Utility Analysis for Decisions in Human Resource Management

Working Paper 88-21

John W. Boudreau
Associate Professor

NYSSILR--Cornell University
393 Ives Hall
Ithaca, New York 14851-0952
(607) 255-2273

Draft of a Chapter for the
Handbook of Industrial-Organizational Psychology
Edited by Marvin Dunnette
to be published by Consulting Psychologist Press
Palo Alto, California

December, 1988

This research was carried out with support from the U.S. Army Research Institute, contract SFRC #MDA903-87-K-0001. the views, opinions and/or findings contained in this chapter are those of the author and should not be construed as an Official Department of the Army policy, or decision.

Introduction

The questions studied by Industrial/Organizational (I/O) psychologists are closely linked to the decisions facing managers of people in organizations. Whether they be line managers, human resource management staff, or organizational psychologists, managers of human resources must make decisions about issues affecting the employment relationship--hiring, training, compensation, performance appraisal, and so on--that draw on theories of human work behavior. Analogously, I/O psychologists, as well as other social scientists, find the organizational environment a rich source of information for advancing knowledge and testing employment-related theories. Both scientists and managers benefit from the knowledge gained about the behaviors of individuals in the work place, who can then search for ways to apply that knowledge to achieve individual and organizational outcomes of efficiency and equitable employment.

The similarity of interests between I/O psychologists and human resource management (HRM) professionals has produced some close collaborative relationships, e.g., the many psychologists who consult for industry, conduct studies designed to support HRM decisions, or, through their work, influence the direction of employment policies. Still, the HRM functions of organizations typically lack the influence and visibility of other management functions such as marketing, finance, and operations. The literature for HRM professionals routinely laments the slow implementation of HRM programs in organizations, even though these programs have gained wide acceptance by scientists (cf. Jain & Murray, 1984), and they admonish and instruct these professionals to "sell" their programs by emphasizing their effects on attainment of organizational goals (Bolda, 1985, Fitz-Ens, 1984; Gow, 1985, Jain & Murray, 1984, Sheppeck & Cohen, 1985). With increased competition and evidence from the United States and abroad that competitive organizations are likely to manage their people differently, HRM personnel are more frequently expected to justify their contributions to the employer and to account for their existence.

One must question whether the lack of influence and slow implementation of HRM programs is a rational response by organizations. Could it be that behavioral theories and findings are relevant only to the scientific community and have such little relevance to organizational decisions and outcomes that they can be ignored by successful organizations? If the theories and findings are relevant, then how should they be communicated to decision makers? Do decisions that consider social science evidence produce greater organizational success, and, if so, are the successes great enough to justify the resources necessary to generate and apply the evidence?

This chapter will discuss utility analysis (UA), which attempts to answer such questions by focusing on decisions about human resources. Utility analysis refers to the process that describes, predicts and/or explains what determines the usefulness or desirability of decision options, and examines how that information affects decisions. In HRM and I/O psychology, the focus lies on decisions involving employment relationships and employee behaviors. Thus, I/O psychologists use the term utility analysis to refer to a specific set of models that reflect the consequences, usually performance-related, of programs designed to enhance the value of the work force to the employing organization.

Utility analysis offers great potential for enhancing the link between the theories and findings of I/O psychological research and the human resource decisions of organizational managers. To achieve this potential, however, UA research and applications must proceed from a framework that recognizes the

broad effects of such decisions on the work force and the organization. Such a framework requires an expansive view of the decision tasks facing managers of people in organizations, a view that recognizes the contributions, limits and implicit assumptions not only of psychological models, but of models from other social sciences as well. The UA framework provides both a rationale and a significant new direction for an integration between the science and practice of I/O psychology and other scientific disciplines relevant to organizational employment decisions. This chapter is intended as a step toward such an integrative framework. Thus, it will not only review and describe UA theories and applications, but will propose new and integrative directions that have received little attention. UA research must certainly acknowledge the considerations of related disciplines such as economics, management and sociology. But as a true theory of organizational decision making, it provides a mechanism to go beyond simple acknowledgement, to achieve a mechanism for truly interdisciplinary approaches to employment decisions.

Chapter Outline

This chapter comprises ten sections. The first section introduces and establishes some fundamental concepts, including the nature of utility models, decision options, attributes and payoff functions. It shows where UA models fit within the broader domain of decision models. It further establishes some ground rules guiding subsequent sections.

The second section outlines the historical development of concepts integral to utility analysis, the roots of which can be traced to the earliest stages of I/O psychological research. Not only does this historical outline provide some basic concepts for those not familiar with UA research, it also identifies certain fundamental concepts and assumptions essential to understanding utility analysis, which are sometimes ignored or forgotten in more recent theoretical developments.

The third section summarizes findings from previous studies revealing the effect of I/O psychological interventions on work force consequences. The fourth section critically reviews the research topic commanding the greatest attention to date--measuring the dollar value of performance variability.

The fifth section examines UA research from the perspective of information theory, by examining the role of risk and uncertainty in decision making. Such a perspective suggests that UA models can improve decisions even when information is severely lacking. Methods for identifying risk and uncertainty are described, as well as a technique for identifying when additional information is valuable. The role of UA research in defining statistical and substantive significance is also discussed.

The sixth section presents enhancements to the traditional selection utility models. These include incorporating financial/economic considerations, "intangible" factors such as equal employment opportunity and affirmative action, and the role of "constituencies" (Tsui, 1984; 1987; Tsui & Gomez-Mejia, 1988) in evaluating the usefulness of HRM programs. This section also shows how UA research can link I/O psychology and labor economics. It suggests that UA offers a mechanism for truly interdisciplinary approaches to employment issues, but that this demands that UA models reflect economic considerations, stocks and flows.

The eighth section discusses the role of utility analysis in describing consequences of programs that affect the "stock" of existing employees by altering the characteristics of the work force or work situation. Recent research is reviewed, suggesting implications for extending utility analysis research to important new areas.

The ninth section presents a unified utility model reflecting outcomes of HRM decisions affect the composition of the work force by changing the "flows" of employees into, through, and out of organizations--an employee movement utility model. Important links are proposed between recruitment, selection, turnover, and internal staffing. Empirical simulation analyses are described that suggest that the actual consequences of HRM decisions are likely to reach far beyond those reflected in current models addressing only the consequences of selection. It demonstrates the need for a fully integrated framework for considering the consequences of changing both the stocks and flows of employees, which can lead to greater synergy in planning and implementing employment programs.

Finally, the tenth section presents a matrix to guide future UA research, emphasizing the need to move beyond selection models and measurement issues, and toward a broader understanding of HRM program decision making.

Concepts and Definitions

Utility Analysis as a Subclass of Multiattribute Utility Analysis

Multiattribute utility (MAU) models are "decision aids" (edwards, 1977; Einhorn & McCoach, 1977; Einhorn, Kleinmuntz & Kleinmuntz, 1977; Fischer, 1976; Huber, 1980; Keeney & Raiffa, 1976) that provide tools for describing, predicting and explaining decisions. MAU models share certain characteristics and requirements. To apply such models, one must:

(1) Identify a set of *decision options* that represent the alternative programs or courses of action under consideration;

(2) Identify a set of *attributes* that reflect the characteristics of the options that are important because they represent the things that matter to the decision makers and/or the relevant constituents.

(3) Measure the level of each attribute produced by each option using a *utility scale* for each attribute;

(4) Combine the attribute values for each option using a *payoff function* reflecting the weight given each attribute and combination rules for deriving an overall total utility value for each option.

Insert Table 1 Here

Table 1 illustrates an extremely simple application of MAU analysis. Suppose productivity is below desired levels among sales people. Two *decision options* might be identified, involving two different training programs called Program A and Program B. Three *attributes* are of interest: (a) Effects on sales levels, (b) Resources required to develop and implement the program; and (c) Effects on sales person job satisfaction. Attributes (a) and (b) use a *utility scale* of dollars, while Attribute (c) uses a rating scale from 1 to 7. The *payoff function* consists of multiplying the level of attributes (a) by 1, multiplying the level of attribute (b) by -1, multiplying the level of attribute (c) by 3,000, and adding the results to produce a total utility value. We could construct a Multiattribute utility matrix like that shown in Table

1, with the cells of the matrix containing the expected level of each attribute for each option, and the total utility values below each option computed using the payoff function. Although Program B has the higher first-year dollar payoff, the high weight given to attribute (c), Job Satisfaction, combined with Program B's lower Job Satisfaction cause it to attain a lower utility value than Program A, and thus to be less preferred. Obviously, MAU models can encompass a variety of decision options, numerous and diverse sets of attributes reflecting many different constituents, and very complex payoff functions, but they generally share the characteristics shown in the simple example of Table 1.

MAU models can assist decision makers in overcoming "limits on rationality" (March & Simon, 1958) by providing a simplified, structured framework within which to consider a number of decision options. Huber (1980, pp. 61-62) identifies five advantages of MAU models over less systematic and structured decision systems:

(1) Because they make explicit a view of the decision situation, they help to identify the inadequacies of the corresponding implicit, mental model;

(2) The attributes contained in such models serve as reminders of the information needed for consideration of each alternative;

(3) The informational displays and models used in the mathematical model serve to organize external memories;

(4) They allow the aggregation of large amounts of information in a prescribed and systematic manner; and

(5) They facilitate communication and support to be gained from constituencies.

As a subclass of MAU models, UA models also serve as decision aids, and can provide the advantages listed above. Unfortunately, very little theoretical or empirical research has approached utility analysis from this decision-making perspective. Nonetheless, a keen appreciation of the role of UA models in the decision process suggests some very different research questions and directions. These will be emphasized throughout the chapter. Unlike the generic MAU model describe in Table 1, UA models focus on a particular type of decision option, a restricted set of attributes, and a defined mathematical formula for attribute weights and combination rules. The next sections examine these MAU components, and how they apply to UA models.

The Decision Options: HRM Productivity-Enhancement Programs

Any MAU model requires a focus of analysis--the decision options considered. For example, an MAU model for deciding where to build a new hospital might focus on options reflecting different types of facilities, combined with different locations, combined with different service offerings. Each combination would constitute a decision option. Utility analysis has focused on HRM programs designed to enhance work force productivity. Such programs include selection testing, recruitment, training, and compensation--all of which affect the organizational value of the work force, whether they are explicitly chosen using decision models or evolve implicitly over time (Milkovich & Boudreau, 1988). Utility analysis involves describing, predicting and explaining the consequences of such program options, their desirability, and the decision processes leading to choices among them. Thus, while the focus of UA is more specific than generic MAU models, it covers a wide array of options relevant to organizational

goals. As we shall see, the majority of UA research has focused only on selection programs, but we now have the theoretical models to apply to virtually any HRM program.

Decisions about individuals versus decisions about programs. Utility analysis models might seem to focus on decisions about individuals, rather than programs. For example Cascio (1980, p. 128) stated, "all personnel decisions can be characterized identically. In the first place there is an individual about whom a decision is required. Based on certain information about the individual (for example performance appraisals, assessment center ratings, a disciplinary report), decision makers may elect to pursue various alternative courses of action." In MAU terms, the decision options are different courses of action for each individual.

However, closer examination shows that UA models are intended to apply to decisions about the *programs* that guide the countless decisions about individuals made by human resource managers. The options under consideration are the procedures, rules or "strategies" (Cronbach & Gleser, 1965, p. 9) meant to be used with many individuals, and evaluated by their "total contribution when applied to a large number of decisions" (p. 23). Decisions about whom to hire depend on what programs of recruitment and testing have been chosen to generate applicants and information about them. Decisions about how much to pay individuals depend on what compensation programs and rules have been chosen for that work force. Decisions about assigning individuals to new jobs depend on what career development and training programs have been chosen to generate skills and forecast future needs. Thus, UA models focus on the more strategic and tactical decisions about programs, rather than the operational decisions about each individual.

Because program decisions affect many individuals throughout their tenure with the organization, the impact of even a single program decision on future work force consequences can be quite large. A selection program that affects the hiring decisions for 1,000 people, each of whom stays for 5 years affects 5,000 person-years of organizational behavior. If a more correct program decision produces even a modest work force quality increase of $10 per person-year, it's impact can be $50,000. Of course, this also suggests that the consequences of wrong decisions have large potential negative effects. Utility analysis uses information from social and behavioral sciences to attempt to improve such important decisions.

Two types of programs addressed by UA models. It is useful to group the variety of HRM programs that can be addressed by UA models according to whether they affect employee "flows" or employee "stocks". First, programs affecting employee *flows* change the composition or membership of the work force through "employee movement" (Boudreau, 1988; Boudreau & Berger, 1985a, 1985b; Milkovich & Boudreau, 1988). For example, *selection programs* allow additions to be made to the work force, *retention programs* determine which employees are retained when separations take place; and *internal staffing programs* determine which employees move between positions within an organization (Milkovich & Boudreau, 1988, Chapters 10-13). UA models applied to such programs focus on the process used to determine which individuals are chosen to move or to remain, and the program's consequences reflect the effects of having a different set of employees in the work force. UA models are typically applied to decisions about this type of program, with external selection programs receiving the greatest attention.

Second, programs affecting the employee *stock* change the characteristics of the existing set of

employees, in their current positions. For example, *training programs* operate by altering knowledge, skills, attitudes, or other employee characteristics; *Compensation and reward programs* operate by altering the relationship between behaviors/outcomes and rewards; *Performance feedback and goal setting programs* operate by altering employee perceptions of the consequences of their behaviors. Such programs work to the extent that they lead to different behaviors by existing employees, which lead to more valuable organizational outcomes. UA models address decisions about such programs by focusing on options representing different kinds of programs affecting the stock of existing employees.

The Attributes of Programs In UA Models

Once a set of decision options is defined, MAU models specify the set of attributes reflecting the outcomes of concern to the decision makers and relevant constituents, and the level of each attribute achieved by each decision option. For example, the decision about where to build a hospital might include attributes as diverse as the environmental impact of the facility, speed of treatment in emergencies, and impact on local property values, reflecting the concerns of constituents as diverse as community planners, potential patients, nearby property owners and the future medical staff.

UA models focus on decisions about HRM programs, so the attribute set is more focused, but still quite broad. Cronbach & Gleser (1965, p. 22) defined the attribute domain as "all the consequences of a given decision that concern the person making the decision (or the institution he represents)." HRM program attributes may be placed in two categories--efficiency and equity (Milkovich & Boudreau, 1988): *Efficiency* attributes reflect the organization's ability to "maximize outputs while minimizing inputs", such as labor costs, job performance, sales volumes, revenues, profits, market share and various financial/economic indicators of organizational strength. *Effectiveness* attributes reflect the "perceived fairness" of organizational procedures and outcomes, such as employee attitudes, labor relations, minority and female representation, compliance with legal requirements, and community relations.

To date, most UA research and applications have focused on a very small set of efficiency-related attributes reflecting the productivity consequences of HRM program decisions. Although UA models can become mathematically complex, all existing UA models reflect just three basic attributes (Boudreau, 1984c, 1986, 1988; Boudreau & Berger, 1985a, 1985b; Milkovich & Boudreau, 1988):

(1) *Quantity*; the number of employees and time periods affected by the consequences of program options;

(2) *Quality*; the average effect of the program options on work force value, on a per-person, per-time-period basis;

(3) *Cost*; the resources required to implement and maintain the program option.

The program options addressed by UA models encompass a potentially large set of attributes reflecting both efficiency and equity, but existing model development has focused on a subset of the efficiency-related attributes reflecting program costs and employee productivity. Thus, like all models, UA models simplify reality by omitting or ignoring some factors. Models, by definition, are deficient because it is impossible to accurately reflect all the potential attributes affected by decisions. As we shall see, examining the nature of the attributes that are and should be included in utility models is one of the most critical issues facing UA research. Defining the domain of appropriate attributes offers fruitful

opportunities for further debate and development. We will discuss these opportunities in some detail as we review existing research.

The Utility Scale for Attributes in UA Models

With the attributes identified, an MAU model must assign a value for each attribute in each decision option. This requires establishing a utility scale for each attribute, as well as determining the particular level of each attribute associated with each decision option. For example, in deciding where to locate a new hospital, the attributes are quite diverse (e.g., environmental impact, speed of treatment, facility cost, community satisfaction, etc.), and might be measured in units as diverse as dollars, time, number of complaints, ratings or rankings.

UA models focus on HRM programs, and therefore face a more limited set of attributes. Yet, even the relatively simple example in Table 1 had attributes measured in dollars (Costs and Productivity) as well as ratings (Job Satisfaction). UA models can potentially include a variety of efficiency and equity-related attributes, requiring diverse payoff scales. However, most UA models have focused primarily on productivity-related outcomes, striving to measure them in units relevant to managerial decisions. Attributes reflecting Quantity are usually measured in person-years, and those reflecting Cost are usually measured in dollars. The appropriate scale for the Quality attributes has been subject to some debate, as we shall see, but the majority of research has been devoted to scaling Quality in dollars per person-year.

Attaching a level of each attribute to each option often reflects a process using both subjective and objective information. When evaluating past programs, it may be possible to determine the actual levels of each attribute achieved by different options. But UA models are planning tools, used to anticipate future consequences and support current decisions, so attaching attribute levels involves predictions and forecasts. Indeed, one major motivation for UA models was to better express statistical forecasts in terms understandable to managers. The predictive nature of attribute measurement means that utility estimates possess uncertainty and risk. While uncertainty and risk take prominence in general MAU research, UA research has largely ignored them. As we shall see, mechanisms exist to promote further research in this important area.

The choice of attribute utility scales and derivation of attribute levels is important, and has received too little attention in UA research. Throughout this chapter we will highlight controversies where additional debate and research attention can be fruitful.

Combining Attributes Using a Payoff Function for UA Models

The fourth component of an MAU model is the payoff function, which specifies how the attribute levels are to be combined into an overall utility value. Deciding where to locate a new hospital might produce very diverse attributes measured on very different scales (e.g., dollars, time, and ratings/rankings). Payoff functions for such decisions must specify both the weights attached to each attribute level, as well as the rules for combining the weighted attribute levels to produce an overall utility value. Such rules might range from a simple numerical weighting and addition of the weighted values, to more complex non-linear weighting schemes and quadratic combination rules.

