AFHRL-TP-90-54

AIR FORCE

AD-A224 496

H
U
M
A
N

R
E
S
O
U
R
C
E
S

LABORATORY

APPLICABILITY OF UTILITY MODELS TO THE EVALUATION OF MILITARY MANPOWER AND PERSONNEL RESEARCH PROGRAMS: A CRITICAL REVIEW AND ILLUSTRATIONS

DTIC
ELECTE
JUL 3 1 1990
S   B   D

Michael D. Matthews

Drury College
Behavioral Science Department
900 N. Benton Avenue
Springfield, Missouri 65802

Larry T. Looper
Sheree K. Engquist, Capt, USAF

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

July 1990

Final Technical Paper for Period June 1987 - June 1990

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

90   031

NOTICE

WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Division

ROGER W. ALFORD, Lt Col, USAF
Deputy Chief, Manpower and Personnel Division

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE July 1990	3. REPORT TYPE AND DATES COVERED Final Paper - June 1987 - June 1990

4. TITLE AND SUBTITLE

Applicability of Utility Models to the Evaluation of Military Manpower and Personnel Research Programs: A Critical Review and Illustrations

5. FUNDING NUMBERS

PE - 62205F
PR - 7719
TA - 20
WU - 20

6. AUTHOR(S)

Michael D. Matthews
Larry T. Looper
Sheree K. Engquist

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

Manpower and Personnel Division
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235-5601

8. PERFORMING ORGANIZATION REPORT NUMBER

AFHRL-TP-90-54

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

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A major problem in the manpower and personnel research community is the transfer of technology from the laboratory to operational settings. One solution to this problem involves translating statistical findings that may be difficult to interpret into economically meaningful statements. Utility analysis is one way of accomplishing this translation. This paper provides a critical review of recent developments in utility analysis and assesses how applicable these developments are to military research programs. Illustrative examples of the application of utility models to Air Force personnel research projects are given. Recommendations are made for areas of future research.

14. SUBJECT TERMS

personnel model   utility assessment
utility analysis   utility model

15. NUMBER OF PAGES
30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

APPLICABILITY OF UTILITY MODELS TO THE EVALUATION OF MILITARY MANPOWER AND PERSONNEL RESEARCH PROGRAMS: A CRITICAL REVIEW AND ILLUSTRATIONS

Michael D. Matthews

Drury College
Behavioral Science Department
900 N. Benton Avenue
Springfield, Missouri   65802

Larry T. Looper
Sheree K. Engquist, Capt, USAF

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed by

David E. Brown, Lt Col, USAF
Chief, Force Management Systems Branch

Submitted for publication by

Roger W. Alford, Lt Col, USAF
Deputy Chief, Manpower and Personnel Division

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

A major problem faced by researchers in the military manpower and personnel domain is transferring technology from the laboratory to operational settings. Like their counterparts in the civilian sector, military researchers often meet considerable resistance from management personnel when attempting to make this transfer. One reason that such resistance is encountered is that scientists and managers often communicate in a different language.

One very actively researched method of dealing with this problem in the past ten years has been a decision-theoretic approach known as "utility analysis" (Hunter & Schmidt, 1983). This method uses information concerning the validity of a selection test, estimates of job performance in dollars and cents terms, and costs incurred with the selection test or intervention. Until recently, estimating certain parameters in these equations was so cumbersome as to severely limit the practical application of utility analysis to the evaluation of research. Recently, however, improved ways of providing these estimates have been advanced.

This study is a review of recent developments in utility analysis and evaluation of their applicability to the assessment of human resource research program. Two applications of these utility concepts are presented. The first being the Vocational Interest Career Examination and the second a Methodology for Generating Efficiency and Effectiveness Measures. The appendix contains an alternate method of estimating standard deviation of job performance in dollars. Finally, recommendations for further research are presented.

i

PREFACE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Applicability of Utility Models to the Evaluation
of Military Manpower and Personnel Research Programs:
A Critical Review and Illustrations

I. INTRODUCTION

The fundamental objective of industrial and organizational psychology is ultimately to improve the effectiveness and efficiency of work organizations. In past decades numerous and substantial advances have been made toward this end. The use of ability tests and measures of other personal attributes has allowed organizations to select better qualified workers and to assign them more effectively to specific jobs within the organization. Improved training programs and other post-selection organizational interventions such as goal setting and feedback have enhanced individual worker performance and group productivity. Strong scientific evidence suggests that the optimal application of these techniques would result in more satisfied employees who would work harder and be less likely to leave the organization prematurely (Katzell & Guzzo, 1983).

Despite convincing evidence of the benefits of research dealing with the selection, training and treatment of employees, human resources specialists have often met with considerable resistance from management when trying to implement their programs and procedures in operational settings. A major impediment to this transfer of technology from laboratories to work organizations is one of poor communications. The criteria which a scientist uses to evaluate the worth of a research project are not the same as those used by management personnel. As Hunter and Schmidt (1983) point out, "to assess the practical impact of findings, one must translate such arcane psychological jargon as $p < .01$ into economically meaningful statements such as '10% increase in output' or 'a reduction of $100 million in labor costs'."

This technology transfer problem is particularly disturbing in the face of widely publicized concerns regarding national productivity. The U.S. appears to be losing ground to some other nations in this domain. The seriousness of the situation is reflected in a 1985 presidential memorandum (Reagan, 1985) which stated:

> ....I have asked the Congress to demonstrate its
> support for a government wide program to improve
> Federal productivity by passing a joint resolution
> declaring productivity a national goal. Viewed in
> conjunction with the Gramm-Rudman-Hollings Act,
> enhancing productivity is thus becoming a major
> national objective.

The methods needed to translate the impact of personnel selection tests and other organizational interventions into terms of economic gains have existed for many years (e.g., Brogden, 1949; Cronbach & Gleser, 1965; Taylor & Russell, 1939). These techniques and more recent ones based upon them (e.g., Cascio, 1982; Schmidt, Hunter, McKenzie, & Muldrow, 1979) are known in the literature as "utility analysis," which is defined as "...the determination of institutional gain or loss (outcomes) anticipated from various courses of action" (Cascio, 1982) or "...the assessment of the economic or social impact of organizational programs" (Hunter & Schmidt, 1983). Thus, the emphasis of

1

utility analysis is upon the interpretation of research findings and the translation of scientific information into terms understandable by managers.

It should be noted that utility analysis is very similar in objective to cost benefit analysis. Both attempt to assess the returns expected from a given approach to a problem against the costs of using that approach. While the phrase "cost benefit analysis" may seem more descriptive, work in this area grew more out of the field of economics (e.g., Sassone & Schaffer, 1978), and it focuses less on psychological variables than does the work reported in the area of utility analysis. Indeed, it appears that because of its emphasis on the interpretation of psychological manipulations such as selection tests and training programs, the field of utility analysis is particularly relevant to manpower and personnel researchers in the DoD environment. Because of this degree of relevance, the current paper will focus primarily upon utility analysis.

