dimensions correlate with judgments of overall worth. Accordingly, and as implied by Equations 1 and 2 (page 2), researchers are encouraged to expand their focus beyond the role of validity and correlations to the intersection of validity, measurement of the standard deviation of performance, and organizational values -- that is, to the study of utility.

FOOTNOTES

¹ The use of a wartime scenario in the rating task was chosen because of its relevance to the Army's primary mission, which is to be prepared to respond to catastrophic, wartime events.

² It should also be noted that correction for range restriction requires the assumption that the regression coefficients between these variables and the criterion, utility, are the same in the sample as in the population. Assuming that the focus is on regression weights, corrections to correlations which are then used to estimate regression coefficients become meaningless.

³ An alternative, but similar, approach to choosing regression subsets involves the use of Mallows' C statistic (cf. Seber, 1977). Examination of Mallows' statistic for the present regression analyses indicated precisely the same cutoff points for the models as those presented in Table 5.

⁴ An analysis in which all 95 (19 MOS X 5 performance levels) criterion values were regressed onto the nine independent variables and performance level provided further support that these six variables were the most important. Performance level was the first variable to enter (the simple correlation between performance level and utility was

.91). Results revealed that the six variables identified above were precisely those retained in addition to performance level as the best 7-variable model. Further, the 7-variable model was optimal according to the Wherry test selection method, with the cross-validated R^2 decreasing upon inclusion of 8 variables.

REFERENCES

- Bellak, A. O. (1984). Specific job evaluation systems: The Hay Guide Chart-Profile Method. In M. L. Rock (Ed.), Handbook of wage and salary administration (pp. 15/1-15/16). New York: McGraw-Hill.
- Bobko, P. (in press). Utility analysis: A primer and application to organizational stress interventions. In A. Riley, S. Zaccaro, & R. Rosen (Eds.), Occupational stress and organizational effectiveness. New York: Praeger Press.
- Bobko, P., Karren, R., & Parkington, J. (1983). Estimation of standard deviations in utility analyses: An empirical test. Journal of Applied Psychology, 68, 170-176.
- Bobko, P., & Schemmer, F. M. (1980). Note on standardized regression estimators. Psychological Bulletin, 88, 233-236.
- Bobko, P., & Schemmer, F. M. (1984). Eigenvalue shrinkage in principal components based factor analysis. Applied Psychological Measurement, 8, 439-451.
- Brady, D., & Rappoport, L. (1973). Policy-capturing in the field: The nuclear safeguards problem. Organizational Behavior and Human Performance, 8, 253-266.

- Brogden, H. (1949). When testing pays off. Personnel Psychology, 2, 171-183.
- Burke, M. (1984). Validity generalization: A review and critique of the correlation model. Personnel Psychology, 37, 93-113.
- Burke, M., & Frederick, J. (1984). Two modified procedures for estimating standard deviations in utility analyses. Journal of Applied Psychology, 69, 482-489.
- Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. Journal of Applied Psychology, 65, 543-558.
- Campbell, J. P. (1983). Sample selection. In N. K. Eaton & M. H. Goer (Eds.), Improving the selection, classification, and utilization of Army enlisted personnel: Technical appendix to the annual report (Research Note 83-37, pp. 17-40). Alexandria, VA: U. S. Army Research Institute.
- Cascio, W. F., & Silbey, V. (1979). Utility of the assessment center as a selection device. Journal of Applied Psychology, 64, 107-118.
- Christal, R. E. (1968). Selecting a harem - and other applications of the policy-capturing model. The Journal of Experimental Education, 36, 35-41.

- Claudy, J. G. (1972). A comparison of five variable weighting procedures. Educational and Psychological Measurement, 32, 311-322.
- Cronbach, L., & Gleser, G. (1965). Psychological tests and personnel decisions. Urbana: University of Illinois Press.
- Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. American Psychologist, 34, 571-582.
- Department of the Army. (1984). Army Regulation 611-201: Enlisted career management fields and military occupational specialties. Washington, DC: Author.
- Department of the Army, DAPC-PLP. (1983). Time in service statistical report (Fiscal Year 1981). Washington, DC: Author.
- Department of the Army, Programs Management Branch. (1982, January). Format 6: Enlisted personnel receiving selective reenlistment bonus by DoD occupational area and military specialty, by zone of reenlistment for period 1 Oct. 80 thru 30 Sep. 81. Washington, DC: Author.