Because UA models focus on decisions about HRM programs, their attributes and payoff scales are more limited, and the payoff functions are often simpler. Still, any payoff function must reflect both the importance of each attribute and its underlying scale. The example in Table 1 adopted a relatively simple combination rule that takes the difference between increased productivity and costs, and then adds the Job Satisfaction level multiplied by 3,000. Obviously, the choice of weights and combination rules can have large effects of resulting utility values, and should reflect the values of the decision makers and relevant constituencies.

UA research has usually focused on productivity-related outcomes, and thus has adopted payoff functions reflecting dollar-valued productivity and program costs. The payoff function may be considered a variant of the cost-volume-profit models used in other managerial decisions to invest resources. The utility of an HRM program option is derived by subtracting Cost from the product of Quantity times Quality, with the program exhibiting the largest positive difference being preferred.

It is typical to refer to UA models as cost-benefit analysis models, and to categorize attributes as either Costs or Benefits. Simply put, Costs represent attributes that reduce overall utility values, while Benefits represent attributes that increase overall utility values. Depending on the decision, a given attribute (e.g., reduced employee separations) may represent either a cost or a benefit. Rather than attempt a classification, this chapter will proceed from the more general position that costs and benefits are defined by the attributes, their utility scales, and the payoff functions used to combine them. It is appropriate to question whether such a payoff function is adequate or even appropriate to UA research, and we will explore this issue at length.

Summary

UA research is a subclass of more general MAU research, and the structure of MAU models provides a useful framework for organizing and understanding UA models. As we have seen, UA models reflect a set of decision options, attributes, utility scales and payoff functions, just as any MAU model does. UA models have historically focused on a particular set of options (usually selection programs), attributes (Quality, Quantity and Cost), utility scales (Dollars) and payoff functions (Quantity times Quality, minus Cost). Measuring the payoff in UA research has been characterized as the "Achilles' heel" of UA research (Cronbach & Gleser, 1965, p. 121). As we have seen, such measurement reflects three MAU components: The attributes included; the utility scale used to measure them and attach a value to each option; and the payoff function specifying the combination rules across attributes. These components reflect implicit and explicit assumptions about the appropriate decision makers, constituents and consequences to be considered. Throughout the chapter, we will use these MAU concepts to organize and analyze existing and needed future UA research.

We have also seen that UA models, like all models, strike a balance between simplicity and realism. All UA models are deficient by definition, and much research debate has centered on whether and how to reduce that deficiency. But we will never develop a UA model that completely reflects all relevant attributes with perfect accuracy. Does this mean that UA research is unlikely to provide any real information about the effects of HRM program decisions on organizations? If the ultimate objective of UA is to *measure the impact* of program decisions on organizations, then the answer might be "yes", and

we could declare a moratorium on UA research. However, like all MAU models, UA models are decision aids, not just measurement tools. A decision aid's usefulness lies in *its ability to describe, predict, explain and improve decisions.* Such value is assessed by asking whether the model allows the best decision to be made with the given body of information, whether it helps to determine if gathering more information would permit better decisions, and whether it helps to determine how much different decision procedures contribute to decision quality (Cronbach & Gleser, p. 21). Depending on the cost and value of the next best alternative decision aid, even a very deficient or inaccurate UA model might prove effective in improving decision processes or outcomes. Thus, this chapter will approach UA research less from a measurement perspective and more from a decision making perspective.

Historical Development of Utility Analysis Models[1]

Though utility analysis is applicable to virtually every HRM program decision, present models resulted from a concern with selection (and later, placement or classification) decisions. Indeed UA models can be characterized as responses to the inadequacies of traditional measurement and test theory in expressing the usefulness of tests.

> "The traditional theory views the test as a measuring instrument intended to assign accurate numerical values to some quantitative attribute of the individual. It therefore stresses, as the prime value, precision of measurement and estimation. The roots of this theory lie in surveying and astronomy, where quantitative determinations are the chief aim. In pure science it is reasonable to regard the value of a measurement as proportional to its ability to reduce uncertainty about the true value of some quantity. The mean square error is a useful index of measuring power. There is little basis for contending that one error is more serious than another of equal magnitude when locating stars or determining melting points: measurement theory is unobjectionable when applied to such appropriate situations.
>
> "In practical testing, however, a quantitative estimate is not the real desideratum. A choice between two or more discrete treatments must be made. The tester is to allocate each person to the proper category, and accuracy of measurement is valuable only insofar as it aids in this qualitative decision. ... Measurement theory appears suitable without modification when the scale is considered in the abstract, without reference to any particular application. As soon as the scale is intended for use in a restricted context, that context influences our evaluation of the scale." Cronbach and Gleser (1965, pp. 135-136).

Therefore, the history of UA will be discussed from a decision-making perspective, focusing on the contributions and implications of UA developments for describing, predicting explaining and enhancing decision processes and outcomes. Because the vast majority of research has emphasized the selection utility model, this will be the focus on the discussion. In this model, the option set involves using a test versus random selection (or choosing between two selection tests), and the utility value reflects only the effects of selection on the first job to which one group of selectees are assigned. Later sections will describe more recent developments that extend utility analysis beyond selection.

---

[1]This section emphasizes developments that set the stage for more recent research and future research directions. Much of this material is drawn from Boudreau (1987, in press). Other historical summaries can be found in Cronbach and Gleser (1965, chapter 4), Hunter and Schmidt (1982), Cascio (1982, 1987).

Defining the Payoff Based on the Validity Coefficient

Description of the Model. The attribute of selection tests that has the longest history is the validity coefficient, or correlation between a predictor measure and some criterion measure of subsequent behavior, usually expressed as $r_{x,y}$. Classical measurement theory suggested this concept as a measure of the "goodness" of a test in predicting subsequent behavior. In addition to the validity coefficient itself, two translations are most commonly cited (e.g., Cronbach & Gleser, 1965, chapter 4; Hunter & Schmidt, 1982), both of them lead to the conclusion that only relatively large differences in the validity coefficient produce important differences in the value of a test. First, one can translate the validity coefficient into the *index of forecasting efficiency* (symbolized as E) using Equation 1 below.

$$E = 1 - (1-r^2_{x,y})^{1/2} \tag{1}$$

This index, emphasized by early statistical texts (e.g., Kelley, 1923; Hull, 1928), indicates the proportionate reduction in the standard error of criterion scores predicted by the test, compared to the standard error of criterion scores predicted using only the group mean. Second, the *coefficient of determination*, or the squared validity coefficient appeared as early as 1928 in Hull's text, and reflects the proportion of variance shared by the predictor and the criterion.

Obviously, very large increases in validity are required to substantially increase these indexes. As Cronbach and Gleser (1965, p. 31) noted, "the index of forecasting efficiency describes a test correlating .50 with the criterion as predicting only 13% better than chance; the coefficient of determination describes the same test as accounting for 25% of the variance in outcome." Yet, correlations as high as .50 may be rare. In short, using these indexes, it appeared that very great improvements in testing would be necessary to have any substantial effect on organizational outcomes.

Evaluation From a Decision-Theory Perspective. As MAU models of a test's usefulness for decisions, such formulas are deficient. Only one attribute of the selection system is considered--the accuracy of prediction, expressed as the shared variance between two normally-distributed variables. From a decision-making perspective, the usefulness of a selection system depends on its ability to provide information that will improve decisions, where decision improvements are measured in terms of valued decision outcomes. Therefore, this model omits selection system attributes such as the quality of the existing selection system, the effect of the proposed selection system information on actual decisions, and the impact of those effects on valued consequences.

The utility scale for attaching attribute values to each option is a statistic that measures squared deviations from a predicted linear function. Thus, both positive and negative prediction deviations from the linear function are equally undesirable. This implies that a decision maker would consider overpredicting a qualified candidate's future performance just as costly as underpredicting it. In fact, of course, the important deviations from predictions are the ones that result in selection errors (i.e., selecting a candidate who should not have been hired, and/or failing to select a candidate who should have been hired). These models adopt an implicit payoff function that assigns equal value (or loss) to inaccurate predictions at all points in the predictor-criterion space (Wesman, 1953). Because there is only one attribute, there is no payoff function for combining different attributes. The statistic serves as the sole

utility value.

These models fail to reflect most of the three basic program attributes (i.e., quantity, quality and cost). They reflect neither the quantity of time periods affected by the selection decisions, nor the quantity of employees affected in each time period. Though these models reflect one statistical quality of the predictor, this is only indirect evidence of that predictor's effect on work force quality. Finally, they fail to acknowledge the costs to develop and apply tests. Though the deficiencies inherent in these formulas are apparent when viewed from a decision-making perspective, the fundamental notion of expressing the relationship between a predictor and a criterion in terms of the correlation coefficient remains a basic building block of UA models. Future models began to explore ways to imbed the correlation coefficient within a set of decision attributes that made it easier to interpret.

Defining Payoff Based on the Success Ratio

Description of the UA model. These utility models reflected a new utility concept--the *success ratio*, or proportion of selected employees who subsequently succeed. Taylor and Russell (1939) proposed a UA model designed to reflect the fact that the usefulness of a test depends on the situation in which it is used. Unlike models based solely on the validity coefficient, the Taylor-Russell model reflects three attributes of the decision situation: (1) the *validity coefficient*; (2) the *base rate*, scaled as the proportion of applicants who would be successful if selection were made without the proposed predictor; and (3) the *selection ratio*, scaled as the proportion of applicants falling above the hiring cut-off on the predictor.

The payoff function combining these attributes assumes a linear, homoskedastic, and bivariate normal relationship between the predictor (or predictor composite) and the criterion, and uses formulas for the area under a normal curve to derive the success ratio. The Taylor-Russell model assumes fixed-treatment selection (i.e., each applicant will either be hired or rejected) and a dichotomous criterion (i.e., selectee value is classified as either successful or unsuccessful). Total utility under the Taylor-Russell approach is the difference between the success ratio predicted for a specific combination of validity, selection ratio, and base rate, minus the success ratio that would result without using the proposed predictor (i.e., the base rate). The combination producing the greatest improvement is the preferred option. Taylor and Russell derived extensive tables indicating the predicted success ratio for various combinations of base rates, validity coefficients, and selection ratios (Cascio, 1987 reprints these tables).

To apply the model, a decision maker would choose the criterion (e.g., job performance) and determine the level of criterion performance that represents the dividing line between acceptable and unacceptable (or successful and unsuccessful) selectees. Then, s/he would estimate the current base rate implied by this criterion level in the population of individuals on which the proposed predictor would be applied (perhaps by examining the current success rate, if the predictor is to be added to those already in use). Finally, s/he would use the Taylor-Russell tables to determine the expected change in the success ratio under various assumptions about validity and selection ratios.

Detailed summaries of the Taylor-Russell model are provided elsewhere (Taylor & Russell, 1939; Cascio, 1980; Cascio, 1987, chapter 7). According to the Taylor-Russell tables, when other parameters are held constant: (1) higher validities produce more improved success ratios (because the more linear the

relationship, the smaller the area of the distribution lying in the false-positive or false-negative region); (2) lower selection ratios produce more improved success ratios (because lower selection ratios mean more "choosy" selection decisions, and the predictor scores of selectees lie closer to the upper tail of the predictor distribution); (3) base rates closer to .50 produce more improved success ratios (because as one approaches a base rate of zero, none of the applicants can succeed, so selection has less value; as one approaches a base rate of 1.0, all applicants can succeed even without selection, so selection has less value).

Evaluation from a decision-making perspective. The Taylor-Russell model reflects three attributes, rather than only the validity coefficient, but it still provides a limited description of selection program utility. Like its predecessors, this model ignores both the number of employees affected and the number of time periods during which that effect will last. the model's measure of Quality (proportion successful) is also troublesome because it does not reflect the natural units of value such as sales, productivity or reduced errors. Finally, the model excludes attributes reflecting program costs (Cascio, 1980, 1987), but cost differences will occur, especially as the selection ratio is changed by screening more/fewer applicants.

Scaling the base rate as a dichotomous criterion (i.e., success/fail) will often lose information because the value of performance is not equal at all points above the satisfactory level, nor at all points below the unsatisfactory level (Cascio, 1982, p. 135; Hunter & Schmidt, 1982, p. 235, Cronbach & Gleser, 1965, pp. 123-124, 138). More typically, performance differences exist within the two groups, so a continuous criterion scale could be more appropriate. Cascio (1982, 1987) suggests it may be more appropriate for truly dichotomous criteria (e.g., turnover occurrences), or where output differences above the acceptable level do not change benefits (e.g., clerical or technician's tasks), or where such differences are unmeasurable (e.g., nursing, teaching, credit counseling). Combining the attributes by assuming bivariate normality and linearity implied in the payoff function may also be unrealistic in some selection situations.

Some have proposed that the choice of the criterion cutoff is "arbitrary" (Cascio, 1982, p. 133; Hunter & Schmidt, 1982, p. 235; Schmidt, Hunter, McKenzie & Muldrow, 1979) because it is set by management consensus or because objective information on which to base such a decision is rarely available, and that changing this "arbitrary" cutoff will change the base rate, and thus substantially alter the conclusions from the model. If indeed there is no objective method of setting the performance cutoff, then the Taylor-Russell utility model is inappropriate. However, the concept of a criterion cutoff is not arbitrary, nor does the Taylor-Russell model imply that arbitrary changes in that cutoff are to be regarded as legitimate methods of enhancing the success ratio. Rather, the criterion cutoff (and the base rate it implies) should be based on the relationship between the selection situation (i.e., the level of minimally-acceptable criterion levels) and the applicant population (i.e., the proportion of the population that would exceed that level if hired). This concept is essential to evaluating the effects of recruitment on staffing utility, and should not be abandoned by labelling it "arbitrary."

Variations on the Taylor-Russell model. The models discussed next add program costs to the model and/or redefine the attribute utility scales to include dollar-scaled consequences of different selection mistakes. Cascio (1980, p. 35) noted that Smith (1948) provides a method of adjusting the Taylor-Russell results to reflect pre-existing selection ratios and validities. Technically, if current-employee characteristics are used as inputs to the model, this assumes that current employees are similar

to the applicant population to which the new predictor system will be applied. This is appropriate if one is adding the new predictor to an existing set of predictors, and if the base rate, selection ratio and validity coefficient reflect this situation. However, if the predictor will replace a previous predictor, then one should use the table corresponding to the observed success ratio given current selection ratios and validities.

Sands (1973) proposed the CAPER model (Cost of Attaining PErsonnel Requirements). It's payoff objective is a recruiting and selection strategy that minimizes total costs of recruiting, inducting, selecting and training enough new hires to meet a set quota of satisfactory employees. This model adds the notion of the costs involved in hiring and recruiting, but it suffers from the same weaknesses in the payoff function as the Taylor-Russell model.

Mahoney and England (1965) noted that success and failure probabilities on a new predictor are conditional on the success and failure probabilities existing in the applicant population after previous methods have been employed. They proposed that previous decision rules (Stone & Kendall, 1956) and Meehl and Rosen (1955) implicitly assumed that these probabilities are .50. They defined the cost of selection mistakes to include not only false positives (hires who do not succeed) as in the Taylor-Russell model, but also false negatives (rejected applicants who would have succeeded), which could be important where high-quality rejected applicants are hired by competitors and reduce the organization's competitive advantage (Guttman and Raju, 1965). Mahoney and England simulated various values for the selection ratio on the proposed predictor, the selection ratio on previous predictors, the existing failure probability, the failure probability under the new system, the ratio of recruitment costs to selection mistakes (i.e., 05, .10, .30, and .50), and the ratio of predictor costs to selection mistakes (i.e., .05, .10, and .30). They concluded that a new predictor's value exceeds its cost only when the probability of selection mistakes is quite low (i.e., less than .30), and that "the opportunities for developing and installing predictive measures that are worth the additional cost appear relatively restricted" (p. 375). This conclusion conflicts with more recent evidence based on newer UA models. One explanation is that their ratios of costs to mistakes were really quite large. Because selection mistakes may reduce performance for many years, and predictors can cost less than a few hundred dollars, it is difficult to imagine situations where the ratio would exceed .10, and it would probably frequently fall below .01.

Hunter and Schmidt (1982) also note a number of studies based on the notion of a dichotomous criterion (Alf & Dorfman, 1967; Curtis, 1966; Darlington & Stauffer, 1966; Schmidt, 1974). While obviously deficient, if the dichotomous-criterion model is easier to implement, then a more complex model (such as those discussed subsequently) must prove its value based on its ability to improve decisions over the simpler model.

Defining Payoff Based on the Standardized Criterion Level

Description of the Utility Model. The major criticism of the Taylor-Russell model was that it used a dichotomous notion of total utility (i.e., success/fail) that failed to reflect the true range of variation in selectee performance. The next version of the selection utility model attempted to remedy this by scaling total utility on a continuous scale. Brogden (1946a, 1946b) showed that the correlation coefficient is the proportion of maximum predictive value obtained using a predictor (where maximum predictive value is

what would hypothetically be obtained if the criterion itself were used to select employees). Moreover, he used the principles of linear regression to demonstrate the relationship between the correlation coefficient and increases in a criterion (measured on a continuous scale). Brogden's logic serves as the basic building block for virtually all subsequent UA research.

Assuming a linear relationship between criterion scores (y) and predictor scores (x), the best, linear unbiased estimate of the criterion score associated with a predictor score is:

$$E(y) = A + B(x) \tag{2}$$

The intercept (A) and the slope (B) of this line reflect the linear relationship between x and y as well as the units in which each of them was originally scaled. However, because predictor and criterion scales vary from study to study, it is difficult to compare these parameters or to use them in a general model. However, if we transform both the y and x variables into standardized (Z-score) units (i.e., Z_x and Z_y), we can write Equation 2 as follows:

$$Z_y = (r_{x,y})(Z_x) \tag{3}$$

Therefore, if we knew the average standardized predictor score of a selected group of applicants (i.e., $\bar{Z}_x$), our best prediction of the average standardized criterion score of the selected group (i.e., $\bar{Z}_y$) would be the product of the validity coefficient and the standardized predictor score, as shown in Equation 4.

$$\bar{Z}_y = (r_{x,y})(\bar{Z}_x) \tag{4}$$

The validity coefficient was well established. One way to estimate the average standardized test score of the selected group would be to actually observe the value after applying a selection device. However, Kelley (1923) suggested that if one assumes that the predictor scores are normally distributed and that one ranks applicants by test score and selects from the top down, then the average standardized predictor score is a function of the proportion of the applicant population falling above the predictor cutoff score (i.e., the selection ratio). However, if one assumes the predictor is normally distributed, then Equation 4 holds only if one also assumes normally distributed criterion scores as well.