Roach (1984) conducted a very thorough historical review of utility analysis models, and identified 16 models from both the psychological and economics literatures. Although Roach concluded that utility analysis is "well suited to handling complex decision problems requiring value judgments and cost considerations" (p. 20), he also noted evidence of resistance to its application both by researchers and more traditionally minded practitioners. Moreover, Roach concluded that "there is no optimum selection model," rather, which of the 16 models to use depends on the problem being addressed in a given situation.

This paper extends the historical review began by Roach. Section II discusses recent developments in utility models along with issues involved in applying utility models to Air Force manpower and personnel research. Section III presents illustrative examples of the application of such models to military personnel research programs, focusing on those conducted by the Air Force Human Resources Laboratory (AFHRL). The appendix contains an alternative method for estimating a key utility assessment parameter.

II. RECENT DEVELOPMENTS IN UTILITY ANALYSIS

Overview

The objective of utility analysis is to allow researchers to translate sometimes esoteric research findings into practical terms. Said another way, utility analysis is one method of evaluating the costs versus the benefits of a given selection test or post-selection organizational intervention. The rapid development in this area of inquiry has yielded many elaborations to the basic utility models and concepts first espoused by Taylor and Russell (1939), Brogden (1949), and Cronbach and Gleser (1965). Because the majority of these elaborations are based on the application of utility models to for-profit organizations, their relevance to large non-profit organizations like the military must be examined. The very basis of utility analysis involves viewing utility as a function of returns and costs. Cronshaw and Alexander (1985) express the model as

$$\text{Utility (U)} = \text{Returns (R)} - \text{Costs (C)}$$

Cronbach and Gleser's (1965) utility model estimated R by the quantity

$$R = N_s SD_y r_{xy} Z_x$$

where SD_y refers to the standard deviation of job performance expressed in dollars, r_{xy} is the correlation between a selection test and work performance, and Z_x is the mean standard score for selectees. C was simply

$$C = N_s C_i$$

or the number (N_s) of individuals given the selection test times the cost (C_i) of each administration.

Modifications to the basic conceptual model that have been made over the past decade have focused on measures of benefit (Returns), expanding them to account for turnover, post selection interventions, and economic variables, as well as refining and empirically testing key components of the equation. What follows is a systematic and critical analysis of each of these modifications in terms of their applicability to the evaluation of military manpower and personnel research.

Employee Flow (Turnover)

Early utility models assumed a more or less stable workforce. Obviously, however, the gains expected from a given selection test or other intervention will be affected by the flow of employees into and out of the organization. For example, suppose a training program was administered to a given work organization consisting of 100 workers. If the average turnover rate is 25 employees per year, then after one year only 75 workers would remain who had received the treatment, after 2 years only 50 would remain, and so on. Boudreau (1983b) traced the development of utility theory and developed equations to take into account the flow of employees into and out of an organization. This development, however, was based on the application of utility models to for-profit industries.

At a glance, the role of employee flow into and out of an organization would seem particularly relevant to military research. The average tenure in an Air Force assignment does not usually exceed four years and many tours are substantially less. Thus, even with a four year tour, an organization will show substantial gains and losses of personnel in any given year. So, if a training program is given once at time X in a unit, then by X + 1 year, only 75 percent of the personnel in the unit will have received the treatment, and by X + 4 years, none will remain who received the treatment. Thus, returns resulting from the treatment will decline rapidly, or repeated treatments may be necessary.

Boudreau (1983b) reformulated the utility model of Schmidt et al. (1979) to account for turnover and then reanalyzed Schmidt et al.'s data. He added a term to the utility equation which expresses the number of treated employees in the workforce K periods in the future following the intervention

$$N_k = \sum_{t=1}^{k} (N_{At} - N_{St})$$

where N_k refers to the number of treated employees that will remain in the workforce k periods in the future, t is the time period through which the flow occurs, N_{At} represents the number of employees added to the workforce in period t, and N_{St} symbolizes the number of treated employees that leave the workforce during period t.

The important question is to what extent does this parameter apply in military settings? It may or may not be of critical importance for several reasons. First, many military human resources programs are ongoing, thus all subjects in the workforce will be treated. For example, all new recruits take the Armed Services Vocational Aptitude Battery (ASVAB). Second, and most importantly, while military organizations do experience substantial flow, as discussed above, many of the accessions into a given unit will have had a very similar or identical job at another duty assignment. So, employees flowing into the organization are not as naive, as a group, as employees entering a typical civilian work organization. Because of these considerations, further developments in the turnover parameter are necessary before it can be meaningfully used in the military context. The refinement of this parameter to account for the military situation should be a high priority in tailoring utility models to the evaluation of manpower and personnel research within this context.

Estimating SD_y

The main parameter of the utility equation that poses a major problem in performing utility analysis is SD_y. Prior to 1979, this value could be obtained only in organizations where accurate cost-accounting information was maintained. Even where this information was available, the estimation of SD_y was a burdensome task. Schmidt et al. (1979), however, report the development of a way for estimating this value based on interviews with persons knowledgeable of the job performance of the target individuals within the organization. Basically, this process involves asking the supervisors of such persons to estimate the value to the organization, in dollar terms, of (1) workers at the 50th percentile of performance and (2) those at the 15th and 85th percentile of performance. If job performance is normally distributed, then workers at the 15th and 85th percentiles should be one standard deviation from the mean. Since the estimates are made in dollar terms, then subtracting the worth of the 50th percentile worker from that of the 85th percentile worker, and/or subtracting from the worth of the 50th percentile worker that of the 15th percentile worker, provides an estimate of SD_y. This procedure is referred to in the literature as the "global estimation procedure."

The development of this procedure for estimating SD_y spurred considerable research in the field of utility analysis, both in evaluating the procedure itself and in refining general utility models. It also encouraged the application of utility analysis to the evaluation of organizational interventions. The importance of this development is well stated by Roach (1984, p. 18) who says "..their global estimation procedure was a milestone in the development of the decision-theoretic approach to personnel selection since it provided a means for determining the dollar criterion. This determination had eluded personnel researchers for a quarter of a century."

In further studies of SD_y, Schmidt found that the value of this parameter usually falls between 40 and 70 percent of the mean annual salary for the job incumbents being studied (Hunter & Schmidt, 1983; Schmidt, Mack, & Hunter, 1984). Thus, if one wishes to make a quick estimate of SD_y, a conservative estimate is provided by computing the value of 40 percent of the mean annual salary of the sample. This technique is referred to in the

literature as the "salary percentage technique." Furthermore, several studies have compared Schmidt et al.'s (1979) original global estimate procedure with his 40 percent method and with procedures suggested by other researchers. Weekley, Frank, O'Connor, and Peters (1985) examined SD_y estimates using both Schmidt techniques and compared them to those obtained using an alternative procedure, CREPID, developed by Cascio (1980). Results indicated that Schmidt's global estimation procedure generated estimates of SD_y approximately 1.8 times larger than the other techniques. Methodological constraints, however, precluded any conclusions regarding which technique for estimating SD_y is "better." Burke and Frederick (1986) obtained estimates of SD_y in a sample of mid-level sales managers, and confirmed that this estimate falls within 40 -70 percent of the mean base salary of job incumbents.