- Dorans, N., & Drasgow, F. (1978). Alternative weighting schemes for linear prediction. Organizational Behavior and Human Performance, 21, 316-345.
- Eaton, N. K., Goer, M. H., & Zook, L. M. (1984). Introduction to current Army selection and classification research. In N. K. Eaton, M. H. Goer, J. H. Harris, L. M. Zook, (Eds.), Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1984 fiscal year. (Tech. Rep. No. 660). Alexandria, VA: U. S. Army Research Institute.
- Eaton, N. K., Wing, H., & Mitchell, K. J. (1985). Alternate methods of estimating the dollar value of performance. Personnel Psychology, 38, 27-40.
- Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 13, 171-192.
- Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. American Psychologist, 23, 483-496.
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. Journal of the American Statistical Association, 49, 732-764.
- Hackman, J. R., & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. Journal of Applied Psychology, 60, 159-170.

- Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. Organizational Behavior and Human Performance, 16, 250-279.
- Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. Psychological Bulletin, 57, 116-131.
- Hoffman, P. J., Slovic, P., & Rorer, L. G. (1968). An analysis of variance model for the assessment of configural cue utilization in clinical judgment. Psychological Bulletin, 69, 338-349.
- Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, and Army Research Institute. (1983a). Improving the selection, classification, and utilization of Army enlisted personnel. Project A: Research plan (Research Rep. No. 1332). Alexandria, VA: Authors.
- Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, and Army Research Institute. (1983b). Improving the selection, classification, and utilization of Army enlisted personnel. Project A: Annual report (Research Rep. No. 1347). Alexandria, VA: Authors.
- Hunter, J. E., & Schmidt, F. L. (1982). The economic benefits of personnel selection using psychological ability tests. Paper presented at the Conference of the

Bureau of National Affairs on Current Directions in Productivity: Evolving Japanese and American Practices, Houston, TX.

Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on job performance and work force productivity. American Psychologist, 78, 473-478.

Karr, C., & Fischbach, G. T. (1984). Specific job evaluation systems: American Association of Industrial Management. In M. L. Rock (Ed.), Handbook of wage and salary administration (pp. 13/1-13/12). New York: McGraw-Hill.

Karren, R., & Bobko, P. (1983, August). Conducting utility analyses: Some methodological concerns. Paper presented at the meeting of the Academy of Management, Dallas, TX.

McLaughlin, D. H., Rossmeissl, P. G., Wise, L. L., Brandt, D. A., & Wang, M. M. (1984). Validation of current and alternative ASVAB area composites, based on training and SQT information on FY 81 and FY 82 enlisted accessions (Tech. Rep. No. 651). Alexandria, VA: U. S. Army Research Institute.

Mitchell, J. L., & McCormick, E. J. (1979, July). Development of the PMPQ. A structured job analysis questionnaire for the study of professional and

managerial positions. (PMPQ Rep. No. 1). West Lafayette, IN: Department of Psychological Sciences, Purdue University.

- Morris, J. W. (1973). Principles and practice of job evaluation. London: Heinemann.
- Patten, T. H., Jr. (1977). Pay: Employee compensation and incentive plans. New York: The Free Press.
- Perrow, C. (1970). Organizational analysis: A sociological view. Belmont, CA: Wadsworth.
- Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. Journal of Applied Psychology, 68, 382-395.
- Robinson, D. D., Wahlstrom, O. W., & Mecham, R. C. (1974). Comparison of job evaluation methods: A "Policy Capturing" approach using the Position Analysis Questionnaire (PAQ). Journal of Applied Psychology, 59, 633-637.
- Rosse, R. L., Borman, W. C., Campbell, C. H., & Osborn, W. C. (1983, October). Grouping Army occupational specialties by judged similarity. Paper presented at the Annual Conference of the Military Testing Association, Gulf Shores, AL.
- Sadacca, R., & Campbell, J. P. (1985, March). Assessing the utility of a personnel/classification system. Paper

presented at the meeting of the Southeastern Psychological Association, Atlanta, GA.

- Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. Educational and Psychological Measurement, 31, 699-714.
- 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., McKenzie, R., & Muldrow, T. (1979). The impact of valid selection procedures on work force productivity. Journal of Applied Psychology, 64, 609-626.
- 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.
- Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. Personnel Psychology, 32, 257-281.
- Schneider, B. (1980). The service organization: Climate is crucial. Organizational Dynamics, 9, 52-65.

- Seber, G. A. F. (1977). Linear regression analysis. New York: John Wiley.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.
- Stone, E. F. (in press). Job scope-job satisfaction and job scope-job performance relationships. In E. Locke (Ed.), Generalizing from laboratory to field settings: Research findings from Industrial-Organizational Psychology, Organizational Behavior, and Human Resources Management (pp. 189-206). Lexington, MA: Lexington/Heath Books.
- Stone, E. F., & Gueutal, H. G. (1985). An empirical derivation of the dimensions along which characteristics of jobs are perceived. Academy of Management Journal, 28, 376-396.
- Taylor, R. L., & Wilsted, W. D. (1974). Capturing judgment policies: A field study of performance appraisal. Academy of Management Journal, 17, 440-449.
- U. S. Army Finance & Accounting Center. (October, 1983). Military occupational speciality training cost handbook (MOSB). Washington, DC: Author.