Brogden (1949, Equation 6) and Cronbach and Gleser (1965, p. 309) make use of this approach to derive their models. If we symbolize the "ordinate of the normal distribution" corresponding to the standardized predictor cutoff score as lambda (i.e.,λ), and the selection ratio corresponding to the standardized predictor cutoff as SR (it has also been symbolized by the greek letter ϕ), then, Equation 3 can be re-written:

$$\bar{Z}_y = (r_{x,y})(\lambda / SR) \tag{5}$$

The "ordinate of the normal distribution" is an important variable, multiplicatively related to the average standardized predictor score, and a statistically sophisticated concept. It is sufficient, however, to

understand that the ordinate is simply a mathematical value that is completely determined by the selection ratio, and (when divided by the selection ratio) can be used to compute the expected average standardized predictor score of those selected using that selection ratio. Computing the relationship between the selection ratio and the average standardized predictor score of the selected group was made even easier by Naylor and Shine (1965) who computed extensive tables showing, for each selection ratio, the corresponding standardized predictor cutoff score, the corresponding ordinate of the normal distribution, and the corresponding average standardized predictor score under the assumptions noted above.

Evaluation From a Decision-Making Perspective. The attributes of the Naylor-Shine utility model still include the validity coefficient and the selection ratio, but their contributions appear through a different payoff function. The validity coefficient now has a constant multiplicative effect on expected standardized criterion levels at all selection ratios. The selection ratio still reflects the "choosiness" of the selection program, but is now used to derive a new attribute--the standardized predictor score of selectees ($\bar{Z}_s$). The lower the selection ratio, the greater the predictor score required to meet selection standards, and the greater the resulting standardized predictor score of those meeting the selection standard. Unlike the Taylor-Russell model, the base rate no longer appears as an attribute because the standardization used to go from Equation 2 to Equation 3 defines the average value of the applicant pool as zero.

The utility model of Equations 3, 4 and 5 addresses one shortcoming of the Taylor-Russell model by using a total utility concept based on a continuous scale. Utility is defined as the difference in average standardized criterion score between those selected using a test and those selected without it. The translation from Equation 2 to Equation 3 requires that the utility concept be expressed in standardized units, which are difficult to interpret in units more natural to the decision process (e.g., performance ratings, dollars, units produced, reduced costs, etc.). Also, this utility concept reflects only the difference between the average standardized criterion score of those selected using the predictor and the average standardized criterion score that would be obtained through selection without the predictor. The absolute utility from the program is not computed, only the increment over not using the predictor. Finally, the model assumes that selection occurs as if applicants were ranked based on their predictor scores, and then hired from the top down until the desired selection ratio is reached, which may or may not describe a realistic selection approach.

Considering the three basic utility model concepts (i.e., quantity, quality and cost), the Naylor-Shine utility model reflects the effects of selection on per-person, per-time-period quality on a continuous criterion. The quantity of employees and the number of time periods affected are not explicitly reflected, nor are program costs. However, the next section will demonstrate that they can be easily added.

Defining Payoff in Terms of Dollar-Valued Criterion Levels

Description of the utility model. The most obvious drawback of the Naylor-Shine UA model is that standardized criterion levels are difficult to interpret in "real" units. Correlation-based statistics are useful when predictor and criterion scales vary from study to study (as in selection research) because the standardized scale underlying the correlation coefficient allows direct comparison between studies. However, when one wishes evaluate utility in units relevant to a particular situation, such standardized scales create problems.

Actual selection makers usually face choices among selection strategies. Each strategy carries with it a set of activities required for development and implementation, as well as the possibility of various outcomes resulting from more accurate selection. The development and implementation activities are often expressed as costs (i.e., the value of required resources) usually scaled in dollars. Therefore, the question becomes whether it is worthwhile to spend that dollar amount to produce the selection consequences. With a standardized criterion scale, one must ask questions such as: "Is it worth spending $10,000 to select 50 people per year, in order to obtain a criterion level 0.5 standard deviations greater than what we would obtain without the predictor?" Many HRM managers may not even be familiar with the concept of a standard deviation. They would find it difficult to attach a dollar value to a 0.5 standard deviation increase in the criterion, particularly because the decision makers may never actually observe the population of applicants to which the predictor would be applied.

These limitations suggest modifying the UA model for selection to be expressed in dollar terms. Both Brogden (1946a, 1946b, 1949) and Cronbach and Gleser (1965, pp. 308-309) eventually derived their utility formulas in terms of "payoff" (often expressed in dollars) rather than standardized criterion scores. Also, they both included the concept of costs. In fact, Brogden's (1949) treatment explicitly computed utility values in dollar terms, and attempted to derive guidelines for testing costs. Brogden and Taylor's (1950) formula introduced a scaling factor to translate standardized criterion levels into dollar terms. The scaling factor is the dollar value of a one-standard-deviation difference in criterion level (e.g., σ_y, σ_ϵ, and SD_y). The cost attribute is usually expressed as the cost to administer the predictor to a single applicant (usually symbolized as C). Finally, the utility value is symbolized as $\Delta \bar{U}$, to indicate that it represents the *difference* between the dollar payoff from selection without the predictor and the dollar payoff from selection with the predictor (this is usually called the "incremental" utility of the predictor). The resulting utility Equation may be written as Equation 6.

$$\Delta \bar{U} = (SD_y)(r_{x,y})(\bar{Z}_x) - C/SR \qquad (6)$$

The per-applicant cost (C) is divided by the selection ratio (SR) to reflect total cost of obtaining each applicant (e.g., if the selection ratio is .50, then one must test 2 applicants to find each selectee, and the testing cost per selectee is 2 times the cost per applicant). Sometimes, the entire formula is simply written in terms of per-selectee outcomes, and the symbol C is used to denote the cost per selectee. Equation 6 depicts the incremental dollar value ($\bar{U}$) produced by using a predictor (x) in a population of applicants where the validity coefficient is $r_{x,y}$; a one-standard-deviation difference in dollar valued criterion levels equals SD_y; the average standardized predictor score of those selected is $\bar{Z}_x$; and the per-selectee cost of using the predictor equals (C/SR).

To express the total gain from using the predictor to select N_s selectees, we simply multiply by the number selected, change the symbol for incremental utility from $\Delta \bar{U}$ to ΔU, and multiply the per-applicant cost by the number of applicants (N_{app}) as shown in Equation 7.

$$\Delta U = (N_s)(SD_y)(r_{x,y})(\bar{Z}_x) - (C)(N_{app}) \qquad (7)$$

This formula is stated in terms of the per-selectee incremental criterion level multiplied by the number selected (Brogden, 1949), but Cronbach & Gleser (1957, 1965) derived their formulas in terms of the per-applicant incremental criterion level, which can be derived by dividing the total utility by the number of applicants, expressed as Equation 8:

$$\Delta\, U/\text{applicant} = (N_s/N_{app})(SD_y)(r_{x,y})(\lambda/SR) - C \qquad (8)$$

In Equation 8, the term (/*SR*) has been substituted for the average standardized test score.

If we note that the term (N_s/N_{app}) equals the selection ratio (*SR*), we can cancel terms and produce the Cronbach & Gleser equation for per-applicant incremental dollar-valued utility, shown in Equation 9.

$$\Delta\, U/\text{applicant} = (SD_y)(r_{x,y})(\lambda) - C \qquad (9)$$

Cronbach and Gleser (1965, p. 39) also developed a utility formula for comparing the usefulness of two tests (one producing lower validity and lower costs, the other producing higher validity with higher costs). They recommended computing the difference in utility between the two tests, which simply involves substituting the difference in validities for $r_{x,y}$ and the difference in costs for *C* in Equations 6 through 9.

Recent embellishments of the B-C-G model have explicitly incorporated the duration of the effects of better-selecting one group by multiplying the value component (i.e., the component containing $r_{x,y}$) by the expected average tenure of the hired group (i.e., *T*).

These equations have come to be known as the Brogden-Cronbach-Gleser (B-C-G) selection utility model.

Evaluation From a Decision Making Perspective. The B-C-G selection utility model reflects the same attributes as Naylor-Shine, but adds the attribute of dollar-valued criterion standard deviation (i.e., SD_y). It also adds attributes reflecting the duration of selection effects (i.e., *T*) and the program costs (i.e., *C*). In terms of the overall utility concept, scaling the per-person, per-time-period incremental criterion level in dollars seems more in keeping with organizational objectives evaluated in dollars. The model continues to focus on the incremental utility added by using the predictor versus not using it. Thus, all utility values are scaled as differences from an unknown utility level that would be attained without the predictor.

Table 2 summarizes the results of the Schmidt, Hunter, McKenzie & Muldrow (1979) application of the B-C-G model for entry-level computer programmers in the U.S. Government. The application reflects the consequences of hiring one group of 618 computer programmers, assumed to stay for 9.69 years and then leave. The utility computation is organized according to the Quantity, Quality and Cost components developed earlier. Unlike earlier models, the B-C-G model incorporates all three concepts. Although modifications to this basic model have recently been proposed, the B-C-G model has been the dominant framework for studying HRM program utility.

Insert Table 2 Here

Assumptions of the B-C-G Selection Utility Model. The payoff function translating the attributes into utility values reflects certain assumptions (Cronbach and Gleser, 1965, p. 307):

(1) Decisions focus on an indefinitely large population of "all applicants after screening by any procedure which is presently in use and will continue to be used." Thus, the appropriate population for deriving the validity coefficient, SD_y, and the selection ratio depends on the decision situation. If one is contemplating adding a new procedure to a group of previously-used procedures, then it is the "incremental" validity coefficient and the pre-screened population SD_y and selection ratio that count. If, however, one contemplates replacing an old procedure with a new one, then the parameters should reflect the unscreened population.

(2) Regarding any person, one can decide only to accept or reject them. Thus, no adaptive decisions can be made to reflect different predictor scores (e.g., training those who achieve a moderately high score, in order to bring them to minimally qualified levels).

(3) Predictor (or "test") scores are standardized to zero mean and unit standard deviation.

(4) The "payoff" resulting from accepting a person has a linear regression on predictor score, and the predictor is scored so that validity is positive.

(5) The payoff resulting from rejecting a person is unrelated to predictor score, and is set to zero. Thus, it is assumed that the organization is indifferent to the consequences of rejection, regardless of the qualification level of those rejected.

(6) The average cost of administering the predictor ("testing") a person is C, and C is greater than zero. In practice, it is often easier to separate this cost into its fixed components (i.e., one-time development costs) and its variable components (i.e., ongoing per-applicant administration costs). Also, if the decision options include the possibility of testing more or fewer applicants, then the differences in recruiting costs necessary to provide different quantities of applicants should be included (Boudreau & Rynes, 1985; Hunter & Schmidt, 1982. p. 241).

(7) The strategy for selection is to set a predictor cutoff score so that the desired proportion (selection ratio) of the applicant group falls above it. All applicants scoring above that level are accepted, those below it are rejected. This is equivalent to ranking applicants from the top down on predictor score, and then hiring by rank order until the established quota of new hires is met (assuming there are no rejected offers). When such hiring does not take place, the effective selection ratio is different.

Validity for the Dollar-Valued Criterion versus Proxy Criteria. Adopting the SD_y scaling factor carries with it some assumptions about observed and implied correlations. There is no clear consensus regarding the meaning of y (we will discuss this after reviewing empirical attempts to estimate SD_y), but it undoubtedly reflects a wide variety of employee behaviors and attributes that affect dollar-valued organizational outcomes. If it were possible to measure such a criterion, the best utility model would simply reflect the regression equation of y on the predictor score (similar to Equation 2). In reality, however, predictors are not validated on such a dollar-valued criterion because it cannot be directly measured. Thus, UA models substitute a validity coefficient ($r_{z,y}$) that reflects the regression of one or more proxy criteria (e.g., performance ratings, tenure, sales, etc.) on the predictor, with all variables standardized to Z-scores. This substitution not only assumes that dollar-valued criterion levels are linearly related to predictor scores, but that the proxy criterion and unobserved dollar-valued productivity are also linearly related.

Hunter & Schmidt (1982) and Schmidt, Hunter, McKenzie and Muldrow (1979) proposed that many mistakenly believe that utility equations are of no value unless the data exactly fit the linear homoskedastic model, and all marginal distributions are normal. They state that the B-C-G model only introduces the normality assumption for "derivational convenience" (Hunter & Schmidt, p. 243) because it provides an exact relationship between the selection ratio and the average standard test score of selectees. They further state that the only critical assumption is a linear homoskedastic relationship between

predictor and criterion, and they present evidence in support of this relationship using observable proxy criteria. They argue (Schmidt, et al., 1979, p. 613) that the relationship between the proxy and employee dollar value will be linear or that ceiling effects on proxy measures will make the correlation between the proxy and the predictor underestimate the correlation between the dollar value and the predictor. Raju, Burke & Normand (1987) note that equality between these correlations implies a correlation close to unity between the proxy and dollar value. Evidence of low correlations between typical and maximum performance (Sackett, Zedeck & Fogli, 1988) suggests that validity might differ depending on whether dollar value reflects typical or maximum performance. Evidence that test validity may be higher at higher predictor score ranges (Lee & Foley, 1986) suggests that the level of test scores in the applicant population may also moderate incremental utility values. We have no direct evidence regarding the correlation between predictors and dollar-valued utility, but small estimation errors may not seriously reduce the utility model's ability to improve decisions (compared to less sophisticated decision models).

Hunter & Schmidt (1982) and Schmidt, et al. (1979) also state it is a mistake to believe that test validities are situationally specific, making application of utility analysis possible only when a criterion-related validity study has been performed in the particular situation. "Validity generalization" research (Hunter, Schmidt, & Jackson, 1982), which allows data from many studies to be analyzed together, strongly suggests that much of the variability in validity coefficients observed across studies is due to artifacts of the studies (e.g., different sample sizes, different criterion reliabilities, different range restrictions, etc.), rather than real differences in the predictor-criterion relationship. Moreover, the variability that does remain after correcting for these artifacts may be so small that it does not seriously reduce the utility model's ability to enhance decisions. Indeed, it has been suggested that selection validities might usefully be estimated by experts or even less experienced judges (Schmidt, Hunter, Croll & McKenzie, 1983; Hirsh, Schmidt, & Hunter, 1986).

The role of testing costs. Both Brogden (1949) and Cronbach and Gleser (1965) portrayed testing costs as a fundamental characteristic of their UA models. The cost attribute recognizes that improvements in validity and/or reductions in selection ratios are not infinitely desirable. At some point, additional costs will offset gains from improved employee quality. At the extreme, it seems unlikely that pursuit of selection systems with validities close to unity would be cost effective. Cronbach and Gleser (1965) discussed the importance of the cost of testing in deciding between competing predictors (p. 39), in determining optimum test length (p. 323-324) and in determining the optimum predictor cutoff score (p. 308). Brogden (1949) noted that considering the cost of testing can show that higher selection ratios (i.e., testing fewer applicants and being less choosy) can be preferable to low ones if the testing cost is high. He concluded that "the ratio of cost of testing to the product of the validity coefficient and SD_y (in dollar units) should not exceed .10. It would be desirable to hold it below .05" (p. 177). Below .05, lower selection ratios contribute to higher utility. Brogden presented an example for hosiery loopers, and used a one-year payoff duration. His analysis indicated that testing costs above $5.00 per person decreased utility at low selection ratios. As we shall see, in actual applications SD_y (per person, per year) is usually fairly large compared to testing costs. Moreover, testing costs occur once, but benefits usually accrue over the selected group's tenure. Thus, the value of SD_y when considered over the group's tenure is larger, and testing costs become less likely to detract from utility except at very low selection ratios (Hunter & Schmidt, 1982, p. 240).

However, omitting such costs from the UA model, or assuming they equal zero removes much of the justification for dollar-valued utility estimates. Faced with a costless selection procedure, any non-negative validity coefficient must produce positive utility because N_s, SD_y, and $\bar{Z}_s$ must always be positive (see Equation 7), so a utility model based solely on the sign of the validity coefficient would suffice. In reality, implementing employee selection programs may require time, energy and other resources that could be used to implement other managerial programs. If so, the lost value of the foregone programs represents a legitimate cost of the selection program, so actual costs (i.e., the true investment necessary to implement the selection program) may be much higher than testing costs alone.

The Appropriate Applicant Population. Cronbach and Gleser (1965, p. 34-35) stated "we use 'validity' subsequently to refer to a correlation computed on men who have been screened on whatever *a priori* information is in use and will continue to be available," and that the appropriate utility calculation depends on the situation in which the selection program will be used. They noted three possible situations: (1) all prior information will continue to be used and the new system added to it; (2) the new system will be substituted for some of the prior information; or (3) a composite of previous and new information will be used. Each has different implications for the UA model. The *incremental* program contribution is key. Moreover, any new program should be compared to the efficient use of information already available. Some have concluded that the B-C-G model presumes "concurrent validity" (e.g., Cascio, 1980, p. 39), but the precise assumption is that selection devices be evaluated in light of the conditions under which they will actually be applied. In fact, such conditions may indicate a population less restricted than current applicants (e.g., if the predictor is to be substituted for an existing predictor and applied to unscreened applicants) or it may imply a more restricted population (e.g., if the new predictor is not only going to be added to an existing screening system, but the existing system will be improved before adding it).

Mueser and Maloney (1987) argue that validity coefficients used in utility analysis may be severely overstated if test validation data arise from situations where composite predictors are already in use, and validity estimates fail to correct for multivariate restriction in range on those composites versus test scores. Applicant population characteristics also affect the selection ratio and SD_y (Boudreau & Rynes, 1985). Determining the appropriate population requires assumptions that have important implications for integrating additional staffing processes (e.g., recruitment, turnover) into the selection utility model, as discussed subsequently.