One problem with the Schmidt technique of estimating SD_y is that when asked to estimate the value of an incumbent in dollars, there is a wide range of responses. Each person queried seems to have their own idiosyncratic metric. Bobko, Karren, and Parkington (1983) describe a methodological solution to this problem, which involves averaging responses from the 50th percentile estimates and giving this mean figure back to raters as the basis against which to estimate the worth of workers at the 15th and 85th percentiles. An alternative to the Schmidt technique is provided in the appendix.

While Schmidt's procedure for estimating SD_y has held up fairly well to comparisons against other methods, its applicability to organizations in which money or profit is not the primary objective is questionable. For example, it might be difficult for a supervisor in an Air Force security police squadron to quantify, in dollars and cents terms, the worth of an incumbent at any given percentile of performance. Eaton, Wing, and Mitchell (1985) recognized this problem in applying the Schmidt technique in the evaluation of Army research. They considered two alternative methods of estimating SD_y using a sample of tank commanders. This sample should be considered representative of other military units in that performance in financial terms is difficult to conceptualize. The first technique considered by Eaton et al. is called the "Superior Equivalents Technique." "Instead of using estimates of dollar value of 85th percentile percentile performance, however, the technique uses estimates of how many superior (85th percentile) performers would be needed to produce the output of a fixed number of average (50th percentile) performers" (Eaton et al., 1985, p. 29). Results indicated that the Superior Equivalents Technique yielded consistent, although lower, estimates of SD_y compared to other techniques.

The second technique studied by Eaton et al. was called the "System Effectiveness Technique." It called for estimates of the number of units comprised of superior performers needed to perform at the level of a fixed number of units comprised of average performers. In comparing the two methods, it was concluded that "the Superior Equivalents Technique appears to be the method of choice in situations where supervisors are more accustomed to dealing with performance in output terms, or in relative output of individuals, rather than in dollar terms." (Eaton et al., 1985)

5

In terms of its applicability to the military, further research appears to be necessary in order to address some of the psychometric and conceptual concerns mentioned above. At the current time, it is suggested that integrating the Bobko et al. (1983) and Eaton et al. (1985) variations of the original Schmidt procedure would hold the greatest promise for military applications.

Economic Variables

In order to develop accurate utility models, Boudreau (1983a) argues that certain economic variables such as discounting, variable costs associated with fluctuations in production, and taxes must be entered into utility equations. While economists have traditionally attended to these factors and others (e.g., sensitivity analysis, break-even analysis, risk and uncertainty), they have done so within the context of traditional cost benefit analysis, which rely primarily on resource intensive validity studies. Consequently, the present review will consider only those applications of economic variables done within the context of utility analysis, and will refer the reader to other sources (e.g., Sassone & Schaffer, 1978) for a review of the economist's perspective.

The major work dealing with economic variables within the context of utility analysis is that of Boudreau (1983a). He proposed three economic variables that would modulate the expected utility of any given human resources intervention. These were discounting, variable costs, and taxation. Briefly, discounting refers to the fact that the worth of capital does not remain constant over time. Forces such as interest accrued to invested monies, inflation and the like alter the value of capital over time. For example, if a human resources intervention was under consideration for placement in a given organization, the benefits resulting from its implementation (in terms of increased productivity, decreased turnover, etc.) must be weighed against the growth of the money needed to implement the program if it were invested instead of used in the intervention. Taxation refers to increased tax liability resulting from enhanced productivity, and variable costs refer to increased costs for raw materials as a function of an increase in productivity. Boudreau discusses these factors within the context of private for-profit industries, and develops equations for each variable to add to the basic utility equation defined by Cronbach and Gleser (1965).

The notion that economic considerations may impact the value of utility estimates is no doubt a valid one, but the economic variables operating within a military context are not the same as those found in the private sector. For example, taxation of military resources is not an issue. Additionally, since many or most military operations are based on human performance rather than production factors, the impact of variable costs is lessened. And, given the nature of military funding procedures, one could question the relevance of discounting as well. That is, military organizations are usually given a set budget, and will spend that money one way or the other in order to justify that budgeting level. So, while a command may find it in its best interest to implement a human resources intervention in order to enhance its combat readiness or productivity, the decision not to implement that program does not suggest that the monies so saved will be invested.

6

Thus, it is questionable that any of the economic variables explored to this point in the utility analysis literature are relevent to the military context. Therefore, it is suggested that they be withheld from utility equations used in evaluating military manpower and personnel research.

Expanding Beyond Selection Tests

Another recent major advance in utility analysis involves the expansion of the basic utility model from selection tests to allow for the evaluation of post-selection organizational interventions, such as training programs or the introduction of job performance feedback systems. This is an important extension from the perspective of military applications, because many military manpower and personnel research programs focus on post-selection interventions. A good example is the Methodology for Generating Efficiency and Effectiveness Model (MGEEM). The MGEEM is a procedure for developing indices of group or organizational productivity, and, when used in conjunction with feedback and goal setting, to enhance productivity (Tuttle & Weaver, 1986).

Landy, Farr, and Jacobs (1982) and Schmidt, Hunter, and Pearlman (1982) suggest very similar ways of expanding utility formulas to account for post-selection interventions. Both suggest substituting a term that expresses the difference in performance between an experimental group (that is, the one receiving the treatment) with a control group (which does not receive the treatment). Landy et al. refer to this term as d_t, while Schmidt et al refer to it as d_y. A drawback to the use of these two methods is that the researcher must know the magnitude of difference between treated and non-treated groups. Consequently, in many cases an experiment must be performed to estimate the magnitude of this difference before the utility analysis may proceed. However, this may not pose a problem for researchers in the R and D environment, who (presumably) will have performed several studies on an intervention prior to reaching the stage of applying a utility analysis to the procedure in question.

This area of utility analysis has received the least attention from scientists involved in developing utility approaches. Both Schmidt et al. and Landy et al. calculate d_t as

$$d_t = \frac{\bar{X}_E - \bar{X}_C}{SD\ (R_{yy})^{1/2}}$$

where $\bar{X}_E$ is the mean performance for the group receiving the experimental treatment and $\bar{X}_C$ is the mean performance for group not receiving the treatment. SD is pooled variance and the reliability of the criterion measure is R_{yy}.