- U. S. Army Recruiting Command. (1984). U. S. Army pocket recruiting guide. Ft. Sheridan, IL: Author.
- Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 83, 213-217.
- Wexley, K. N., & Yukl, G. A. (1977). Organizational behavior and personnel psychology. Homewood, IL: Richard D. Irwin, Inc.
- Wherry, R. J., Sr. (1975). Underprediction from overfitting: 45 years of shrinkage. Personnel Psychology, 28, 1-18.

Appendix A

MATERIALS PROVIDED TO FIELD GRADE OFFICERS FOR
EVALUATING THE UTILITY OF 19 MOS

WARTIME SCENARIO

Hostilities have broken out in Europe and your Corps' combat units are engaged. Your Corps' mission is to defend, then re-establish, the host country's border. Pockets of enemy airborne/heliborne and guerilla elements are operating throughout the Corps sector area. Limited initial and reactive chemical strikes have been employed but nuclear strikes have not been initiated. Air parity does exist.

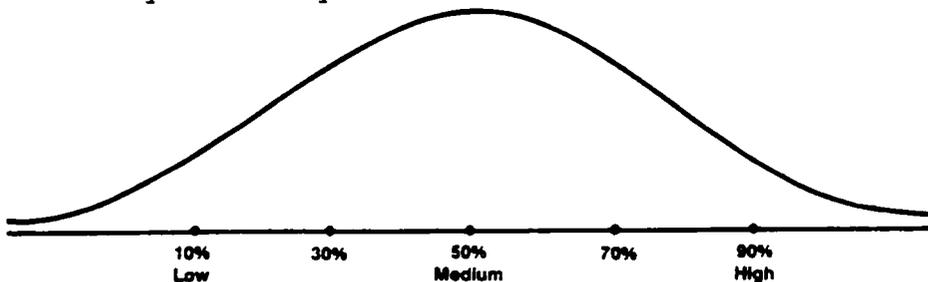
INTRODUCTION TO THE RATING PROCEDURE

Soldiers in different MOS operating at different performance levels can have different utility or overall worth to the Army in wartime. In order to determine how best to measure this utility, we are trying out various methods of obtaining judgments of experienced Army officers. In the procedures to be tried out today you will be given a set of 95 cards. On each card there is a short description of a different first tour soldier. The descriptions contain the following information about each soldier:

- (1) A brief summary of the soldier's MOS and associated job duties; and
- (2) The percentile rank of the soldier on an overall measure of the effectiveness of the soldier in wartime derived from combining hands-on performance scores, knowledge test scores, ratings and administrative indexes to form one composite score.

The summary MOS statements were taken from AR 611-201, which gives descriptions of all Army MOS. The overall effectiveness percentiles for each MOS were derived by first combining the separate measures into a composite using expert judgments to weight the components for that MOS. Then the percentiles were calculated for the composite scores using a large, representative sample of first tour soldiers in each MOS.

The soldiers on the cards are from 19 MOS. Please assume that the soldiers described on the 95 cards are at either the 10th, 30th, 50th, 70th, or 90th percentile on the overall performance measure for their respective MOS. Please also assume that the spread or amount of variation in performance is equal in each MOS and that the scores are approximately normally distributed:



Note: 10th percentile indicates low overall performance and 90th percentile indicates high performance.

DIRECTIONS FOR ASSIGNING UTILITY VALUES

Setting the value of a
50th percentile Infantryman - 11B at 100

In this procedure you will assign a numerical utility value to each of the soldiers you previously ranked. The values that you assign will be proportionate to the value assigned to a 50th percentile Infantryman - 11B. In other words, the overall worth or utility of the 50th percentile Infantryman will be used as a yardstick and the worth of all other soldiers will be judged in relationship to this Infantryman. (This is comparable to the use of a given platinum bar as the defined length of 1 meter or 100 centimeters in the metric system.) In this case the value of the 50th percentile Infantryman will be set at 100.

To make the judgments please follow these steps:

- (1) Write the value, 100, on the 50th percentile Infantryman card which is on top of the deck you just received. (The other cards in the deck are in random order.)
- (2) Then take each of the other cards in turn and assign the soldier a utility value which reflects the worth of each soldier given that the 50th percentile Infantryman has a worth of 100. You may assign higher values than 100 or even negative values to one or more soldiers if you wish. Write the values you assign directly on the cards.
- (3) When you have gone through the deck once, please arrange the cards in numerical order from lowest to highest value.
- (4) Then go through the cards once more and change any assigned value that you feel is out of line with the others (with the exception of the value of 100 assigned to the 50th percentile Infantryman.)

In making your judgments please keep in mind the same European wartime scenario described earlier.