Several enhancements to the B-C-G model have been proposed and applied, but the vast majority of empirical UA research has focused on selection systems, using the B-C-G model. Therefore, we will now review empirical research based on the B-C-G model, and discuss the enhancements and their empirical findings subsequently. Existing UA applications have produced two kinds of findings: Evidence of the utility values from selection programs, and evidence of differences in SD_y.

Utility Values for Selection Programs

Insert Table 3 Here

Table 3 summarizes the utility values reported in existing literature. Twenty-one empirical studies were located, with utility values for 48 interventions. Two of these studies reported results for non-selection activities (Florin-Thuma & Boudreau, 1987; Mathieu & Leonard, 1987), but the utility model used by these studies is sufficiently similar to include their results here (the utility model for non-selection programs will be discussed subsequently). Several studies used enhanced utility models incorporating additional attributes (Burke & Frederick, 1986; Cronshaw, et al., 1986; Florin-Thuma & Boudreau, 1987; Mathieu & Leonard, 1987; Rich & Boudreau, 1987). The symbols at the top of the table stand for the parameters of the utility model. N_s is the number selected or treated; T is the tenure of the selectees, or F is the analysis period; SR is the selection ratio; $\overline{Z_x}$ is the estimated average standardized predictor score of selectees; $r_{x,y}$ is the validity coefficient; SD_y is the dollar-valued standard deviation of performance among the applicant population (or the untreated group for non-selection programs); $Cost$ is the total program cost; and ΔU is the total utility of the program over all treated employees and all time periods. The last two columns contain an equation expressing total utility as a function of SD_y, as well as the "Break-Even" (B-E) SD_y value necessary for the program's total returns to equal its costs (Boudreau, 1984, and as discussed subsequently).

The overwhelming conclusion from Table 3, is that selection programs pay off handsomely. Virtually every study has produced dollar-valued payoffs that clearly exceeded costs (Van Naersson, 1963 did report that improved selection to reduce accidents did not pay off because accident frequency and damages were already quite low). Even the earliest studies that reported utility per person (or per person, per hour in the case of Roche, 1961) found that the payoff exceeded costs. In studies dealing with more employees, multiple-year tenure, and occurring more recently (which include the effects of inflation) the utility estimates are always positive, and have ranged into the millions (e.g., Schmidt, et al., 1979; Cascio & Ramos, 1986; Cronshaw, Alexander, Weisner & Barrick, 1986; Schmidt, Hunter, Outerbridge & Trattner, 1986; Rich & Boudreau, 1987). The clear positive payoff from selection programs remains evident in studies with both small and large SD_y values, and with selection ratios as high as 81% (Van Naersson, 1963). The largest utility values occur where large numbers of individuals are affected by the program, and N_s is large.

Many of the studies were designed to examine whether substituting a more-valid selection method for a less-valid one (usually an interview) produced greater dollar-valued payoff (Cascio & Silbey, 1979; Schmidt, et al., 1979; Ledvinka, Simonet, Neiner & Kruse, 1983; Schmidt, et al., 1984; Cascio & Ramos, 1986; Burke & Frederick, 1986; Rich & Boudreau, 1987). In these cases, Table 3 reports a utility value for each selection method separately and for the difference between them. As shown, in every case the more valid (and usually more costly) selection procedure produced the greater estimated utility. However, even the interview produced positive utility despite its cost and low validity. This is not an argument in favor of less-valid selection, but it does illustrate that even modestly valid selection programs may produce substantial utility values.

The utility values measured by the B-C-G model appear to be quite high. Moreover, the estimated costs of improved selection are often minuscule compared to the benefits. A $10 per-applicant testing cost might produce over five times greater validity if the PAT is substituted for the interview (Schmidt, et al., 1979). As noted earlier, testing costs are unlikely to reflect the full range of resources required to

implement top-down selection based on more valid predictors, but even inflating costs by a factor of 10 or 100 often would not change the positive utility values. According to these findings, the economic impact of improved selection might well surpass many more traditional investment opportunities, such as plant, equipment, marketing, financial, etc. Such a conclusion seems at odds with the observations reported earlier (and verified by many HR managers) that human resource management's contribution is often ignored, that HR issues are not considered in organizational planning, and that debate continues over whether HR activities are really an appropriate use for organizational resources. This suggests several important research issues regarding the decision processes of managers, and how payoff information about HRM programs is interpreted and evaluated. However, only one research issue has received substantial attention in the I/O psychology literature--the accuracy, psychometric quality, and proper measurement method for SD_y.

Research Measuring SD_y

The standard deviation of dollar-valued job performance in the applicant population (SD_y) was characterized as the "Achilles' Heel" of utility analysis by Cronbach and Gleser (1965, p. 121). The amount of recent research aimed at estimating this elusive concept suggests that many of today's UA researchers agree. Moreover, researchers often regard accurate SD_y measurement as fundamental for useful UA research (Burke & Frederick, 1984, 1985; Weekley, et al., 1985; DeSimone, Alexander & Cronshaw, 1986; Greer & Cascio, 1988). This section reviews this research from a decision-theory perspective, focusing its contribution toward better describing, predicting, explaining and enhancing HRM program decisions. The review will focus on four decisions that must be made in measuring SD_y: (1) the definition of utility (i.e., y); (2) the focus population; (3) the setting of study; and (4) the operational measurement method used. From a decision-making perspective, these decisions should be guided by how well the analysis will describe, explain, predict and/or enhance HRM program decisions. SD_y measurement is fundamentally linked to the decision context in which the measure is applied. However, existing research seldom explores whether utility analysis and SD_y measures affect decisions or reflect decision maker objectives and values. Instead, research tends to pit one measure against others, often advocating a particular measure, with quality usually defined psychometrically (e.g., consistency with other measures, reliability across estimators, consistency with distributional assumptions). Such research provides interesting tests of measurement principles. However, its value in describing, predicting, explaining and enhancing decision processes is difficult to determine, because most SD_y studies don't reflect actual decisions.

UA models were spawned by the limitations of measurement theory and correlational statistics to fully capture the decision processes and consequences of selection programs (see Cronbach & Gleser, 1965, p. 135-137). It is ironic that the resurgence in UA research should focus once again on measurement issues. UA research (including SD_y measurement research) should focus clearly on the ultimate purpose of UA models to describe, predict, explain and enhance decision processes. This focus is frequently absent in the rush to develop and test each new SD_y measure.

Insert Table 4 Here

Table 4 summarizes existing SD_y measurement research. The studies are arranged chronologically, with each study described in terms of its setting and sample, utility scale, estimation method and research findings. The research findings are described in terms of the mean SD_y estimate derived (i.e., *MEAN*), the standard deviation of the SD_y estimate in the sample (i.e., *SD*), the standard error of the mean SD_y estimate (*SE*), the percent of average salary represented by average SD_y, and the percent of the mean payoff estimate represented by average SD_y. In studies estimating dollar-valued payoff (*y*) directly (e.g., Desimone, et al., 1986; Day & Edwards, 1987; Greer & Cascio, 1987; Edwards, et al., 1988) Table 4 reports the actual average payoff estimate (Mean *y*) as well as the estimate of the standard deviation (SD_y). Thirty-four studies were located, producing over 100 individual SD_y estimates (the results shown in Table 4 sometimes represent averages of groups of estimates derived by the authors). The trend in research activity is clearly evident, with only five studies between 1953 and 1978 but with 29 studies between 1979 and 1988.

The Utility Scale

Viewing UA models as special cases of MAU models suggests that utility will be largely in the eye of the beholder. Generic MAU models often rely on subjectively-scaled payoff functions, measured by having decision makers indicate their preferences for different levels of certain attributes on a scale of zero to 100 (see Huber, 1980 for a number of examples). The nature of the decision situation and the decision makers determine the payoff function, and the MAU model makes the values, assumptions and priorities explicit.

Because UA models serve (in part) to translate HRM program consequences into units that managers understand (usually dollars), UA research has used more focused utility scales. Equations 6 through 9 clearly indicate that the utility scale reflects the expected average increase in employee dollar value due to the selection program, on a per-person, per-year basis. Little consensus exists regarding the meaning of "dollar value." The variety of criteria available for evaluating HRM decisions (see Smith, 1976, Milkovich & Boudreau, 1988 or other introductory textbooks) virtually guarantees that different researchers will adopt diverse definitions of the payoff scale, as we shall see. Still, a broad concept of the utility scale must be maintained, to avoid basing decisions on a dangerously narrow perspective.

Defining the meaning and scale of the criterion is important to advancing UA research and applications (Day & Edwards, 1987; Desimone, et al., 1986; Steffy & Maurer, 1988). While a single definition will not apply in all situations, this section will attempt to develop a framework for categorizing existing definitions and developing new ones. At the very least, such a framework will allow researchers to clearly identify the objectives and assumptions underlying various studies. Eventually, it may aid understanding of the appropriate utility scales for different situations.

The Utility Concept. A general definition of payoff for utility analysis is "*all consequences of a given decision that concern the person making the decision (or the institution he represents)*" (Cronbach & Gleser, 1965, p. 22; Boudreau, 1987; in press). Some of these consequences may be positively valued (often referred to as "benefits") and some may be negatively valued (often referred to as "costs"). This definition suggests several implications:

(1) Utility may reflect different outcomes (e.g., productivity increases, labor cost reductions, affirmative action goal attainment, improved organization image, consistency with fundamental organizational beliefs, high levels of financial return, etc.) consistent with the desires and objectives of decision makers and the constituents they serve (see Cronbach & Gleser, 1965, p. 23).

(2) Utility measures should reflect the decision context. Work force quality improvements will have different value depending on how they are used by the organization. For example, improved work force quality may be used to increase the number of units produced, to increase their average quality, or to reduce costs. As we will discuss subsequently, the dollar implications of these strategies are quite different.

(3) Increased measurement precision will not always improve decision quality. For example, if a simple (and inexpensive) payoff measure implies positive program utility, but a more accurate (but also more expensive) measure leads to the same decision, then the more accurate measure does not improve decision quality.

A Framework for the Payoff Scale Defining SD_y. The payoff scales in UA research usually focus on the economic consequences of programs that increase labor force quality. Yet, there are many ways an organization might employ a higher-quality work force (Cronbach & Gleser, 1965, p. 23), and the payoff from HRM programs depends on how the organization uses the quality enhancements they produce. The Quantity, Quality and Cost concepts introduced earlier provide a useful framework. Among other objectives, organizations aim to increase economic value. They can do this through some combination of: (1) producing high *quality* per unit of product sold (in order to generate high prices/revenue from selling each product unit); (2) producing and selling a large *quantity* of units; and (3) producing units at low *cost* (i.e., the value of resources in their next best alternative use, Levin, 1983). This framework applies even to non-profit organizations, whose objective is to provide the maximum quantity of service at the minimum cost, with a target profit of zero. This implies three general uses for improved labor force quality: (1) Increasing the quantity of production; (2) Increasing the quality of production; and (3) Reducing production costs. Managers may choose to use labor force quality increases in any combinations of these ways. A payoff scale defined in terms of economic profit can reflect any or all of these uses. A payoff scale defined in terms of quantity will be sufficient to reflect uses affecting product quantity, but it will fail to reflect the other two, and so on. Payoff scales reflecting revenue enhancements (through higher quality or quantity) and cost reductions dominate the UA literature, though profit-based scales are emerging.

Payoff as cost reduction. Most of the earliest UA applications focused on cost reduction from improved selection. Doppelt and Bennett (1953) focused on reductions in training costs. Van Naersson (1963) focused on reductions in driving accident and training costs. Lee and Booth (1974) and Schmidt & Hoffman (1973) focused on reduced costs of replacement (e.g., recruitment, selection and hiring costs) when turnover is reduced. More recently, Eaton, Wing & Mitchell (1985) and Mitchell, Eaton and Wing (1985) measured payoff in terms of the avoided costs of additional tanks to achieve a given military objective. Boudreau (1983a, p. 555) noted that utility models including variable costs applied to situations where cost reduction is an important selection outcome. Schmidt and Hunter (1983, p. 413) noted that increases in work force productivity might be used to reduce "payroll costs" by producing the same amount of output with a smaller number of employees. Arnold, Rauschenberger, Soubel & Guion (1982), DeSimone, et al. (1986) and Schmidt, et al. (1986) emphasized cost reduction from hiring fewer employees to do the same amount of work. These payoff functions are also consistent with the "behavioral costing" approach to HRM program analysis described by Cascio (1982, 1987) in which HRM

program effects are evaluated according to their ability to reduce costs associated with undesirable employee behaviors. A few authors (Mahoney & England, 1965; Sands, 1973) have incorporated not only the costs of replacing employees, but also the costs of false negatives (i.e., costs of mistakenly rejecting applicants who would have been successful if hired).

Cost-based payoff functions reflect an important element of economic payoff, but they can be misleading in those situations where programs that reduce costs also reduce revenue. For example, improved selection may identify employees who stay longer and reduce separation expenses, but if they stay because they are mediocre performers and have few employment opportunities, the reduction in replacement costs may be offset by a reduction in productivity. Although this danger is less apparent with a payoff scale reflecting reduced training time costs (because training success is likely to positively relate to subsequent job performance), training cost reductions may understate selection utility. Where cost reduction is the dominant consideration, cost reduction alone may represent a useful payoff scale. However, its deficiencies have led researchers to explore further options.

Payoff as the "value of output as sold". Schmidt, et al. (1979) proposed an SD_y measure that asked estimators to consider the "yearly value of products and services", and the "cost of having an outside firm provide these products." This payoff scale reflects the product of price and quantity sold, or the "sales value" (Boudreau, 1983a) of productivity. Hunter and Schmidt (1982, pp. 268-269) interpreted the payoff function as the value of "output as sold," or what the employer "charges the customer." As Table 4 indicates, much research has focused on similar payoff scales (Cascio & Silbey ,1979; Bobko, et al., 1983; Ledvinka, et al., 1983; Burke & Frederick, 1984; Schmidt, et al., 1984; Wroten, 1984; Bolda, 1985; Burke, 1985; Eaton, Wing & Lau, 1985; Eaton, Wing & Mitchell, 1985; Eulberg, O'Connor & Peters, 1985; Mitchell, Eaton & Wing, 1985; Reilly & Smither, 1985; Weekley, et al., 1985; Burke & Frederick, 1986; Cascio & Ramos, 1986; Cronshaw, et al., 1986; DeSimone, et al., 1986; Schmidt, et al., 1986; Day & Edwards, 1987; Greer & Cascio, 1987; Mathieu & Leonard, 1987; Rich & Boudreau, 1987; Edwards, 1988).

The "sales value" payoff scale implies that the appropriate benefit from improved HRM programs is the increased revenue generated by higher-quality employees. Its widespread adoption reflects, in part, the strong endorsement it originally received. For example, Hunter & Schmidt (1982) characterized Roche's payoff definition (contribution to company profits) as "deficient on a logical basis", because it subtracted costs of production from the value of output as sold. This view is difficult to reconcile with the general payoff definition originally proposed by Brogden & Taylor (1950) and Cronbach & Gleser (1965), both of whom included the notion of revenue minus costs (or simply cost reduction) as part of the payoff function.

Boudreau (1983) and Reilly & Smither (1986) proposed that the practice of asking estimators to consider both the "value of products and services" and the "cost of having an outside firm provide these products" may be confusing in an economic sense because a firm will pay an outsider a maximum of the *internal costs* of providing a service, not their value. Day & Edwards (1987) found that when the latter instruction was dropped, SD_y values were slightly higher for Account Executives, and much higher for Mechanical Foremen, though the inter-rater variability of the estimates was also higher.

As Boudreau (1983a, p. 553; 1983b; 1987; in press) noted, the value of output as sold produced by employees can be a deficient payoff definition for organizations using traditional financial investment

decision models. When other organizational investments are being evaluated based on profit contribution, evaluating HRM investments based on revenue contribution (without considering associated costs of production) can cause HRM program value to be relatively inflated. Hunter, Schmidt & Coggin (1988, p. 526) adopted a similar position, stating that "increase in the dollar value of output as sold is the most relevant index when the concern is with sales figures, total firm income, market share and so forth," noting that this is a different payoff definition from profits.

Payoff as Increased Profits. The initial attention to the payoff function for utility analysis proceeded from the notion that the payoff scale should be applicable to business decisions, and generalizable across business organizations. Brogden and Taylor (1950) proposed the "dollar criterion," providing a number of computations for dollar-valued criterion measures. All of them share notion that each unit produced (e.g., square feet of flooring laid) represents some value to the organization. That value reflects the sales revenue generated when the unit is sold, less any costs involved in producing that unit. Brogden & Taylor list a number of elements to be considered in such a criterion, including:

(1) Average value of production or service units,
(2) Quality of objects produced or service accomplished;
(3) Overhead--including rent, light, heat, cost depreciation, rental of machines and equipment, etc.;
(4) Errors, accidents, spoilage, wastage, damage to machines or equipment due to unusual wear and tear, etc.;
(5) Such factors as appearance friendliness poise, and general social effectiveness where public relations are involved;
(6) The cost of time of other personnel consumed.

Roche (1961) explicitly followed Brogden and Taylor (1950) in developing a dollar criterion that would convert "production units, errors, time or other personnel consumed, etc. into dollar units" (p. 255).

Cronbach and Gleser (1965) provided a very general payoff concept, including all consequences important to decision makers. Thus, their payoff concept is consistent with a "profit" definition, though it can encompass even broader definitions. Cronbach's comments on Roche's dissertation (Cronbach & Gleser, 1965, p. 266) seem to suggest that the concept of profit (revenue less costs) fits their definition. Indeed, Cronbach suggested a formula for hourly profit that reflected revenue less variable and fixed costs.

More recently, Cascio and Ramos (1986, p. 20) discussed the concept of "the difference between benefits and costs" as their payoff function, and Greer & Cascio (1987) used "contribution margin" to reflect a similar concept. Hunter, et al. (1988, p. 526) also endorse the profit concept, noting that "when the focus of concern is with pretax profits, that would be the most relevant index."