This may be a somewhat simplistic method of estimating the impact of an independent variable. Most recent psychological research involves more than one level of an independent variable, and many involve more than two independent variables. Multiple criterion measures are not unusual. Expansion of the d_t term to accommodate the additional information obtained from more complex experimental designs would improve utility estimates

obtained, and be of greater use to military researchers. Nevertheless, the developments in this area are of value and may be used in evaluating Air Force manpower and personnel research programs.

Other Developments

Two other modifications to utility models have been suggested. Boudreau and Rynes (1985) argued that recruitment practices may alter the pool of applicants and to the extent this is true, may affect the expected utility of a given selection test or organizational intervention. While this may be an important consideration in smaller organizations, its relevance to the military situation is questionable. The Air Force's recruitment methods aim at the general population of 18-21 year old citizens. Because of the size of the Air Force's recruiting efforts, it is unclear what effect changes in recruiting practices might have on the applicant pool. This is often not the case. For example, applicants to graduate school often apply to more than one institution. The most highly qualified applicants are accepted by several programs. If half of a graduate school's first choices for admission turn them down, then potential selections remaining in the pool are not representative of the original pool. Their selection test scores will be lower because of the nonrandom attrition of more highly qualified applicants. To the extent this is true, expected utility should decrease.

Murphy (1986) offers modifications to the model to account for this. Additional thought needs to be given to this issue in the military context. An analysis of recruiting records should reveal whether rejected offers are substantially altering the profile of selectees. Until this research and conceptualization is accomplished, it is recommended that the role of rejected offers not be considered essential to utility models as applied in military settings.

Application of Utility Models

Relatively few applications of utility analysis have been reported in the literature. This is due, largely, to difficulties in estimating SD_y. Two recent papers applied utility analysis to assist in determining the utility of using selection tests in hiring employees. These studies are important because they illustrate the value of utility analysis in evaluating personnel procedures and also because they provide a "how to" example to prospective users of utility analysis procedures.

Schmidt, Mack, and Hunter (1984) evaluated the use of the General Aptitude Test Battery (GATB) in selecting U.S. Park Rangers. Hunter and Hunter (1984) provided information about the validity of aptitude tests. They concluded that the correlation between aptitude tests and job performance is approximately .51. SD_y was estimated using the Schmidt et al. (1979) procedure via interviews with 114 first-line supervisors of park rangers. Information on other variables (for example testing costs) in the utility equation was gathered. Results of the utility analysis indicated that the incorporation of the GATB into the selection procedure could result in savings as high as $3,764,000 per year, if 130 rangers were hired annually. Estimated utility varied as a function of number of selectees and the longer their predicted tenure, the greater the expected utility. Also, utility was maximized using a top down selection model (that is, selecting subjects with

8

the highest GATB scores) versus the use of cutoff scores at or below the mean on the GATB.

Mathieu and Leonard (1987) recently provided an example of applying the Schmidt et al. (1982) utility formula for evaluating a training program. The effect of a training program on the supervisory skills of bank employees was examined. The authors also used vectors for taxes, variable costs and discounting, as suggested by Boudreau (1983a). Comparing employees who had received the training program with a matched group who had not, an estimate of d_t of .3146 was obtained. Estimates of SD_y were made and the results of the utility analysis indicated substantial dollar benefits to the organization by using the training program.

III. ILLUSTRATIVE EXAMPLES

Introduction

In this section the utility model defined above will be applied to two ongoing AFHRL research projects. One application will illustrate the use of utility analysis in evaluating a classification test (the Vocational Interest Career Examination, or VOICE). The other application will illustrate the use of utility analysis in evaluating a post-selection intervention (the MGEEM). As accurate information as possible will be used in the illustrations, but the values of some parameters will have to be estimated.

Application to a Selection Test

The VOICE is an interest battery designed to assess interests among recruits for blue-collar jobs typical of Air Force enlisted occupations. The basic interest scales have been validated against job satisfaction (Alley, Wilbourne, & Berberich, 1976). Predicted job satisfaction (PJS) scores based on the basic interest scale scores have been validated against attrition (Matthews, 1982), technical school attrition (Matthews & Ballentine, 1983), and rated work performance (Berry & Matthews, 1982). A thorough description of the VOICE, its development, and its psychometric characteristics is provided by Alley and Matthews (1982).

The intended use of the VOICE is in the classification of recruits to first-term Air Force assignments. It would be used to help place recruits into jobs in which they would be likely to enjoy and be successful. Unlike the ASVAB, it would not be used to exclude a recruit from a given career field. Rather, it would be used more as a counseling tool to help the recruit make an informed choice regarding a selection of a job among those available to him or her.

The first assumption that is required to use utility analysis is that the VOICE will be administered to all Air Force enlisted accessions (about 55,000 per year). For the purposes of the utility model, a discount rate of 10 percent is assumed. SD_y may be estimated by using the observation that its value is usually between 40 and 70 percent of the mean annual salary of the target workers. In this case, the target sample is first term airmen, whose average annual salary including benefits and allowances for food, housing, and clothing is approximately $15,000 per year. Using the most conservative value

(i.e., 40 %) as the basis for estimating SD_y, it may be estimated to be $6,000.

For r_{xy}, a value may be obtained from data published by Berry and Matthews (1982). They showed that the VOICE correlates with supervisor's ratings of incumbent job performance at a significant (p < .01) but low value of .053. This value may be taken as a conservative estimate of the true value of r, because many lows performers in the Berry and Matthews study had attrited from the Air Force prior to information on their job performance being obtained.

It is also assumed that the average Z score for each applicant is .8. Naturally, the estimated dollar savings resulting from implementation of any selection test will depend on how selective the test is, but this value would seem to be a reasonable one. Finally, it is estimated that the cost for administering and scoring the VOICE is $20 per selectee. The VOICE may be administered in the field at MEPS stations, where the cost could be a little higher due to the one-on-one nature of testing in those environments. It may also be administered during Basic Military Training (BMT), where the nature of testing in those environments costs would be somewhat lower, due to the possibility of testing large numbers of subjects at a time.

Given the utility formula

$$U = N_s\ r_{xy}\ SD_y\ Z_x - N_s C_i$$

U may be estimated by substituting in for the missing values. In this case, it turns out that U = $12,892,000.

So, given the above assumptions and parameter values, the Air Force would gain almost 13 million dollars annually through increased job performance by implementing the VOICE into its classification system. This savings does not even include that which might result from decreased attrition resulting from VOICE implementation.

Another point to be made about this illustration is that it is conservative. The true value of r_{xy} may be higher. It is also possible that the value of SD_y is higher than $6,000. If either of these parameter values were larger, U would be much larger. For example, if r was .15 instead of .053, U would be $38,500,000.

It is also interesting to note that the actual increase in utility per individual is rather small. In the first example given above, the value of U for each individual is $234.40, or less than two percent of the worker's annual salary. Such increments in performance are so small as to be nearly inconsequential in the case of an individual, but spread over large numbers of workers, the returns can be substantial.