Thank you again for your cooperation.

SAMPLE RATING STIMULI

INFANTRYMAN

SUMMARY: Leads, supervises and serves as member of an infantry activity employing individual weapons and machineguns in offensive and defensive combat operations.

DUTIES: Closes with and destroys enemy personnel weapons and equipment.

OVERALL EFFECTIVENESS: 90 percentile

MOTOR TRANSPORT OPERATOR

SUMMARY: Supervises or operates wheel vehicles to transport personnel and cargo.

DUTIES: Operates wheel vehicles.

OVERALL EFFECTIVENESS: 70 percentile

UTILITY HELICOPTER REPAIRER

SUMMARY: Supervises, inspects, or performs maintenance on utility helicopters, excluding repair of systems components.

DUTIES: Assists in organizational, direct, and general support (aviation unit, intermediate and depot) maintenance of utility helicopters, excluding repair of system components.

OVERALL EFFECTIVENESS: 50 percentile

MEDICAL SPECIALIST

SUMMARY: Supervises dispensary or field medical facilities, administers emergency medical treatment to battlefield casualties, assists with inpatient and outpatient care and treatment, and assists with technical and administrative management of medical treatment facilities.

DUTIES: Performs routine field medical activities and patient care procedures.

OVERALL EFFECTIVENESS: 30 percentile

Appendix B

SAMPLE JOB DESCRIPTIONS PROVIDED TO GRADUATE
STUDENTS WHEN RATING JOB CHARACTERISTICS

UTILITY HELICOPTER REPAIRER

SUMMARY: Supervises, inspects, or performs maintenance on utility helicopters, excluding repair of systems components.

DUTIES: Assists in organizational, direct, and general support (aviation unit, intermediate and depot) maintenance of utility helicopters, excluding repair of system components.

May perform utility helicopter crewchief duties. Assists in removal and installation of subsystem assemblies such as engines, transmissions, gearboxes, rotor hubs and rotor blades. Removes and replaces subsystem components such as starters, generators, inverters, voltage regulators, lights, batteries, pumps, reservoirs, valves, hydraulic cylinders, and hoses. Cleans, services, and lubricates helicopters and subsystems. Prepares helicopters for extensive inspection and maintenance by removing and replacing items such as cowling, inspection plates, panels, floorboards, doors, and auxiliary equipment. Assists in maintenance operational checks. Applies corrosion control preventive measures. Assists in preservation and depreservation techniques in preparing helicopters for entry into and removal from storage and in preparing and loading for surface and airborne shipment. Assists in removing helicopters from transit shipping status and placement in flyable status. Prepares requests for turn-ins and repair parts. Uses and performs user maintenance on common and special tools and ground support equipment required for helicopter maintenance and ground handling. Employs certain helicopter maintenance forms and records as related to MOS.

UNIT SUPPLY SPECIALIST

SUMMARY: Supervises or performs duties involving request, receipt, storage issue, accounting for, and preservation of individual, organizational, installation, and expendable supplies and equipment.

DUTIES: Receives, stores, issues, accounts for, and preserves supplies in unit.

SUPPLY HANDLING. Receives, loads, unloads, segregates, stores, issues and delivers organizational and installation supplies and equipment. Verifies units of issue, description, and quantity of requested materiel against issue/turn-in documents. Moves property to issue or transfer points. Prepares storage facilities to facilitate issue, inventory, and segregation of supplies by type of class within fire, safety, and security requirements. Inventories supplies and equipment and records results. Identifies, sorts, and cleans supplies. Processes individual and organizational laundry. **SUPPLY ADMINISTRATION.** Keeps administrative files. Posts supply regulations, publications, records, and forms used. Uses supply catalogs in identification, assembly, and breakdown of supplies. Prepares request and turn-in documents. Processes supply requests. Prepares follow-up and cancellation requests. Prepares, reviews, and corrects hand receipt property listings and annexes. Posts transactions to record of demand. Prepares clothing and equipment records. Prepares abstract of organizational clothing and equipment. Prepares laundry rosters and lists. Checks or prepares payroll deduction form for laundry. Prepares inventory adjustment documents. Uses typewriters, adding machines, calculators, copier, and microfiche viewers in performance of duties. Uses supply catalogs in identification, assembly, and breakdown of supplies. **CLOTHING RECORDS.** Inspects clothing and equipment record of newly assigned personnel to determine shortage and unauthorized items. **ARMORER DUTIES.** Maintains key control register for weapon storage area. Issues and receives small arms. Controls weapons and ammunition in arms room. Secures and performs organizational maintenance on unit arms. Turns in unserviceable weapons for repair. Prepares required weapons and ammunition reports for commander.

**The two page vita has been
removed from the scanned
document. Page 1 of 2**

**The two page vita has been
removed from the scanned
document. Page 2 of 2**