Reilly & Smither (1985) compared several different payoff definitions for SD_y, including profits. Their results suggest that the graduate students in their simulation differed most in their SD_y estimates when they were asked to consider "net revenue" rather then "new sales," or "overall worth". The results of Bobko, et al. (1983) may reflect a similar phenomenon, in that their sales counselor supervisors exhibited much greater variability in their SD_y estimates when attempting to estimate "yearly value to the company" rather than "total yearly dollar sales." Greer & Cascio (1987) estimated SD_y based on contribution margin, the revenue generated by better-quality workers less the costs associated with them, and found that the average value of y and SD_y were both higher than SD_y estimates derived by scaling average salary levels (CREPID), but were only slightly higher than those based on revenue (Schmidt, et al., 1979).

Summary. Although costs, sales and profits have enjoyed some attention as payoff functions, any payoff function's usefulness should be judged in terms of its ability to better describe, predict and explain and enhance decisions. Because UA models focus on the consequences of improving the quality of an organization's labor force, a fundamental consideration is how the organization uses quality improvements. In some organizations, improved work force quality may reduce costs (e.g., through reduced staffing levels), but maintain the same quantity, quality and price of output. In other applications, the quantity of output may be increased, while maintaining the quality and price. A revenue-based utility scale (e.g., "output as sold") will reflect the objectives in the latter situation but not in the former (because revenue doesn't change in the former), and vice versa for a cost-based utility scale. A profit-based utility concept encompasses the objectives of both, because quantity, quality and cost are included (though they may not vary in every decision). One can use a revenue-based payoff function in estimating SD_y, and add other parameters to the UA model so that overall utility values reflect a profit focus. In Table 3, several studies have adopted a total utility function reflecting profit contribution (Burke & Frederick, 1986; Mathieu & Leonard, 1987; Rich & Boudreau, 1987), though their SD_y measures reflected revenue increases. This financial/economic approach (Boudreau, 1983a) will be discussed subsequently.

Comparing the psychometric characteristics (e.g., inter-rater consistency, distribution shape, mean) of SD_y estimates resulting from different payoff functions can provide interesting insights about measurement, but research addressing decision processes and outcomes efforts should place the estimates within a decision context, and apply them to actual decisions. All existing payoff scales reflect a concern with productivity-based outcomes, virtually ignoring other factors that might be affected by selection decisions (e.g., community relations, work force attitudes, adherence to a code of ethics). Thus, every payoff function is deficient in some way. While deficiency is a characteristic of all models (because models simplify reality), research should address the effects of incorporating these broader outcomes into actual decisions.

Effects of Jobs Studied

A wide variety of occupations has been examined, with the occupation usually determined by the research setting presented to the researchers (SD_y studies often occur within a validation study). Different occupations should exhibit different SD_y values. Occupations in which workers exercise more discretion regarding production and/or where variation in production has large implications for organizational goals should exhibit higher SD_y values than jobs without these characteristics. However, this effect may be reduced if variability in skills and motivations among applicants is negatively related to discretion and variation in productivity. Even jobs with high discretion and variability may emerge with lower SD_y values when their employees/applicants have low ranges of skill and motivation. Most SD_y studies examine only one job, making across-job comparisons difficult (because jobs, measurement methods, settings and time periods are confounded). Five studies employed more than one job (Wroten, 1984; Eaton, Wing & Lau, 1985; Mitchell, Eaton & Wing, 1985; Day & Edwards, 1987; Mathieu & Leonard, 1987). Wroten (1984) did not statistically test the effect of jobs on SD_y estimates, but his results indicate different rankings of jobs based on SD_y level for each estimation method. Similarly, although Eaton, Wing and Lau (1985) found a significant effect of MOS (job type) for their GLOBAL estimation

technique, they found no such significant effect for the EQV technique. Nor did they find that military rank or the interaction between rank and MOS were significantly related to SD_y. Mitchell, Eaton and Wing (1985) found very similar results for Crewman and Transport Operators. Day & Edwards (1987) found higher SD_y values for Account Executives than for Mechanical Foremen, with the differences more pronounced for more subjective and global estimation methods. Mathieu & Leonard (1987) found similar SD_y levels for bank Head Tellers and Operations Managers, but substantially higher values for Branch Managers. Thus, in studies estimating SD_y in different jobs, there is mixed evidence of across-job variation in SD_y.

Hunter, Schmidt & Judiesch (1988) studied the effects of occupational complexity, defined using Hunter's (1980) system, on performance variability. Instead of examining SD_y estimates, however, they examined the ratio of the standard deviation of output to mean output (SD_p). Across many studies, they used the actual reported ratio of the standard deviation to the mean. For low or medium complexity jobs that reported only the ratio of highest-performer output to lowest-performer output, they used a formula that assumed normality Schmidt & Hunter (1983, p. 408). For some high complexity jobs (attorneys, physicians and dentists) they used the mean and standard deviation of income. They corrected observed distributions to reflect a constant time period. They conclude that for incumbents in routine clerical or blue-collar work, SD_p is about 15%, in medium complexity jobs it is 25%, and in high-complexity jobs it is 48%. For life insurance sales it was 97%, and for other sales it was 42%. After correcting for selective hiring (assuming applicants were hired using general mental ability), they estimate that the progression from low to medium to high-complexity jobs is 20% to 30% to 50%. For life insurance sales it is 123.8% and for other sales it is 54.2%. To the extent that dollar-valued productivity is linearly related to these productivity measures, one could expect SD_y to rise with job complexity as well.

Pitfalls of job descriptions as group identifiers. Every study used job titles to distinguish the group for analysis. By using job titles to identify employees holding similar job duties and tasks, existing research may be inadvertently including across-job differences in the SD_y measure. For example, although computer programmers may all hold the same job title, certain programming jobs may involve primarily transcribing flowcharts into computer code, while other programming jobs may involve designing the logic of the program (Rich & Boudreau, 1987). Clearly the latter job has more potential for both valuable positive contributions and/or costly mistakes. Yet, existing SD_y measurement methods would include both groups in the SD_y estimate. If the selection test will primarily be used to select programmers assigned as coders, this will overstate SD_y (and vice versa). Still, the Hunter, et al. (1988) study suggests that even job titles may be sufficient to detect consistent differences in SD_y according to job complexity.

The Focus Population

The focus population is the population of individuals over which variability occurs. Virtually all SD_y measurement methods focus on job incumbents. The incumbent population is most familiar to job supervisors who provide the SD_y estimates, and it is the only population on which actual output information exists. However, the incumbent population is not strictly the appropriate population of interest for most utility models.

For selection utility models, the appropriate population is the applicant population to which the

selection procedures will be applied. This population may differ from the incumbent population for a number of reasons. First, certain procedures may operate to make the incumbent population a restricted sample of applicant job performance (for example, promoting the best performers and dismissing the worst performers), as discussed by Hunter, et al. (1988) and Schmidt, et al. (1979). Such a situation would make SD_y estimated on job incumbents a downward-biased estimate of the applicant population. Second, the applicant population may change over time due to different recruitment procedures or labor market influences (Becker, 1988; Boudreau & Rynes, 1985). Such influences may operate either to increase or decrease performance variability among applicants, and produce applicant SD_y levels either higher or lower than the variability among job incumbents. Third, SD_y estimates based on job incumbents encourage estimators to consider all of the incumbents in their experience. This group includes incumbents with very different tenure levels. If performance varies with tenure, then job incumbent variability will reflect this. However, each cohort of hired applicants will have equal tenure throughout their employment, removing this source of variability within cohorts of selectees. Thus, where job tenure and performance are related, SD_y estimates based on job incumbents will include variability not present in applicants and will tend to overestimate. However, Greer & Cascio (1987) examined this possibility in a sample of beverage salesmen, and found that though wide variations in tenure existed, tenure was not significantly correlated (r=0.118) with dollar-valued output estimates. Fourth, as noted earlier, UA studies have grouped employees with similar job titles to form the focus population. If task assignments or work environments differ within the same job, the variability of performance may differ as well. This is not a problem if selected individuals are assigned to tasks and environments in the same proportion as the incumbent group. If, however, entering employees tend to be assigned to specific tasks or environments (perhaps with less chance for error), then SD_y estimates based on incumbent populations may be inaccurate reflections of the actual SD_y in the selection system (Bobko, et al., 1983; Boudreau & Rich, 1987). Most authors argue that incumbent-based SD_y estimates are conservative due to restricted range. However, there is no evidence regarding the possible biasing effects of different recruiting approaches, or different labor market conditions.

Measurement Techniques

Without doubt, the research question most addressed by existing UA research is whether using different SD_y measures produces different SD_y values. Table 4 attests to this fact, indicating that the vast majority of studies compare one SD_y estimation method to another. Authors customarily argue that because SD_y was characterized as the Achilles' Heel of utility analysis by Cronbach and Gleser (1965) and because differences in SD_y can cause such large differences in total utility estimates (because SD_y is multiplied by many other factors in the selection utility formula), it is important to develop better SD_y measures.

SD_y measurement methods fall into four categories:

(1) *Cost Accounting*, which refers to methods in which accounting principles are used to attach a value to units of performance or output for each individual, with the standard deviation of these individual performance values representing SD_y (e.g., Roche, 1961; Van Naersson, 1963; Schmidt & Hoffman, 1973; Lee & Booth, 1973; Greer & Cascio, 1987);

(2) *Global Estimation*, where experts are asked to provide estimates of the total yearly dollar valued performance at two, three, or four percentiles of an hypothetical performance distribution, and average differences between these percentile estimates represent SD_y (e.g., Cascio & Silbey, 1979; Schmidt, et al., 1979; Hunter & Schmidt, 1982; Bobko, et al., 1982; Burke & Frederick, 1984; Schmidt, Mack & Hunter, 1984; Wroten, 1984; Bolda, 1985; Eaton, Wing, & Lau, 1985; Eaton, Wing & Mitchell, 1985; Mitchell, Eaton & Wing, 1985; Weekley, et al., 1985; Burke, 1985; Burke & Frederick, 1985; Mathieu & Leonard, 1987; Rich & Boudreau, 1987);

(3) *Individualized Estimation*, which refers to methods in which some measurable characteristic of each individual in the sample (e.g., pay, sales activity, performance ratings) is translated into dollars using some scaling factor such as average salary or average sales, with the standard deviation of these values representing SD_y (e.g., Janz & Dunnette, 1974; Cascio, 1980; Arnold, et al., 1982; Dunnette, et al., 1982; Bobko, et al., 1983; Ledvinka, et al., 1983; Burke & Frederick, 1984; Reilly & Smither, 1985; Cascio & Ramos, 1986; Eulberg, O'Connor & Peters, 1985; Greer & Cascio, 1987).

(4) *Proportional Rules*, which involve multiplying the value of some available productivity-related variable (e.g., average wage, average sales, average productivity value) by a proportion to arrive at an SD_y estimate (e.g., Hunter & Schmidt, 1982; Schmidt & Hunter, 1983; Eaton, Wing & Lau, 1985; Weekley, et al., 1985; Cascio & Ramos, 1986; Eulberg, et al., 1985; Mathieu & Leonard, 1987; Schmidt, et al., 1986).

Cost accounting. As noted above, the initial concept of a payoff function measured in dollars (Brogden & Taylor, 1950) proposed using cost accounting to attach a dollar value to production units based on their contribution to organizational profit. Then, the number of units produced by each individual in a sample (over a constant period of time) is recorded, and each unit produced is multiplied by its profit contribution, producing a dollar-valued productivity level for each individual. The standard deviation of these values is used as SD_y.

One early study attempted this technique (Roche, 1961, in Cronbach & Gleser, 1965). In summarizing the method, Roche notes that "many estimates and arbitrary allocations entered into the cost accounting" (Cronbach & Gleser, 1965, p. 263), and Cronbach's comments note that it is possible the accountants did not fully understand the utility estimation problem (p. 266-267). Cascio and Ramos (1986, p. 20) also discuss the difficulties they encountered in applying a cost-accounting approach to SD_y estimation for telephone company managers. Greer & Cascio (1987) applied cost accounting to estimate productivity of route salesmen in a midwestern U.S. soft drink bottling company. Their method involved estimating the "contribution margin" (revenue less variable costs) associated with selling cases of different sizes and types, multiplying that by the number of cases sold by each salesman, and then multiplying that by the percentage of sales attributable to route salesman effort on each route. This produced an estimate of the contribution margin for each route salesman, and the standard deviation of these values represented SD_y. The difficulty and arbitrariness of the cost-accounting methodology has frequently been cited as arguing in favor of simpler methods (e.g., Cascio, 1980; Cascio & Ramos, 1986; Hunter & Schmidt, 1982; Schmidt, et al., 1979), because although cost accounting methods are complex, costly and time consuming, they are still prone to arbitrary estimation and subjectivity, especially in jobs for which there is no identifiable production unit, such as managerial jobs.

Global Estimation. This SD_y measurement method, first proposed by Schmidt, et al. (1979), involves having experts estimate the dollar value of several points on a hypothetical distribution of performance (usually the 15th percentile, the 50th percentile, and the 85th percentile). If the average difference between the 15th and 50th percentile is not significantly different from the difference between the 85th and 50th percentiles, the presumption of normally-distributed payoff levels is accepted, and the

average of the two differences is used as the SD_y estimate. This procedure has the advantage of being relatively simple and straightforward to administer (Schmidt, et al., 1979; Cascio, 1980). Schmidt, et al. (1979) proposed that obtaining estimates from a sample of experts would cancel any individual biases. Studies using such global SD_y estimates have generally produced large SD_y values and resulting large utility values, so the global estimation procedure has been the subject of substantial study by researchers interested in testing its reliability and construct validity, producing several controversies.

First, subjects frequently find the task of estimating the dollar value of performance distribution percentiles somewhat difficult. Some respondents gave inconsistent percentile estimates, found the task difficult, refused to do the task, or provided percentile estimates extremely different from the others (e.g., Bobko, et al., 1983; Mitchell, Eaton & Wing, 1985; Reilly & Smither, 1985; Mathieu & Leonard, 1987; Rich & Boudreau, 1987). In these cases, and even where authors do not report subject difficulty, the inter-rater variability in SD_y estimates is usually as large or larger than the mean SD_y estimate. The SD_y Values column of Table indicates the within-sample standard deviation of SD_y estimates (abbreviated *SD*), where reported. Such high inter-rater variability is disturbing, of course, because it suggests that the measures may be capturing bias or error. Desimone, et al. (1986) explicitly examined the inter-rater and temporal stability of their SD_y estimates, and found both of them to be low. Weekley and Gier (1986) also noted inconsistencies in Global estimates across a three-month period. This has led some researchers to suggest new measurement methods (discussed below), others have suggested or investigated variations on the global estimation method designed to improve consensus. The most frequently used tactic is to provide an anchor for the 50th percentile (e.g., Bobko, et al, 1983; Burke & Frederick, 1984; Wroten, 1984; Eaton, Wing & Mitchell, 1985; Burke, 1985; Burke & Frederick, 1986) which is supported by evidence of a high correlation between 50th percentiles and SD_y (e.g., Bobko, et al., 1983; Schmidt, Mack & Hunter, 1984; Wroten, 1984; Edwards, et al., 1988). Research comparing the anchored method to the unanchored method generally suggests that providing anchors reduces inter-rater variability (e.g., Burke & Frederick, 1984; Wroten, 1984), but that the value for the anchor is positively related to the SD_y values that result. Another frequently-used tactic is to have groups of raters provide consensus judgments of different percentiles (e.g., Burke & Frederick, 1984; Wroten, 1984). Some researchers (e.g., Burke & Frederick, 1984; Mathieu & Leonard, 1987) simply drop inconsistent or outlier values on the assumption that they represent error, though there is no theory or empirical data to suggest how inconsistent or unusual an estimate must be to qualify for deletion as an outlier.

A second controversy involves the underlying assumption of normality inherent in the global estimation approach. Averaging the differences between the 15th and 50th percentiles with the differences between the 85th and 50th percentiles presumes a normally-distributed dollar-valued performance distribution. This assumption is often justified by failing to reject the hypothesis that the means of the two differences are significantly different, but this amounts to accepting a null hypothesis. In view of the large inter-rater variability associated with these measures, it seems possible that failure to reject this hypothesis may be due to measure unreliability rather than to an underlying normal distribution. Some studies have suggested non-normal performance distributions or significantly different percentile estimates (e.g., Bobko, et al., 1983; Burke & Frederick, 1984; Schmidt, Mack & Hunter, 1984; Burke, 1985; Rich & Boudreau, 1986) However, other studies found no significant differences, and there is evidence that actual performance distributions follow a normal distribution (Hunter & Schmidt, 1982).

Some researchers examined this issue by including an additional percentile estimate (the 97th percentile), which would be expected to be equally different from the 85th percentile as the 85th and 15th percentiles are different from the 50th. Bobko, et al. (1983) and Burke & Frederick (1984) found the difference between the 97th and 85th percentiles significantly smaller than the other two, suggesting either a non-normal underlying distribution or that estimating the 97th percentile taps a different estimation process than estimating the other three percentiles.

A third controversy arises because the initial research on the global estimation method provided no information to indicate what processes are used in arriving at the SD_y estimates (e.g., what anchors respondents use, what performance attributes they consider, and whether similar anchors and attributes are considered by different experts). This has prompted a few researchers to investigate the judgment processes underlying SD_y. Bobko, et al., 1983 noted that sales managers reported using pay as an anchor for their estimates of "overall worth." Burke & Frederick (1984) gathered anecdotal data following their main study, and found that supervisors of sales managers reported using five dimensions: (1) management of recruiting, training and motivating personnel; (2) amount of dollar sales achieved; (3) management of sales coverage; (4) administration of performance appraisal; and (5) forecasting and analyzing sales trends. Burke (1985) found that supervisors of clerical workers followed job evaluation dimensions in their judgments, with salary-related factors most frequently used.