Note that the above estimates are based on the notion of comparing the VOICE to random selection. This would not be the case in the actual application of the instrument in the Air Force. Measures of aptitude already exist and are used in addition to education level and other personal and demographic information. Suppose that the multiple R between the vectors of the existing classification equation and performance was .6. Using the values assumed earlier, U for this existing equation would be $157,300,000. This is

10

compared to random sampling from the population of applicants. Suppose further, however, one could increase the multiple R to .65 by adding the VOICE to the equation. The value of U now becomes $12,100,000, with the existing system as its reference point (that is, the U associated with the multiple R of .65 minus the U associated with the multiple R of .60). Thus, it would seem that utility analysis may be used in the co text of complex prediction equations in use in military environments.

Application to a Post-Selection Intervention

AFHRL recently sponsored the development of a procedure for estimating group productivity (Tuttle, Wilkinson, & Matthews, 1985). This procedure, the MGEEM, utilizes group dynamic procedures to identify key aspects of a unit's mission and ways of quantifying and measuring this mission critical feature. These indicators of productivity are then scaled such that they can be summed to provide an overall index of unit productivity. Commanders may then feed this information back to the managers of the organization and set organizational productivity goals. Tuttle et al. report that, in a field test of the MGEEM on Air Force units, the procedure is highly acceptable to users, is cost-effective, and satisfied the commander's need for a comprehensive measure of organizational productivity.

The MGEEM, of course, provides an example of a post-selection intervention. A utility analysis using the techniques described by Landy et al. (1982) and Schmidt et al. (1982) may be applied to it. To illustrate such an application, consider the following scenario. The target organization is an aircraft maintenance shop with 1000 employees. Recall that the utility formula for these situations is

$$U = N_s d_t \, SD_y - N_s C_i$$

In this case, lets assume that SD_y is $5000 and costs (see Tuttle and Weaver, 1986, p. 14) total 3000 dollars. Assume further that d_t is .65. Given these assumptions, U equals a value of 3,247,000 dollars.

While such an estimate may seem disproportionately large, consider that the mean salary of the workers may be $25,000 per year. The total budget just for salary then is 25,000,000 dollars. Thus, the U value obtained represented only 13 percent of the total budget for salary. If other budgetary factors are also considered such as materials, travel expenses, etc., the value of U would be an even smaller proportion of the total budget for the organization, but would represent a sizeable sum in absolute terms.

IV. CONCLUSIONS AND RECOMMENDATIONS

The objective of this paper was to review recent developments in utility analysis and evaluate their applicability to the assessment of human resource research programs in military settings. Because the additions to the basic utility equation first proposed by Cronbach and Gleser (1965) pertaining to economic variables, turnover, recruitment, and rejected offers do not appear relevant to the military as currently defined, it is suggested that military manpower and personnel researchers continue to use the Cronbach-Gleser formula (as described earlier in the paper) until such modifications to the additional

variables are developed. Thus, the formula suggested for application to military research is

$$U = N_s r_{xy} SD_y Z_x - N_x C_i$$

where the variables are defined as they were previously, allowing for discounting of the time value of money as U is summed over time.

Illustrative examples show that the types of research most often done by AFHRL researchers and indeed all industrial psychologists, can be quantified into dollars and cents terms using utility models. Utility analysis can be very useful in evaluating alternatives. For example, it could help a manager select among two or more alternative selection tests under consideration for use in selecting or classifying workers. Utility analysis would rank order, at the very minimum, which tests should show the greatest return. However, it would be an error to interpret literally the dollar estimates obtained in a utility analysis. Such figures may not be highly valid dollar estimates of returns accruing from a given course action.

A large number of variations and modifications have been added to the basic utility concepts, although not all of these developments are germane to utility applications in nonprofit organizations like the military. More development needs to be done in the area of tailoring utility models to nonprofit organizations and/or situations where there is not a tangible product. Extension of terms, like turnover, to military organizations is needed. For example, the impact of turnover in the military is probably different from that in civilian jobs. While some members do leave the Air Force, many simply move on to highly similar jobs in different units. Thus, an experienced aircraft jet engine mechanic added to a maintenance shop at one air base will not be as costly to the receiving organization in terms of training as is the addition of an inexperienced mechanic.

Utility analysis has a role in estimating not only the worth of research and development efforts but also is a member of a set of tools which can evaluate the costs and benefits of any project or organizational intervention. When used cautiously, it is a powerful tool providing much insight and help in making effective use of scarce Air Force resources.

REFERENCES

Alley, W. E., & Matthews, M. D. (1982). The Vocational Interest
  Career Examination: A description of the instrument and possible
  applications. Journal of Psychology, 112, 169-193.

Alley, W. E., Wilbourn, J. M., & Berberich, G. L. (1976).
  Relationships between performance on the Vocational Interest-Career
  Examination and reported job satisfaction. (AFHRL-TR-76-89, AD-A040 754).
  Lackland Air Force Base, Texas: Air Force Human Resources Laboratory.

Berry, G. A., & Matthews, M. D. (1982). Predicting job satisfaction
  and job performance. Proceedings of the 24th Annual Conference of the
  Military Testing Association. San Antonio, Texas: Air Force Human
  Resources Laboratory.

Bobko, P., Karren, R., & Parkington, J. J. (1983). Estimation of
  standard deviations in utility analysis: An empirical test. Journal of
  Applied Psychology, 68, 170-176.

Boudreau, J. W. (1983a). Economic considerations in estimating the
  utility of human resource productivity improvement programs. Personnel
  Psychology, 36, 551-576.

Boudreau, J. W. (1983b). Effects of employee flows on utility
  analysis of human resource productivity improvement programs. Journal of
  Applied Psychology, 68, 396-406.

Boudreau, J. W., & Rynes, S. L. (1985). Role of recruitment in
  staffing utility analysis. Journal of Applied Psychology, 70, 354-366.

Brogden, H. E. (1949). When testing pays off. Personnel Psychology,
  2, 171-183.

Burke, M. J., & Frederick, J. T. (1986). A comparison of economic
  utility estimates for alterntive SD estimation procedures. Journal of
  Applied Psychology, 71, 334-339.

Cascio, W. F. (1980). Responding to the demand for accountability: A
  critical review of three utility models. Organizational Behavior and Human
  Performance, 25, 32-45.

Cascio, W. F. (1982). Costing human resources. New York: Van
  Rostrand Reinhold.

Cronbach, L. J., & Gleser, G. C. (1965). Psychological tests and
  personnel decisions (2nd ed.). Champaign, Il: University of Illinois Press.

13

Cronshaw, S. F., & Alexander, R. A. (1985). One answer to the demand for accountability: Selection utility as a investment decision. Organizational Behavior and Human Decision Processes, 35, 102-118.

Eaton, N. K., Wing, H., & Mitchell, K. J. (1985). Alternative methods of estimating dollar value of performance. Personnel Psychology, 38, 27-40.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee job performance and workforce productivity. American Psychologist, 38, 473-479.