How accurate is the global estimation technique? Only limited evidence exists, usually based on arguably deficient objective performance measures (sales performance). Bobko , et al. (1983) found that the actual distribution of sales revenue (number of policies sold times average policy value) for sales counselors was normally distributed, and that the SD_y estimate based on the average of the 85th minus 50th and the 50th minus 15th percentiles was not significantly different from SD_y based on the actual sales distribution, although the percentile estimates were quite different DeSimone, et al., 1986 found the opposite results. However, when respondents in the Bobko, et al. (1983) study were asked to consider the "overall worth of products and services" and "what you would pay an outside organization to provide them," the values were only about one-tenth the actual sales standard deviation, and apparently anchored on pay levels rather then sales. Burke and Frederick (1984) also found SD_y estimates of overall worth were lower (about one-percent of the actual sales standard deviations), and anchored on various activities including sales. Reilly and Smither (1986) found that graduate students participating in a business simulation, who had been provided with data to estimate actual standard deviations, produced global SD_y estimates slightly higher than the simulation information for repeat sales and new sales, and much higher than the simulation for net revenue. The SD_y estimate of overall worth was 49% of actual repeat sales, 3.45 times actual new sales, and 1.92 times actual net revenue. DeSimone, et al. (1986) found that the global SD_y estimate for medical claims approvers was 19% of the compensation-weighted standard deviation of actual claims approved. Greer & Cascio (1987) found no significant differences between the global SD_y measure and their cost accounting estimate. Thus, the research comparing global SD_y estimates to objective performance is sparse, and the results are mixed.

Individualized estimation. This is similar to the cost-accounting method in that it attempts to attach a dollar value to the output of each individual in a sample, the standard deviation of those dollar values becoming the SD_y estimate. However, more recent versions of this approach have foregone the complex and costly cost-accounting approach in favor of approaches derived from industrial psychology and HR

management practices. Cascio (1982; 1987) and Cascio & Ramos (1986) have developed the CREPID (Cascio-Ramos Estimate of Performance In Dollars) method. This method breaks a job into important "principle activities." Then, each activity is rated on two dimensions--time/frequency and importance (originally, difficulty and consequence of error were also included), and the ratings multiplied to give an overall weight to the activity. The proportion of total weights becomes the final importance weight assigned to each activity. To assign a dollar value to each activity, average salary for the job is divided among the activities according to the proportional importance weights. After this "job analysis" phase, supervisors are asked to rate a sample of employees in terms of their performance on each principle activity, using a 0 to 200 scale, "with a value of 100 points indicating average performance ('This employee is better than 50% of those I've seen do this activity'). A value of 200 indicated that the employee was better than 99% of those the supervisor had seen do the activity, and a value of 50 indicated that the employee was better than 25% of those he or she had seen do the activity. A value of 0 indicated that the employee was the worst the supervisor had seen do the activity." (Cascio & Ramos, 1986, p. 22). Then, to translate these ratings into dollars, the ratings are divided by 100 (to produce a 0 to 2.0 scale) and these are multiplied by the dollar value assigned to that activity. Finally after each employee has been assigned a dollar value for each activity, these values are summed over activities to provide the total dollar value of yearly performance for that employee. Thus, a person performing better than 50% of the incumbents the supervisor has experienced on all dimensions will receive a dollar value equal to the average yearly salary for that job. A person performing better than 99% of all incumbents will receive a dollar value equal to twice the average salary, and the worst performer each supervisor has experienced will receive a dollar value of zero. Edwards, et al. (1988) modified the basic CREPID procedure applied to District Sales managers by substituting archival data for either performance ratings, job analysis ratings, or both. They found that SD_y levels were similar for the original procedure, and when substituting either performance or job analysis archival information, but much smaller when using archival data for both performance and job analysis (see Table 4).

Janz & Dunnette (1977) also proposed identifying critical job activities. However, rather than allocating salary to each activity based on its time/frequency and importance, the Janz and Dunnette procedure requires job experts to estimate the "relative dollar costs associated with different levels of effectiveness on each of the various job performance dimensions" (p. 120). This requires tracing the consequences of the various levels of effectiveness to determine their impact on activities to which costs and/or value can be attached. For example, different levels of equipment maintenance effectiveness might be traced to breakdowns, which in turn can be traced to repairs, which in turn can be traced to dollar losses due to repair costs and/or lost productivity during repair. Different levels of effectiveness would produce different levels of breakdowns, repair costs and lost productivity. This method was applied to power plant operators by Dunnette, et al. (1982), producing results that supported the high SD_y values derived using the Schmidt, et al. (1979) global estimation method for the same jobs (see Table 4).

Another individualized estimation approach involves having experts directly assign dollar values to individual employees. Bobko, et al. (1983) used this method to derive an SD_y estimate based on sales (sales volume times average policy value) levels, with each person's yearly sales representing the individual value estimate. Burke and Frederick (1984) also used individual sales levels. Wroten (1984) adopted a similar approach, but did not have sales data available. He simply asked his supervisors to

provide a direct estimate of the yearly dollar value of each employee's performance. Ledvinka, et al. (1983) and Desimone, et al. (1986) used total payroll plus benefits divided by the number of insurance claims as the value per claim, and then multiplied this value by the actual standard deviation of claims processed. Greer & Cascio (1987), as noted, multiplied the quantity of cases sold by an estimate of the contribution margin per case. Day & Edwards (1987) proposed a "return on investment" approach that calculated "average annualized investments" for a job as total compensation plus benefits plus 40% overhead. Supervisors estimated the percentage return on this investment represented by each of the seven points on their existing performance appraisal form, with the product of this percentage and the average annualized investment representing the value of that performance rating. Each person's value was estimated according to "ROI" value of their performance rating.

Individualized estimation has the advantage of assigning a specific value to each employee that can be explicitly examined and analyzed for its appropriateness. Such analysis might be useful in determining which individual attributes contribute to differences in payoff values. It may provide a more understandable or credible estimate to be communicated to those familiar with the job. The very limited evidence on this issue is mixed. Greer & Cascio (1988, p. 594) stated that four top managers and an accountant preferred CREPID. Day & Edwards (1987) found no significant differences in managerial confidence ratings for different estimation methods. Edwards, et al. (1988) found supervisors perceived their CREPID job analysis ratings as more accurate than their global utility value estimates, but found the CREPID less "doable"/feasible than Procedure B. These tests do not directly examine the effects of SD_y estimation methods on confidence or accuracy of decisions.

Each method makes certain basic assumptions regarding the nature of payoff. CREPID is based on the assumption that the average wage equals average productivity, a position frequently questioned in economic theory (Becker, 1964; Bishop, 1987; Frank, 1984; Rynes & Milkovich, 1986) and clearly violated in organizations with tenure-based pay systems, pay systems based on rank, hourly-based pay systems, and where training may have different value to different organizations. Sales-based measures are based on the assumption that sales captures sufficient performance differences to be useful (an assumption that may omit important job tasks, such as training, that reduce an individual's sales but increase the group's sales); and the Janz-Dunnette measure assumes that job behaviors' effects on costs and revenues can be accurately traced by managers. Such estimation methods are usually more complex, costly and time consuming than the direct estimation methods, which may provide perfectly adequate SD_y values for many decisions (as discussed below).

Proportional rules. The final SD_y measurement method emerged from observations concerning the relationship between SD_y estimates and average salary levels, and from the desire to provide a straightforward SD_y measurement method. The method involves multiplying average salary in a job by some proportion (e.g., between 40% and 70%) to derive the SD_y estimate for the incumbent employee group.

Hunter and Schmidt (1982, pp. 257-258) reviewed empirical studies for which an SD_y estimate was reported or could be derived. They compared the SD_y estimates to reported average salary levels (or made assumptions about average salary levels), and discovered that on average SD_y was about 16% of average salary in previous studies. They observed that these values "refer to only partial measures of value to the organization" (p. 257) because they generally relied on partial job performance measures

(e.g., tenure or reduced training costs). The authors also reviewed two of their own studies employing the global estimation procedure, where SD_y was 60% of annual salary in one study of budget analysts and 55% of annual salary in another study of computer programmers. They estimated that "the true average for SD_y falls somewhere in the range of 40 to 70% [of average salary]" (p. 258). In a follow-up investigation, Schmidt and Hunter (1983) proposed the following logic: In the United States economy (based on National Income Accounting methods), wages and salaries make up approximately 57% of the total value of goods and services produced. Therefore, if we knew the ratio of SD_y to mean *salary*, we could multiply that ratio by .57 to obtain a predicted ratio of SD_y to mean *output* value. Thus, if the ratio of SD_y to salary ranges between 42% and 60%, the ratio of SD_y to output value should fall between 23% and 34%. To test this logic, the authors reviewed studies reporting empirical data on productivity levels measured in units of output. Their review indicated that for studies examining non-piece-rate situations the average ratio was .185 (standard deviation of .052), for studies examining piece-rate situations, the average ratio was .150 (standard deviation of .044), and for studies with uncertain compensation systems, the average ratio was .215 (standard deviation of .067). Though all three average ratios fell below the lower bound predicted (the values for both the non-incentive and incentive conditions were statistically significantly lower), the authors made five observations: (1) that their method was "intended to apply to jobs without incentive based compensation systems"; (2) that even the lowest mean value is "still 77% as large as the predicted lower bound value" (p. 409); (3) that the studies reviewed "reflect primarily quantity of output; quality of output is probably reflected only crudely in these figures"; (4) that quality and quantity have been found positively correlated in some studies (p. 411), and (5) that the reviewed studies were conducted on "blue collar skilled and semiskilled jobs and lower level white collar jobs", while their studies were conducted on higher level jobs where errors may be more expensive (p. 412). These observations led them to conclude that "researchers examining the utility of personnel programs such as selection and training can estimate the standard deviation of employee output at 20% of mean output without fear of overstatement", and that "the findings of this study provide support for the practice that we have recommended of estimating SD_y as 40% of mean salary" (p. 412). Schmidt, et al. (1986, p. 5) state "the standard deviation of employee output can safely (if conservatively) be estimated as 20% of mean output, or alternatively, 40% of mean salary."

Hunter, et al. (1988) extended this research by analyzing the ratio of output standard deviation to mean output (SD_p) in a larger sample of jobs, including high-complexity and sales jobs as well as those analyzed earlier. They employed new corrections for unreliability that reduced observed SD_p, corrected for restricted range assuming selection on general mental ability, and used variability in salary levels as a proxy for output variability in professional jobs (see the earlier section on the Focus Population). Their findings suggested that as one moves from routine to medium complexity to professional work, the SD_p values progress from 20% to 30% to 50%.

The proportional rules proposed by Schmidt and Hunter are intriguing because they suggest that SD_p estimation may be quite feasible in many applications where job complexity can be estimated, removing a major stumbling block to widespread utility measurement. However, knowing SD_p allows one to estimate only the percentage increase in productivity likely from HRM programs. Determining whether such increases offset dollar costs, or whether to invest program resources in different jobs requires assumptions or estimates of the dollar value of this percentage. The assumption that average salary is equal to about

half the average value of products "as sold" may be violated in tenure-based pay systems, negotiated pay systems, or due to labor market conditions such as unemployment, and internal labor markets (e.g., Becker, 1975). Indeed, National Income accounting used to generate the national GNP and labor cost figures used by Schmidt and Hunter (1983) assigns the same value to both output and wages for jobs where output is not readily measurable (e.g., Government services), producing a ratio of output to wage of 1.0, not .57. Thus, the .57 figure represents an average around which specific jobs may vary.

Existing research provides limited support for the proportional rules applied to output, and less for proportional rules applied to salary. Table 4 shows that of the 44 SD_y values from studies reporting mean productivity values, only two SD_y values fell below 20% of mean productivity, with 13 falling in the predicted 20%-35% range, and 29 falling above 35%. However, of 66 values in studies reporting salaries, 24 fell below 40%, 18 within the predicted 40%-70% range, and 22 above 70%. The values falling above the ranges may reflect high-complexity jobs. In fact, Eaton, Wing & Lau (1985) to conclude that 125% of base pay would be a conservative estimate of SD_y for military personnel. Using 40% of salary may overestimate the SD_y value that would result from other methods, but using 20% of mean output seems to be conservative compared to other measurement methods. Still, overly conservative SD_y estimates may produce severely understated utility estimates, and possible rejection of potentially useful HRM programs. Clearly, the impact depends on the decision situation.

Evidence Directly Comparing Measurement Techniques. Wroten (1984) compared the Schmidt, et al. method, individual subjective payoff estimates, and group consensus percentile estimates for six jobs, with either no anchor, high, low or "accurate" anchors. The means for six unanchored methods, and each of the three anchors are shown in Table 4 for each job. He found that unanchored SD_y estimates had higher variance, that the mean unanchored SD_y estimate was not significantly different from the actual anchored condition, but that it did differ significantly from both the high and low anchored conditions. He also found that individualized estimation usually produced less SD_y variation than the global method.

Eaton, Wing & Lau (1985) compared the Schmidt et al. ("GLOBAL") technique to a variant of the proportional technique called "superior equivalents" (EQV) in which experts estimated the number of 85th and 15th percentile performers it takes to equal the work of 17 average performers (the value of an average performer anchored by either average compensation or the subjective estimate of the 50th percentile). They also used a new "system effectiveness technique" (EFF) in which the standard deviation of payoff is expressed as a proportion of mean payoff (in units of the cost of a tank). The underlying payoff scale of these latter techniques is cost savings (either in terms of payroll or tank costs). Results indicated that as a percent of the GLOBAL value, the EQV salary anchor technique was 66.7%, the EQV global anchor technique was 72%, and the EFF technique was 150%.

Eaton, Wing and Mitchell (1985) compared the GLOBAL technique (using only the 85th and 50th percentiles), the EQV technique and the 40%-70% of salary rule, producing SD_y estimates for 5 military occupations (MOS). Over all 5 MOS, the average SD_y of the GLOBAL technique was $9,387, and for the EQV technique was $14,990. As shown in Table 4, the EQV values were higher for every MOS. The GLOBAL estimates always fell within the 40%-70% range of salary (though they always fell above 35% of the mean y estimate), while the EQV estimates were always higher than 70% of salary. The

EQV technique produced a larger range but a lower between-subject dispersion in SD_y values compared to GLOBAL. The EQV produced no significant differences by MOS, rank, or their interaction. The GLOBAL technique also produced no significant differences by rank or by the interaction of rank and MOS, but it did produce significant MOS differences (with Armor Crewmen having a lower SD_y than Vehicle Mechanics, Medical Specialists, and Radio Operators).

Mitchell, Eaton & Wing (1985) explored whether job incumbents could provide usable SD_y estimates, and studied the jobs of Motor Transport Operator and Cannon Crewman in the U.S. Army. They used the GLOBAL technique, the EQV technique and then the GLOBAL technique after feedback of dollar values for soldiers in other specialties. For both jobs, EQV produced highest SD_y values, GLOBAL next and feedback lowest, as Table 4 shows. The authors also reported that they had respondents delineate job tasks before making estimates and this "seemed to reduce extreme values." Eulberg, O'Connor and Peters (1985) explicitly compared the SD_y estimates provided by supervisors and job incumbents of the medical technician job in the U.S. Air Force. They used the CREPID method, applying the same performance ratings on each job dimension and the same average salary value for both group's estimates. Each group provided its own set of importance ratings for the job dimensions. As Table 4 shows, the SD_y values were quite similar (approximately $3,300 per year) for both methods and for the 40% of salary rule. The authors find this convergence "remarkable", but it reflects only similar rankings of job task importance between both groups (because the same pay levels and performance ratings were used), and the mathematical properties of CREPID suggest it will produce values approximately 40% of pay (Raju, Burke & Normand, 1987; Reilly & Smither, 1986).

Reilly and Smither (1986) provided graduate students taking part in a management simulation with sales data on 10 employees, based on 3 job components (selling established products, selling new products and cost control). They used CREPID methods to obtain importance ratings on the 3 job dimensions and then compared these to the actual simulated data provided. They also obtained SD_y estimates for each job component using Schmidt, et al. (1979) techniques. Both methods caused some confusion among subjects, there were no order effects. The Schmidt, et al. SD_y values for established sales, new product sales, and cost control were significantly correlated (r >.68), but none of these were correlated with the SD_y for "overall worth." The Schmidt, et al. (1979) estimates were slightly higher than actual for repeat sales, 13% higher than actual for new product sales, and 51% higher for revenue less costs. The CREPID SD_y estimate was below all the Schmidt, et al. (1979) estimates, slightly higher than the actual SD_y of new sales, and far lower than the actual SD_y for repeat sales and net revenue. These inconsistencies are interesting because subjects had the information necessary to make exact calculations of the dollar-valued performance for each employee, but apparently failed to use it in their estimates, even under such "ideal" conditions.

Weekley, et al. (1985) compared CREPID to the Schmidt, et al. technique to the 40% of salary rule for convenience store managers. They discovered very high variability in using the Schmidt, et al. method, and this method produced a value almost twice as high as CREPID. The CREPID value was 36% of average salary and the Schmidt, et al. value was 66% of average salary. Cascio & Ramos (1986) also applied the CREPID technique (to telephone company managers) and found that it produced an SD_y value roughly 35% of salary.

Desimone, et al. (1986) found that Global SD_y estimates were much lower than compensation-

weighted deviations in the number of processed claims. Similarly, Greer & Cascio (1987) found "Cost-accounting" SD_y estimates to be slightly higher than Global estimates for soft drink route salesmen. Greer & Cascio (1987) also found the CREPID method produced the lowest SD_y estimate. Day & Edwards (1987) found that SD_y values were highest for the Global and modified Global method, followed by the % ROI, and lowest for the 40%-salary and CREPID methods for Account Executives and Mechanical Foremen. Finally, Edwards (1988) found that the Global method with feedback (Burke & Frederick's, 1984 Procedure B) produced the highest SD_y values, followed by various forms of the CREPID method.

CREPID estimates frequently fall near 40% of salary, and below more Global estimates, prompting the argument that because the CREPID scale is based on salary, it "considers only the contribution of labor not the combined contribution of labor, equipment, capital, overhead and profit, as does a standard based on the value of output as sold" (Greer & Cascio, 1987, p. 593). Edwards, et al. (1988, p. 533) also argue that average salary cost should be increased by the amount of benefits and "overhead" before scaling, to better reflect the "value of the total cost of services." The ROI method (Day & Edwards, 1987) proposes a similar scaling approach. It is undoubtedly true that larger scaling factors would increase CREPID estimates, but it is not clear that such adjustments are justified. As noted above, salary (or salary plus benefits) will not necessarily reflect the average value of employees, and may overstate it. Selecting higher-quality employees often has little effect on expenditures for equipment, capital, and overhead, so including these factors as potential cost reductions seems inappropriate. Moreover, while all of these factors moderate the contribution of higher-quality labor to organizational goals, the SD_y concept always reflects such contributions because it is estimated across employees within a particular mix of capital, equipment and overhead. Salary-scaled estimates may or may not reflect this quality, but the concept is no different whether scaled using salary or some other method. As noted in the discussion of payoff scales, the key question is how the high-quality labor will be used to enhance organizational goals, and this is likely to be situationally specific. SD_y measures and *post hoc* adjustments should make their assumptions explicit. Simple proportional rules or compensation-based scaling factors may not generalize to every situation. Yet, where do such questions fit the decision-theory perspective on utility analysis?