Katzell, R. B., & Guzzo, R. A. (1983). Psychological approaches to productivity improvement. American Psychologist, 38, 468-472.

Landy, F. J., Farr, J. L., & Jacobs, R. R. (1982). Utility concepts in performance measurement. Organizational Behavior and Human Performance, 30, 15-40.

Mathieu, J. E., & Leonard, R. L. (1987). Applying utility concepts to a training program in supervisory skills: A time-based approach. Academy of Management Journal, 30, 316-335.

Matthews, M. D. (1982). Vocational interests, job satisfaction, and turnover among Air Force enlistees. Proceedings, Fourth International Learning Technology Congress and Exposition. Orlando, Florida: Society for Learning Technology.

Matthews, M. D., & Ballentine, R. D. (1983). Predicted job satisfaction and attrition from Air Force technical schools. Proceedings, 25th Annual Meeting of the Military Testing Association. Gulf Shores, Alabama: U.S. Navy.

Murphy, K. R. (1986). When your top choice turns you down: Effect of rejected offers on the utility of selection tests. Psychological Bulletin, 99, 133-138.

Reagan, R. (1985, July 31). Memorandum for the heads of Executive Departments and Agencies. Washington, D. C: The White House.

Roach, B. W. (1984). Decision-theoretic approach to personnel selection: A review. (AFHRL-TP-84-19, AD A143 388). Brooks Air Force Base, Texas: Air Force Human Resources Laboratory.

Sassone, P. G., & Schaffer, W. A. (1978). Cost-benefit analysis: A handbook. New York: Academic Press.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 64, 609-626.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on work force productivity. Personnel Psychology, 35, 333-347.

Schmidt, F. L., Mack, M. J., & Hunter, J. E. (1984). Selection utility in the occupation of U.S. Park Ranger for three modes of test use. Journal of Applied Psychology, 69, 490-497.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 13, 565 - 578.

Tuttle, T. C., & Weaver, C. N. (1986). Methodology for generating efficiency and effectivenss measures (MGEEM): A guide for commanders, managers, and supervisors. (AFHRL-TP-86-26). Brooks Air Force Base, Texas: Air Force Human Resources Laboratory.

Tuttle, T. C., Wilkinson, R. E., & Matthews, M. D. (1985). Field test of the methodology for generating efficiency and effectiveness measures. AFHRL-TP-84-54, AD-A158 183. Brooks Air Force Base, Texas: Air Force Human Resources Laboratory.

Weekley, J. A., Frank, B., O'Connor, E. J., & Peters, L. H. (1985). A comparison of three methods of estimating the standard deviation of performance in dollars. Journal of Applied Psychology, 70, 122-126.

Comparison of Supervisor's and Incumbent's
Estimates of SDy

The most important developments in utility analysis pertain to methods for estimating the standard deviation of job performance in dollars and cents terms. This variable is a key one in utility models, and is symbolized by SD_y. Schmidt et al. (1979) proposed a way of estimating this variable based on obtaining expert judgments of the dollars and cents worth of workers at the 50th, 15th, and 85th percentiles of performance. If job performance is normally distributed, then by subtracting the estimates of the 50th percentile from the 85th, and the 15th from the 50th, then an estimate of SD_y is obtained. A number of subsequent studies examined the validity of this approach (e.g., Burke & Frederick, 1986; Weekley et al., 1985), and have found it to be fairly accurate in the populations studied, although some psychometric questions do persist. Additionally, Hunter and Schmidt (1983) report that estimates of SD_y tend to fall near 40 percent of the mean annual salary for the job incumbents, a technique referred to as the "salary percentage technique," (Eaton, Wing, & Mitchell, 1985).

The Schmidt et al. (1979) method for estimating SD_y relies on obtaining estimates from supervisors concerning the value to the organization of the incumbents. In order to obtain an accurate estimate, a relatively large sample of supervisors is required. For example, Weekly et al. (1985) examined a sample of 196 supervisors employed by a national convenience store. Because of the large number of supervisors required to obtain these estimates, applying utility analysis methods in small organizations could be difficult due to there being an insufficient number of supervisors to provide performance estimates. This could limit the applicability of this technology to relatively large organizations.

A solution to this limitation exists if one could obtain the required estimates from the incumbents themselves. Conceptually, one could question whether or not incumbents have a sufficiently broad view or understanding of their contributions to the organization as a whole to provide the basis for accurate judgments. On the other hand, as Tuttle, Wilkinson, and Matthews (1985) have argued, no one knows their job as well as the workers themselves. Perhaps they are indeed in a good position - even the best position - to make such estimates of worth. This issue may be tested empirically, by obtaining estimates of value of workers from both supervisors and icumbents from the same organization, and comparing these estimates.

Estimates of the dollar value to the organization of workers at the 50th, 15th, and 85th percentiles of performance were obtained from a sample of supervisors and job icumbents. No basis exists for an a priori prediction of the outcome of the study, but the impliciations of the outcome are clear. If supervisors and incumbents were found to make similar estimates, then future utility analysis studies could be conducted in smaller organization, which lack the number of supervisors needed to make accurate estimates, but which have sufficient numbers of incumbents. Indeed, data from supervisors and incumbents could be pooled to allow for a sufficiently large sample to accurately estimate utility parameters.

The subjects used in the current study were task scientists and their supervisors from a government human resources research laboratory. Eighteen incumbents were included in the sample. Their average age was 39.6, 16 were male, and two were female. Their average job tenure was 9.8 years. Their GS ratings ranged from GS-11 to GS-13. Thirteen supervisors were studied. Their average age was 42.2, 10 were males, and three were females. Their average tenure in the organization was 14 years, and they had been supervisors for an average of 6.5 years. Their grade ranged from GS/GM 13 to GS/GM 14.

Each respondent was asked three questions. These were modeled after the questions used by Schmidt et al. (1979), and required that the supervisors and incumbents provide estimates of the monetary worth of a "task scientist" who performed at the 50th, 15th, or 85th percentile. In addition, the respondents were asked their age, gender, time of service in the organization and (for supervisors) how long they had been in a supervisory position.

A master list of the personnel assigned to the target organization, the manpower and personnel research division of a military research organization, was obtained and task scientists and their supervisors were identified. All such personnel assigned to the division were surveyed. The researcher made personal contact with each subject, explained that he was conducting a project designed to develop ways of quantifying the performance of workers in economic terms, reviewed the questionnaire with the subject, and left it with her/him to complete. The researcher would only explain the questions as posed, and provided no guidance to the respondents about how to make their judgments. That is, each respondent, after examining the questionnaire, independently generated a response using her/his own personal metric.

The day following the distribution of the questionnaires, the completed forms were gathered and collated. Three respondents were away from the organization at the time of the initial survey. Copies of the questionnaire were left with their supervisors for them to complete when they returned. All three completed the questionnaire and mailed them to the researcher.

The main findings from the study are shown in Figure A-1, which shows the mean dollar estimates given by supervisors and incumbents for workers at the 15th, 50th, and 85th performance percentiles. Supervisors ($\overline{X}$ = $26,962) and incumbents (X = $24,670) gave similar estimates for the dollar value of the 15th percentile workers ($\underline{t}$=1.02; df=29;p $>$.05). The estimates of the value of 50th percentile workers were less similar, with supervisors ($\overline{X}$ = $75,923) providing a larger estimate than incumbents ($\overline{X}$ = $65,511). The difference between estimates given by incumbents and supervisors for 50th percentile workers was significant ($\underline{t}$=2.88, df=29;p $<$.05). Finally, the greatest disparity between value estimates of incumbents and supervisors was for workers at the 85th percentile. Supervisors provided a mean estimate of $278,150 for these superior workers, while the mean estimates of the incumbents was $160,878. This difference was also significant ($\underline{t}$=18.58; df=29;p $<$.05).
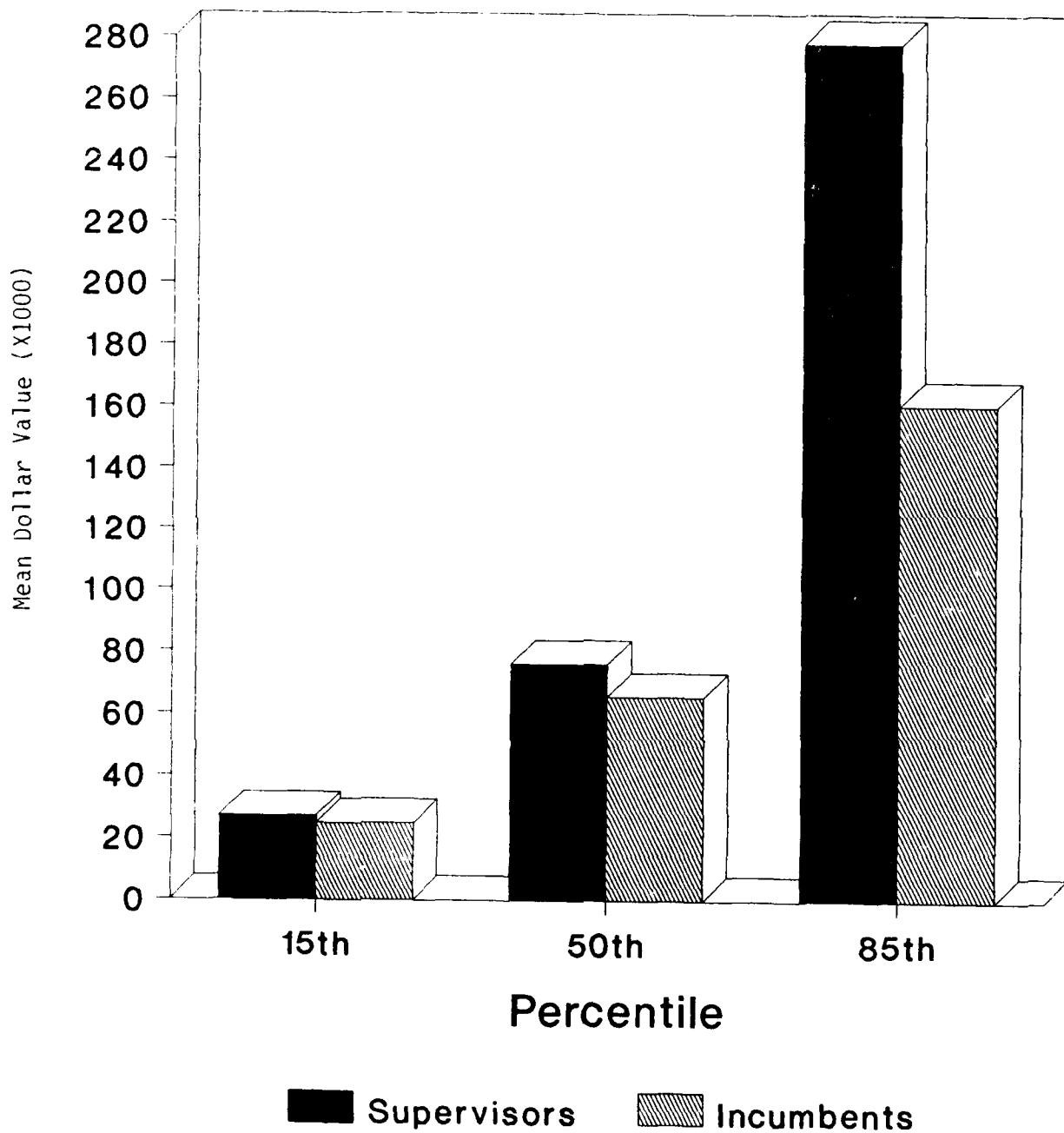
Figure A-1.   Mean Dollar Estimates Given by Supervisors and
             Incumbents

Estimates of SD_y can be obtained by subtracting the mean estimates given for the 50th percentile from those of the 85th percentile, and those of the 15th from the 50th. Such estimates for both supervisors and incumbents are given in Table A-1. These data suggest that estimates of SD_y based on subtracting the 15th percentile from the 50th are similar for both supervisors and incumbents. However, estimates derived from subtracting the 50th from the 85th percentile are quite disparate, with supervisors providing a much larger estimate than incumbents. These estimates suggest that performance, as estimated by the dollar value of workers, is not distributed symmetrically around the mean (i.e., the 50th percentile).