Summary and Conclusion: The Need to Look Beyond SD_y

Differences between SD_y estimates using different methods are often less than 50% (and may be less than $5,000 in many cases). Still, these differences may be multiplied by factors of hundreds or thousands, depending on the number of employees selected, the validity of the device and the selection ratio (as shown in Table 3), in deriving the final total utility value. Even a small SD_y differences multiplied by such large values imply vast total utility differences. The tempting conclusion is that we need substantially more research on SD_y measurement to whittle down such differences and provide more precise total utility estimates.

This conclusion is not encouraging. The unfortunate fact is that I/O psychology and Human Resource Management has produced no well-accepted measure of job performance differences (on any scale, let alone dollars). The task of estimating dollar-valued performance variability has proven confusing and difficult for some subjects and virtually always produces substantial disagreement among raters. When SD_y estimates can be verified against an objective criterion (e.g., sales, units produced), the criterion is arguably deficient leading to the conclusion that any observed differences in SD_y estimation

methods cannot serve as justification for using one measure over another (Day & Edwards, 1987; Burke & Frederick, 1984; Weekley, et al., 1985). Thus, measured against the accuracy of SD_y estimates, the contribution of this research must await development of an acceptable dollar-valued performance measure. Greer & Cascio (1987, p. 594) state "researchers within the accounting profession must develop an objective, verifiable, and reliable method for estimating the standard deviation of job performance in dollars", but accounting systems are neither designed nor intended to reflect this variable. Of course, if a measure of y existed, we could derive SD_y directly, making estimation unnecessary. Indeed, even the SD_y concept would have little value because one could use the slope of the regression line in Equation 2 to predict selection utility. Thus, while SD_y measurement research produces information on the variability across raters, methods or jobs, it is unlikely to provide information on measurement accuracy, nor is it likely to allow us to substantially reduce the uncertainty associated with total utility values.

SD_y estimation research can advance measurement theory. Here, the value of the research rests not on its ability to better describe, predict, explain or enhance decisions, but rather to illuminate new aspects of measurement. This may be a quite useful and legitimate application of the SD_y concept. However, it is very different from utility analysis, and this difference should be made clear by those researchers pursuing measurement theory.

If SD_y research is unlikely to produce the most accurate measure, and unlikely to alleviate uncertainty in utility estimates, then what is the role of SD_y measurement research in advancing UA knowledge? The B-C-G utility model emerged from traditional psychological measurement theory, which focused only on standardized error terms but provided no context within which to evaluate them. UA models were formulated to better account for the decision context facing selection program managers. It seems ironic that after over 30 years, the major research efforts remain focused on measurement, taking little notice of the decision context in which such measures will be used. We have evolved from focusing only on the correlation coefficient, to focusing only on the SD_y value. We must return to describing, predicting, explaining and improving *decisions*, taking into account the context within which those decisions must be made. This suggests several research issues which have been all but ignored in the rush to develop new SD_y measures.

First, the effects of SD_y measures on the perceived quality of the utility analysis should be examined. Though virtually every new measure is justified by proponents because it may produce more credible, understandable, or easily communicated utility values, not one study has directly addressed these issues. If decision makers find the utility values resulting from a relatively simple proportional rule just as credible as complex job-analysis-based methods (e.g., CREPID, Janz-Dunnette) or Global estimation (Schmidt, et al., 1979), they may have little motivation to pursue the latter to increase decision credibility. Of course, even decision-maker preferences are not the real issue. Research (e.g., Kahneman & Tversky, 1972, 1973) shows that decision makers frequently prefer and use heuristics that are detrimental to decision quality. UA research must focus on the quality of decisions as well as decision maker preferences.

Second, and related to the first issue, we have little information on the relative effort and cost required to implement the different SD_y measurement procedures. On their face, the proportional rules seem least complex, followed by the global estimation methods, followed by individual estimation methods, followed by the job-analysis based methods. In a sense, the burden of proof rests with those

who would advocate more complex and costly measures to demonstrate that the improvement in decision quality or our ability to understand the decision process justifies the additional resources necessary to gather the information. The costs of the different SD_y estimation methods has not been computed, though Cascio & Ramos (1986) noted that CREPID ratings took 15 minutes per employee and Edwards, et al. (1988, p. 532) noted that their managers felt the CREPID procedure took "too long".

Third, and most important, comparative SD_y studies seldom estimate overall utility values for actual decisions, producing results that are completely devoid of any decision context. It is often impossible to tell whether the measurement differences detected would have made any difference to actual decisions. Yet, as Table 3 demonstrates, virtually every study that has accounted for the decision context (by computing a utility value) has produced extremely high utility values regardless of the SD_y level. Weekley, et al. (1985) proposed that while break-even SD_y values are low when comparing implementing an HRM program to doing nothing (with zero cost and zero benefits), comparing HRM programs to other organizational investments might produce decision situations where differences in SD_y estimates indeed affect the ultimate decision. Research incorporating such contextual variables could prove quite fruitful. Still, in the absence of any criterion against which to verify SD_y values, one would still be left with little basis for choosing one over another. Further SD_y measurement research seems unlikely to explain how apparently high HRM program payoffs can exist while the HRM function achieves low status and importance in organizational decision making. Answering that question, and developing decision models to alleviate the situation, requires that UA research explicitly recognize organizational decision contexts. The next sections discuss how utility models can better reflect such contextual factors, and links UA research to other fruitful research streams.

The Role of Uncertainty and Risk in Utility Analysis

How is it that UA research can simultaneously produce such clear evidence of HRM program payoff and such a raging debate on the proper measurement method for one utility parameter (SD_y)? Although the expected utility values are quite high, if substantial uncertainty is associated with these utility estimates, and if that uncertainty results from uncertain SD_y values, then reducing SD_y measure uncertainty will improve decisions. However, uncertainty affects all UA parameters, not just SD_y. Just as all models are deficient, all predictions contain uncertainty. UA research cannot ignore this fact, but must instead embrace its implications for advancing understanding of decisions and decision processes.

Frameworks incorporating uncertainty (Alexander & Barrick, 1987; Boudreau, 1984a, 1987; 1988, in press; Cronshaw, Alexander Wiesner & Barrick, 1987; Milkovich & Boudreau, 1988; Rich, 1986; Rich & Boudreau, 1987) change the focus of utility analysis from estimating the expected utility value to estimating both the expected value and the distribution of values. Measurement issues become relevant as they affect uncertainty in the decision situation. This framework emphasizes the role of utility value variability in changing decisions, rather than simply measuring the sources of that variability (e.g., SD_y measurement error) in the absence of a decision context. It is surprising that the issue of uncertainty and risk in utility analysis received little attention for so long, because decision theory has traditionally been concerned with decision making under uncertainty, and has recognized that the riskiness of alternatives plays a role in decision making. this emphasis has been especially evident in the literature on financial

investment decision making (e.g., Bierman & Smidt, 1975; Hertz, 1980; Hillier, 1963; Hull, 1980; Wagle, 1967). If two alternative resource investments offer the same expected value, but offer substantially different risks of large losses (below he expected value) or large gains (above the expected value), rational decision makers should take such risks into account.

Four Alternative Approaches for Estimating Uncertainty

Rich and Boudreau (1987b) provided an initial conceptual framework for uncertainty in UA and empirically compared four alternative methods addressing uncertainty: (1) sensitivity analysis; (2) break-even analysis, (3) algebraic derivation of utility value distributions; and (4) Monte Carlo simulation analysis.

Sensitivity analysis. Though existing utility models contain no parameters reflecting utility value variability, the notion that utility values represent estimates made under uncertainty has not been completely overlooked. Several previous utility analysis applications and demonstrations (e.g., Boudreau, 1983a, 1983b; Boudreau & Berger, 1985a; Cascio & Silbey, 1979; Cronshaw, et al., 1987; Florin & Boudreau, 1986; Schmidt, et al., 1979; Schmidt, et al., 1984) have addressed possible variability in utility parameters through sensitivity analysis. In such an analysis, each of the utility parameters is varied from its low value to its high value while holding other parameter values constant. The utility estimates resulting from each combination of parameter values are examined to determine which parameters' variability has the greatest effect on the total utility estimate. These sensitivity analyses virtually always indicate that utility parameters reflecting changes in employee quality caused by improved selection (i.e., rx,y, $\overline{Z}_x$, SD_y) and the quantity of employees affected (i.e., N_s) have substantial effects on resulting utility values. A variant of sensitivity analysis involves attempting to be as "conservative" as possible in making utility estimates. This approach has led researchers to produce clearly understated SD_y values (Arnold, et al., 1982), or to estimate the 95% confidence interval surrounding the mean SD_y value and use the value at the bottom of this interval in the utility computations (e.g., Cronshaw, et al., 1987; Hunter & Schmidt, 1982; Schmidt, et al., 1979, 1984). If estimated utility values remain positive despite such conservatism, it is presumed they will be positive in the actual application.

Though valuable in assessing the effects of individual parameter changes, sensitivity analyses provide no information about the effects of simultaneous changes in more than one utility parameter (though Boudreau & Berger, 1985a and Boudreau, 1986 expressed utility as a function of changes in several parameter levels, to present the effects of simultaneous changes in utility parameters more concisely). They also provide no information regarding the utility value distribution nor the probabilities associated with particular parameter value combinations (Hillier, 1963, p. 444). Moreover, when all parameters are estimated at their most conservative levels (a statistically unlikely event), one runs the danger of incorrectly concluding that some programs will not pay off.

Break-even analysis. Boudreau (1984a) proposed that a relatively simple and straightforward uncertainty analysis could be carried out by calculating the lowest value of any individual utility parameter (or parameter combination) that would still yield a positive total utility value. These parameter values were termed "break-even" values because they represent the values at which the HRM program's benefits are equal ("even with") the program's costs. Any parameter values exceeding the break-even

value would produce positive total utility values, and vice versa. Such logic is well-known in micro-economic theory and financial management (i.e., Bierman, Bonini & Hausman, 1981). Boudreau showed how to apply break-even analysis not only when considering one program option (i.e., where the alternative is to do nothing, incur no costs, but receive no benefits), but also when multiple alternatives are involved (with more expensive alternatives offering greater potential payoffs). With multiple alternatives, one computes a series of decision rules specifying the range of parameter values that would justify choosing that alternative over the others (e.g., Boudreau, 1988; Milkovich & Boudreau, 1988). The break-even approach is simple and focuses on the decision context. Boudreau proposed that break-even analysis allows decision makers to maximize the value of existing information, determine the critical values for the unknown parameters that could change the decision, and determine whether further measurement effort is warranted. Because controversy surrounded the accuracy and validity of SD_y estimates, Boudreau (1984a) concentrated his analysis on that utility parameter, demonstrating that the break-even SD_y values for the studies by Cascio and Silbey (1979), and Schmidt, et al. (1979) were substantially lower than the expected SD_y value they derived.

Table 3 updates Boudreau's analysis to incorporate additional and more recent utility analyses. The column labeled "Payoff Function," presents incremental utility as a function of SD_y. These payoff functions were derived for each study and each selection device. The coefficient multiplied by SD_y in each equation was derived by dividing the total program payoff (before subtracting program costs) by the SD_y value. The number subtracted from this product is the reported total cost. All equations express payoff in terms of total utility, but for studies that reported only per-person utility values, the payoff function reflects the utility of selecting one individual. The last column of Table 3 computes the break-even SD_y value based on the payoff equation. This is simply the cost figure divided by the coefficient on SD_y. For studies reporting no incremental cost for more valid selection, the implied break-even SD_y value is also zero because any positive return justifies a costless program, so SD_y becomes irrelevant.

The equations and break-even values not only verify the earlier conclusion that HRM program utility is uniformly high, but also shed some light on the SD_y controversy. Compare the reported SD_y values (in the column labelled SD_y) to the break-even values for each study. Without exception, the break-even SD_y values fall at or below 60% of the estimated SD_y value. In many cases, the value necessary to break even is less than 1% of the estimated value. The break-even SD_y value exceeds 20% of the estimated value in only 6 of the 42 analyses. In three of these six cases (Burke & Frederick, 1986; Rich & Boudreau, 1987; Schmidt, Mack & Hunter, 1984), this reflects an interview with low validity. The break-even value determining whether to replace the interview with a more-valid predictor was much smaller in the latter two studies.

In short, the vast majority of utility analysis applications conclude that the more-valid selection device is worth its extra costs. This conclusion would probably have been apparent without ever actually measuring SD_y (or by measuring it in the simplest manner possible) because the break-even SD_y values are so low that they often fall several standard deviations below the expected value. Rich & Boudreau (1987b) found that the break-even SD_y value fell below the lowest value estimated by any of the subjects.

Boudreau's (1984a; 1987; 1988; in press) findings produced a similar conclusion, leading him to propose that future utility analysis research should use break-even analysis to put parameter measurement controversies into perspective. He speculated that many UA applications do not require costly and

complex SD_y measurement, but could simply present decision makers with the break-even SD_y values and ensure that there is consensus that it would be exceeded. Moreover, he proposed that such an approach may prove much less confusing and difficult for decision makers than attempting to estimate an exact point estimate. In other words, the break-even approach suggests a mechanism for concisely summarizing the potential impact of uncertainty in one or more utility parameters. It shifts emphasis away from estimating a utility value, to making a decision using imperfect information. It pinpoints areas where controversy is important to decision making (i.e., where there is some doubt whether the break-even value is exceeded) versus areas where controversy has little impact (i.e., where disagreements about SD_y do not indicate a serious risk of values below break-even). Thus, break-even analysis provides a simple expedient allowing utility analysis models to assist in decision making even when some utility parameters are unknown or uncertain. Measurement research (on SD_y or other utility parameters) is not always unnecessary, but such research must consider the decision context, and report not only the magnitude of the uncertainty but also its likely effect.

Recent research incorporating Boudreau's break-even analysis approach has reached similar conclusions (e.g., Burke & Frederick, 1986; Cascio, 1987; Cascio & Ramos, 1986; Florin-Thuma & Boudreau, 1987; Karren, NKomo, & Ramirez, 1985; Mathieu & Leonard, 1987). Eaton, Wing and Lau (1985) also concluded that HRM program decisions in the military seldom hinge on differences of 10% or 20% of testing costs, so a rough estimate of SD_y may often be sufficient for decision making.

Although relatively simple, break-even analysis is not without limitations. It is more difficult (but quite possible) to conduct break-even analysis when more than two or three utility parameters may vary. Moreover, the distribution of utility values is not estimated, so two programs could have similar break-even values and similar expected utility values, but one might be preferable because its distribution may be more positively skewed. Neither traditional utility analysis, sensitivity analysis, break-even analysis nor algebraic derivation (discussed next) adequately reflect such situations.

Algebraic derivation of utility value variability. Goodman's (1960) equations for the variance of the product of three or more random variables under conditions of independence were adapted by Alexander and Barrick (1987) to produce a formula for the standard error of utility values associated with a one-cohort selection utility model. They demonstrated this derivation using data from the Schmidt, et al. (1979) study, as well as variance estimates for employee tenure, SD_y, validity, and the number selected. Their standard deviations (estimated for various selection ratios) were about 50% of the expected utility values. By assuming a normal distribution of utility values, determining the utility value at the lower end of a 90% confidence interval, and using break-even analysis, the authors concluded that the selection program had a very high probability of producing benefits exceeding costs.

Algebraic derivation reflects simultaneous variability in several utility parameters, and can be useful in estimating the risk associated with utility values. However, it is more complicated than break-even analysis and has limitations. First, the formula can incorporate dependencies between variables, but doing so produces very complex estimation equations and requires information on covariances that is seldom available. Variance estimates become especially difficult when programs can be expanded or abandoned during the project's life, or when variables are related in a non-linear fashion (as the selection ratio and the average standardized predictor score are related in utility formulas). Alexander and Barrick (1987) surmounted this difficulty by holding the selection ratio and average predictor score constant for each

variance estimate. Second, algebraic derivation provides a variance estimate, but it requires assumptions about the distribution shape (e.g., normality) to make strong probabilistic inferences. Existing literature provides no empirical information supporting or refuting the assumption of normality, but Hull (1980) noted that non-normal distributions are likely when: (a) programs can be abandoned or expanded during their life; (b) non-normal components heavily influence the distribution; and (c) there is only a small number of variables. Each of these conditions may characterize utility analysis, as discussed below.

Monte Carlo simulation of utility value variability. Monte Carlo simulation attempts to address limitations of the three previous methods. Simulation describes each utility model parameter in terms of its expected value and distribution shape. In each simulated trial, a value for each utility parameter is "chosen" from the distribution for that parameter, and the combination of chosen parameter values is used to calculate total utility for that trial. Repeated application of this choosing and calculating procedure (using a computer) produces a sample of trials from which describes the distribution properties of utility values. Thus, unlike the other three methods, simulations can vary many parameters at once, can reflect dependencies among the parameters, can acknowledge possible program expansion or abandonment, and can reflect non-normal distribution assumptions.

Rich and Boudreau (1987b) applied Monte Carlo simulation and the other three uncertainty estimation methods to a decision to use the Programmer Aptitude Test (PAT) to select computer programmers in a mid-size computer manufacturer. They used a utility model enhanced to reflect financial/economic factors and employee flows through the work force (these enhancements will be discussed subsequently). They discovered that all of the utility parameters were subject to some degree of uncertainty or variability over time. They also discovered that SD_y variability heavily influenced the utility value distribution and that the distribution of SD_y values was positively skewed as in other studies (e.g., Bobko, et al., 1983; Burke & Frederick, 1984; Schmidt, Mack & Hunter, 1984; Burke, 1985; Mathieu & Leonard, 1987). The simulation suggested greater risk (variability) in utility values than the algebraic derivation because the simulation better reflected dependencies among utility parameters and parameter relationships over time. However, break-even analysis, algebraic derivation and Monte Carlo simulation all led to the same conclusion--The selection program had a very small probability of negative payoff.