Table A-1

Estimates of SD_y Based on Supervisor's and Incumbent's Data

	85th - 50th	50th - 15th
Supervisors	$202,227	$48,961
Incumbents	$ 95,367	$40,841

Because of the apparent asymmetrical nature of these estimates, it could be argued that the mean does not provide the most accurate estimate of central tendency. Accordingly, median values for estimates by supervisors and incumbents of workers at the three performance percentiles were obtained. These medians are shown in Figure A-2. The median values given are substantially less than corresponding mean values. For example, ne median estimate of workers at the 85th percentile given by supervisors was $95,000, compared to a mean estimate of $278,150. Also, while there appears to be some difference between median estimates of supervisors and incumbents, especially for workers at the 15th and 85th percentiles, analysis by Mann-Whitney U test
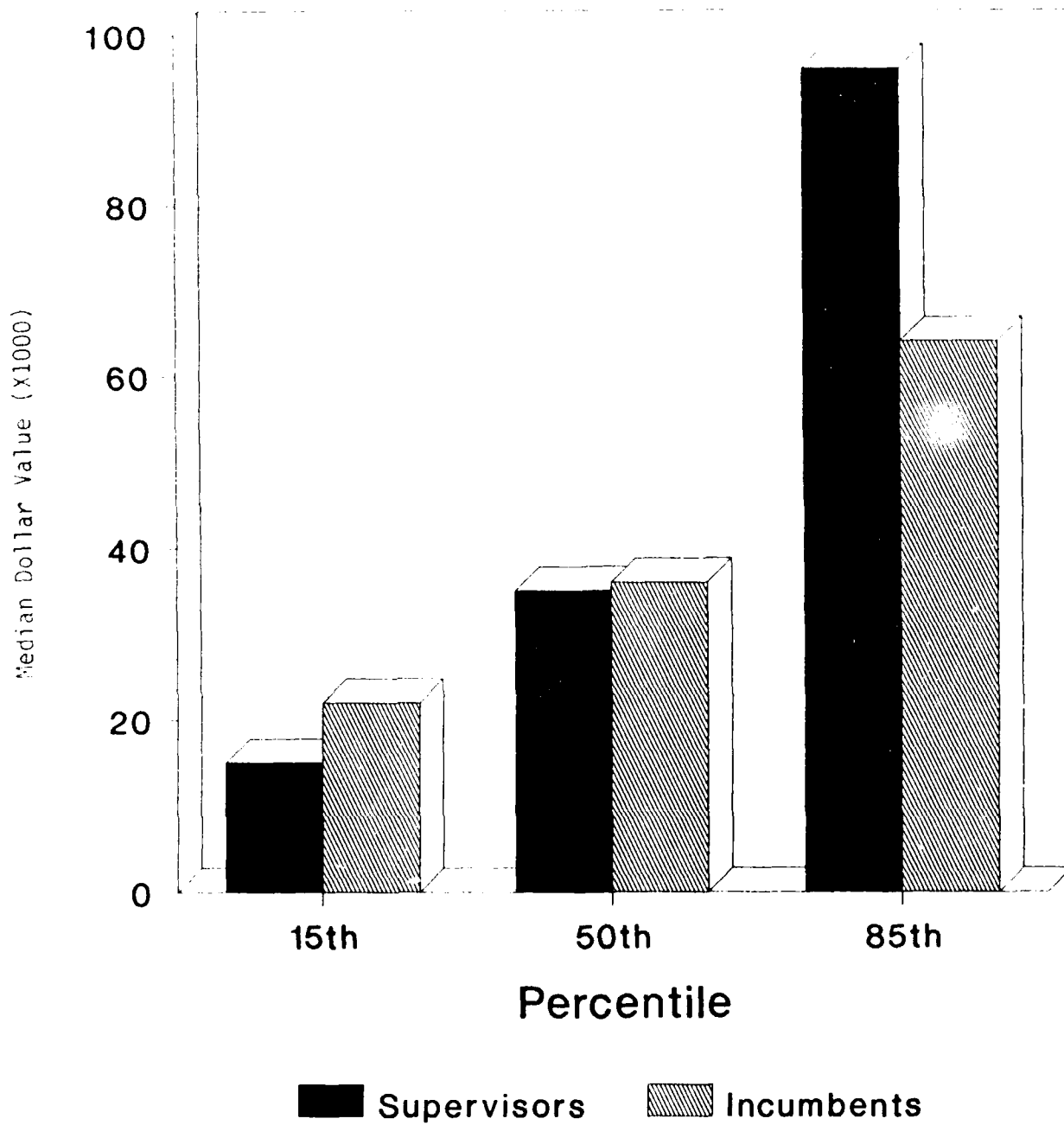
Figure A-2. Median Values for Estimates by Supervisors and
Incumbents

indicates that none of median estimates given differed significantly. (See Table A-2 for summary of Mann-Whitney U test analysis.)

Table A-2

Results of Mann-Whitney U Test Analysis

	U	n_1	n_2	$\underline{P}$
15th Percentile	114	13	18	.05
50th Percentile	120.5	13	18	.05
85th Percentile	87.5	13	18	.05

The interpretation of the results for the current study depend on whether or not a parametric analysis is viewed as appropriate for these data. If it is felt that the assumptions required by the $\underline{t}$ test are met, then it may be concluded that supervisors and incumbents do not render similar judgments of the worth of workers (except at the 15th percentile), with supervisors providing significantly and substantially higher estimates of the dollar value of average and superior workers. If this is true, then it would likely be argued to accept the estimates of supervisors as being more valid, since they should have a more comprehensive picture of the role of workers in the organization as a whole and should be in a better position to compare the output of workers of different performance levels.

However, if it is felt that the assumptions required of parametric analysis are not met, then it would be concluded that supervisors and incumbents provide similar estimates of the value of workers at the performance levels tested. This would allow researchers to view estimates of incumbents with greater credibility, and make utility analysis more amenable to application to smaller organizations which lack the number of supervisors needed to provide accurate SD_y estimates.

If the parametric analysis is accepted, what can be made of the SD_y estimates obtained by the two groups? Hunter and Schmidt (1983) reported that most estimates of SD_y obtained using their procedure fall within a range of 40 to 70 percent of the mean salary of the job incumbents. This approach is known as the "Salary Percentage Technique," and provides a basis for evaluating the estimates obtained in the current study. Information on the actual salary of the incumbent sample in the present study was not obtained.

However, their average GS rating was approximately a mid-level GS-11. The pay for this grade is $32,700 per year. Using the salary percentage technique, then, yields an expected value of SD_y of $13,080 to $22,890. Both of these values are substantially less than SD_y estimates obtained by either the supervisors or incumbents in the current study. This disparity may be a reflection of the skewed nature of the distribution of performance estimates. Alternatively, the larger estimates obtained in the current study may be related to the nature of the job performed by these incumbents. Previous published studies of SD_y, with one exception (Eaton et al., 1985), have all examined workers from private enterprise and most have been blue-collar workers. Additionally, other researchers (e.g., Weekley et al., 1985), report that the salary percentage technique yields estimates of SD_y that are substantially less than those obtained by the original Schmidt et al. (1979) procedure.

Further research comparing estimates of supervisors and incumbents is needed. The population of workers sampled in the current study were highly educated (most with Masters' degrees, many with Doctorates') specialists in behavioral sciences. The current sample of incumbents had considerable experience at their jobs. Moreover, the job of task scientist is multifaceted and often involves regular interaction with supervisors on the planning and execution of research projects and related tasks. It may be that somewhat different results would be obtained if the sample used consisted of workers with less diverse jobs who were also less educated.

It may also be that the methods used in this study are not the most accurate way of generating the desired estimates. Tuttle et al. (1985) described a procedure based on group dynamics principles used to develop productivity indexes for organizations. This method could also be used to generate SD_y estimates. In the standard Schmidt et al. (1979) procedure, respondents produce their estimates independently, and performance estimates are obtained by calculating group averages. A more structured approach using groups of supervisors and incumbents, utilizing methods described by Tuttle et al. might provide even more accurate estimates.