Cronshaw, et al. (1987) also simulated utility values, but their analysis held validity, costs and SD_y constant, and used subjective estimates of optimistic, likely and pessimistic parameter levels, rather than observed distributions. Their analysis also focused only on effects for the first cohort of selectees hired, while Rich & Boudreau (1987b) incorporated effects of subsequent program application and employee turnover (discussed subsequently). Still, Cronshaw, et al. (1987) reached a similar conclusion--The selection program had a very small probability of negative payoff.

Thus, Monte Carlo simulation better reflects factors affecting utility value variability, and indeed suggests that substantial variability existed due to both measurement error and uncertainty regarding future conditions (Rich & Boudreau, 1987). This methodology may prove very useful in describing the behavior of utility value variability in future research. However, existing research also suggests that the simpler break-even analysis procedure may describe the decision situation adequately enough to reveal the correct decision. We should also note that all selection utility models and all of the variability estimation procedures except Monte Carlo analysis presume a linear and constant relationship between utility and the parameters reflecting employee quantity and quality (i.e., Ns, $\bar{Z}_x$, SD_y, and rx,y). Economic theory

suggests this assumption may be questionable in certain situations (as will be discussed subsequently). Therefore, Monte Carlo simulation may have an advantage over the other three methods when such non-linearities are important enough to alter decisions and when they can be quantified sufficiently to be incorporated into a simulation algorithm.

Statistical Hypothesis Testing and Uncertainty in Utility Analysis

The inferential statistics approach. Researchers are familiar with the classic statistical tools of confidence intervals, hypothesis tests, and probability statements. Such tools usually emphasize the probability of Type I error (accepting an alternative hypothesis that is false) by specifying the significance level of the statistical test. A statistical approach uses sample information to estimate the variability of the sampling distribution in a statistic (e.g., t, F, etc.). Then, assuming that the null hypothesis (usually a hypothesis of zero effect) is the mean of the sampling distribution, a decision rule for the statistic is established such that the null hypothesis is rejected in favor of the alternative hypothesis only if the observed effect in the sample is large enough to fall near the tail of the assumed distribution (i.e., in the highest 5% or 1% of the distribution). Of course, this ignores the probability of Type II error (incorrectly failing to reject the null hypothesis when it is false). Although some methods for reducing Type I error (e.g., increasing sample size and measurement reliability) reduce both errors by producing a smaller variance in the sampling distribution, other mechanisms for reducing Type I error (e. g., requiring larger effects before rejecting the null hypothesis) actually increase the probability of Type II error.

Utility models are intimately connected to both statistical inference and to decision making. UA models make use of statistics (e.g., the correlation coefficient) that summarize sample information. However, they also can illuminate the limitations of statistical analysis in a decision making context, and suggest a more complete approach to using statistical evidence in decision making. Several authors have argued for increased emphasis upon substantive significance as opposed to statistical significance (e.g., Campbell, 1982; Rosenthal & Rubin 1985). With its emphasis on decision making, UA research can contribute to formalizing and quantifying this more substantive emphasis. It is beyond the scope of this Chapter to fully debate the philosophical and practical issues surrounding the question of substantive and statistical inference. However, it is important to delineate some important roles for utility analysis in such a debate.

The role of utility analysis in defining substantive significance. Statistical inference emphasizes extreme conservatism in the interest of maximizing confidence in reported findings. Specifically, it sets very stringent standards for new research results to replace previously accepted findings. Consider validation studies, where the correlation coefficient is tested for statistical significance using the inferential model specified above. Assuming the true distribution of correlation coefficients has a mean of zero (and a variance determined by the sample size, the reliability of the measures and other factors), the observed correlation must be large enough that the probability of its occurrence in such a distribution is below 5% before rejecting the null hypothesis of zero correlation. Such an approach amounts to an extremely conservative decision rule, especially because practical sample sizes and measure reliabilities often require quite large sample correlation coefficients to reach statistical significance (Schmidt, Hunter & Urry, 1976). Meta-analysis techniques can help to place the results of many small-sample studies in perspective and

provide a truer picture of the correlation coefficient mean and variability.

The inferential statistical model is usually applied outside of a decision making context. No costs or benefits are attached to the two types of error, and the implied value judgments inherent in statistical testing are accepted implicitly. But suppose the study described above were being conducted in an actual organization, where managers must decide whether to adopt a particular selection device. Costs associated with Type I error--i.e., adopting the selection device when it should have been rejected include test development, test administration and scoring, and possible productivity reductions from using the test instead of (or in addition to) the existing selection system. Adopting a decision rule that rejects the selection device unless the observed correlation is large enough to reach statistical significance "protects" the organization from needlessly incurring these costs. However, Type II error--failing to adopt the selection device when it should have been adopted also brings costs such as the lost productivity enhancements or cost reductions from improved person-job matching. The B-C-G utility model suggests that productivity enhancements and cost reductions are often quite sizable even with very modest correlation coefficients and performance variability. Thus, improved selection systems may often be "worth the risk," because the costs of Type I errors are fairly small, the costs of Type II errors are relatively large, and only a modest validity level is required to produce benefits from the improved selection system. Classical statistics attempts to minimize Type I error even at substantial risk of Type II error, reflecting values that are virtually opposite from these characteristics.

Several authors have made a similar argument, though the link to utility analysis has not been as clear. Rosenthal and Rubin (1985) take issue with the notion that statistical inference is designed to establish facts, proposing that the purpose is to summarize information efficiently. They make three important points: (1) that "when the dependent variable is of some importance and where obtaining additional data is difficult, expensive or unlikely" even non-significant results can contribute to scientific understanding; (2) that by taking the ratio of the probability of Type I error to Type II error, we obtain an index of the "perceived relative seriousness" of the two errors which indicates that in most studies Type I errors were implicitly "from 5 to 95 times more serious than Type II errors" (p. 529); and (3) that the notion of value-free scientific inference is usually inaccurate because investigators use their own values in choosing what statistical tests and contrasts to investigate. Cascio and Zedeck (1983) also suggested computing the ratio of Type I to Type II errors as a measure of the relative importance of decision consequences. Using utility analysis formulas, they demonstrated that less stringent decision rules increase power (the ability to detect non-zero effect sizes), but also increase Type I error. They suggested that researchers adjust alpha levels (i.e., acceptable Type I error levels) downward to increase statistical power.

Fowler (1985) noted Campbell's (1982) lament that statistical significance is often incorrectly taken as substantive significance, and his admonition that researchers argue for the substantive as well as the statistical significance of their findings. Using a definition of substantive significance derived from Cohen's (1977) signal to noise ratio, and Lykken's (1968) observations concerning common variance in psychological variables, Fowler reviewed *Journal of Applied Psychology* articles from 1975 and 1980, finding that 75% of the 1975 effect sizes and 69% of the 1980 effect sizes were "below Cohen's large effect" (p. 217), though they reached statistical significance. Abelson (1985) described the paradox whereby baseball batting skill explains less than 1% of the variance in single-at-bat performance but is

regarded with extreme importance by decision makers in selecting baseball players and assigning them positions (batting opportunities in games). Abelson pointed out that decision makers must consider the season-long performance of a team, not the single at-bat performance. Because any player may have 1,000 at bats in a season, and because scoring rallies are more likely when groups of skilled batters build on each other's skills, even the "modest" explanatory power of batting skill has important implications for team performance (compared to other alternative selection and assignment schemes).

These observations suggest an important role for UA research in explicating the debate on substantive versus statistical significance. Rosenthal and Rubin's first observation supports the earlier conclusion that the potential effect of HRM programs on productivity is important enough that even imperfect information about utility parameters may be quite valuable, because further data gathering may be difficult or expensive. Their second observation, as well as the Cascio and Zedeck observations, suggests that HRM program decision rules should be adjusted so that the ratio of Type I to Type II error probabilities is consistent with the costs and benefits of both types of errors. Rosenthal and Rubin's third observation suggests that both the organizational implications and scientific value judgments should be considered when interpreting statistical tests, and UA models can provide valuable information describing organizational implications. Fowler's observation that a majority of research studies may produce statistically significant findings that have low substantive importance is not unlike our earlier observation that although many of the discrepancies between SD_y measurement methods are quite large in absolute terms, break-even analysis reveals that the discrepancies appear to have little bearing on the quality of HRM program decisions. Fowler's findings also reinforce our conclusion that such research should report the decision context in which SD_y information will be used. Finally, Abelson's (1985) explanation for the variance explanation paradox parallels the break-even analysis of Table 3, suggesting that most organizations (like baseball teams) are more concerned with productivity outcomes reflecting multiple employees and time periods, than with the behaviors of one employee.

Incorporating the Value of Information

The issues of necessary precision, statistical versus substantive significance and uncertainty regarding UA for HRM programs are analogous to similar issues for other organizational investment decisions. Decision makers must (implicitly or explicitly) assess the value of additional information (and the cost of acting with uncertainty) in light of the particular decision context they face, and several models for quantifying these issues are available (cf. Bierman, et al., 1981). Yet, these well-known models are not usually applied to HRM decisions at least partly because of the widespread failure even to attempt to quantify the effects of HRM interventions. UA allows quantification, and thus offers one link to decision models that more explicitly incorporate the value of information. Although it is not possible to fully develop the mathematical and logical arguments inherent in such an information model, we can briefly summarize how UA models can be used for this purpose, and the implications of viewing UA as a component of the larger task of making decisions under uncertainty.

The basic information value model incorporating utility analysis. Information has value when it reduces uncertainty in a way that produces better decisions. Gathering information (such as utility model parameter measurement) is a decision in itself, subject to both desirable and undesirable consequences. In

simplest terms, the value of additional information depends upon: (1) the probability that the information's results can change decisions; (2) the consequences of the changed decisions; and (3) the cost of gathering the additional information. The value of additional information may be considered as the product of (1) and (2), less (3). Additional information has greatest value when the probability that the additional information will change the decision is very great, the consequences of changed decisions are very large, and the information can be gathered at low cost. Additional information has less value under the opposite conditions.

Evaluating information requires: (1) an explicit decision (i.e., the alternatives, their attributes, and the value of the differences in their consequences); (2) a decision rule for using the additional information to alter decisions; (3) assumptions or data regarding the likely results to be revealed by the information; and (4) the cost of the additional information. Two models for evaluating information value are commonly discussed (cf. Bierman, et al., 1981)--"perfect information" and "imperfect information." The two models differ primarily in the way they treat the third factor listed above (i.e., the probable results of the additional information).

Suppose an organization is considering two selection devices. One device is more valid, but also more costly to develop, administer and evaluate. The decision maker realizes that selection utility consequences will be quite different if future conditions produce very large applicant pools (allowing the organization to be very choosy, and achieve a high average selectee test score) as opposed to very small applicant pools (providing less choice and thus less payoff to improved validity). Suppose it has been determined that two selection ratios are possible (i.e., .30 or .70). UA reveals that if the selection ratio is .30, then the more-expensive device offers a utility of $500,000, and the less-expensive device offers a utility of $300,000. If the selection ratio is .70, then the more expensive device offers a utility of $50,000 and the less-expensive device offers a utility of $200,000. Should the decision maker gather additional information (e.g., labor market forecasts, strategic forecasts of labor demand, etc.) to attempt to predict the selection ratio more precisely? Without further information, the decision maker attaches a 20% probability to the .30 selection ratio, and an 80% probability to the .70 selection ratio. Thus, with no additional information the expected values are $140,000 and $220,000 for the more- and less-expensive alternatives, respectively, and the less-valid and less-expensive alternative is preferred. The expected value of this decision is $220,000.

In the "perfect information" model, one assumes that a perfect predictor would foretell the actual selection ratio in advance, and calculates the additional decision value that could be derived from that information. In the example, if the decision maker had perfect information, then there is a 20% chance that the information would foretell a low selection ratio. With this information, the decision maker would switch to the more-expensive alternative and would enjoy the $500,000 utility instead of the $300,000 utility of the less-expensive selection device. However, there is an 80% chance the perfect information will reveal a high selection ratio, in which case the original decision was correct anyway. Thus, the value of perfect information is equal to 20% times the utility difference under the high-selection-ratio condition (i.e., .20 times $200,000), or $40,000. Under these assumptions, this is the upper limit of the value of any information that improves the ability to predict the actual selection ratio. The information value is realized only under the conditions where it changes the decision, and depends on the consequences as well as the probability of that change. Even in the absence of selection ratio

information, it is possible to compute the value for the two probabilities that would change the decision. The more-expensive option is preferred if the probability of the low selection ratio exceeds 43 percent, and the probability of the high selection ratio does not exceed 57 percent.

In the "imperfect information" model, one uses Bayesian probability relationships to determine how imperfect information changes the *a priori* probability estimates, the decisions implied by these changes, and the expected consequences of the decisions under all future conditions and information outcomes. Frequently decision trees can represent the decision situation. However, the main objective is similar--to determine the economic value of information designed to reduce uncertainty. Moreover, the same three factors determine the economic value of additional information.

Variations on these models can be developed that reflect continuous as well as discrete distributions of future conditions, information outcomes and probabilities. Indeed if the distribution of information outcomes is assumed to be normal, it is possible to evaluate the consequences of various statistical decision rules (e.g., setting Type I error at 5%) in light of alternative future conditions (e.g., decision consequences, true values for utility parameters), and determine the economically optimum decision rule and/or the economically optimum sample size for a future study. Moreover, such methods can be applied not just to uncertainty regarding future selection ratios, but to uncertainty regarding any of the utility parameters. Such a framework makes it possible to explicitly consider not only expected utility values, but uncertainty and risk inherent in those values as well as the implications of decision rules derived from inferential statistics or other methods.

Linking the Information Value Model and Emerging UA Research. Recall the three determinants of information value: (1) the probability that the additional information will change decisions; (2) the consequences of the changed decisions; and (3) the cost of gathering the additional information. Despite research on SD_y variability, actual selection device decisions are unlikely to be altered by different SD_y measures because: (1) different SD_y measures have low probabilities of producing SD_y values below break-even, (2) even crude SD_y estimates will often lead to a decision favoring improved selection, and (3) refined SD_y measures may be complex and costly. Boudreau (1984a) developed this point in detail using the "perfect information" model.

When the costs of implementing improved selection are modest compared to the potential benefits, relying on decision rules based on statistical significance may be overly conservative. Failure to adopt improved selection devices because validities do not reach statistical significance may imply a belief that the consequences of erroneous implementation are tens (or hundreds) of times as great as the consequences of erroneous failure to implement. Existing evidence suggests that implementation costs are low and potential productivity benefits are very large, so *HR managers often cannot afford the risk of not trying improved selection devices.* Still, the B-C-G model reflects only a portion of relevant decision factors, and we have no research to suggest what other factors decision makers may consider.

Conclusion

In view of the high variability associated with utility parameter estimates (especially SD_y), it seems plausible that perceived uncertainty and risk associated with utility estimates may explain why HRM programs do not enjoy widespread acceptance and why the utility values may appear larger than many

researchers would have expected (Schmidt, et al., 1979; Hunter & Schmidt, 1982). However, this view may also reflect ignorance of the capacity for improved HR management to affect organizational goals. As the illustration in Table 2 and the break-even analysis of Table 3 vividly illustrate, the *leverage* or quantity of person-years affected by HRM programs can be quite large. Thus, the coefficient on SD_y is often quite large, and even modest levels of performance variability offer substantial opportunities for highly valuable HRM program effects. Schmidt et al. (1986; 1984) have expressed utility values as a percentage of output and wages, and suggested that if decision makers or researchers find utility values "implausible," it may reflect the fact that they do not appreciate the magnitude of their human resource investment.

Uncertainty about SD_y would not have been an important factor in any published utility analysis applications. Further research on the cognitive processes affecting SD_y estimation and further efforts to develop new and more reliable SD_y estimation methods may provide more information on the nature and magnitude of this uncertainty (Bobko, Karren & Kerkar, 1987). Such research should reflect the decision context so that the implications of these findings can be meaningfully interpreted. The information value model suggests that valuable future UA research will address issues likely to alter decisions, in contexts where such alterations carry large consequences.

Recognizing the decision context reveals that UA models reflect an organizational process, not merely the single application of a particular program. The next sections review enhancements to the B-C-G utility model designed to better reflect such organizational processes. Such enhancements can have at least three purposes: (1) To provide more accurate and realistic utility values; (2) To improve the usefulness of UA models in enhancing decisions; and (3) To allow UA research to encompass a broader theoretical domain that advances scientific understanding of decisions about HRM programs. The first objective must be measured against actual or presumed objective values which may not be available. The second objective can be measured against the information value principles noted above. The third objective may be the most important for research, and can be measured against the ability of enhancements to incorporate and integrate fruitful new directions for scientific inquiry.

Expanding the Domain of Attributes In Selection Utility Analysis

UA models are special cases of MAU models, representing some, but certainly not all factors affected by HRM decisions. All MAU models, including UA models are undoubtedly deficient. This deficiency offers another possible explanation for utility values that may be higher than expected, or for the lack or widespread application of the models. The UA model may be missing important variables that are relevant to decision makers. Such deficiencies would be especially troubling if the omitted variables tend to argue against interventions, because UA models could produce positive utility values (suggesting program implementation), while a more complete UA model might reveal reasons against implementation. Moreover, because UA models focus on decisions to invest organizational resources in HRM programs, they implicitly draw on assumptions regarding both financial decision processes and labor market phenomena that interact with such decisions. We now explore how attributes from each of these related domains affect the B-C-G selection utility model.

Financial/Economic Considerations

The dollar-valued payoff function in UA models has led to speculation that UA models can provide a link between Personnel/HRM research and more traditional management functions (e.g., marketing, finance, accounting, operations). For example, Landy, Farr & Jacobs (1982, p. 38) suggested that UA models may be capable of "providing the science of personnel research with a more traditional 'bottom line' interpretation", Cascio and Silbey (1979) called for a "closer liaison" between personnel researchers and cost accountants, and Greer & Cascio (1987) proposed that cost accounting should contribute to defining the criterion in utility analysis. Even the original treatments of Brogden & Taylor (1950) and Cronbach & Gleser (1965) reflected a concern with the profit contribution of enhanced work force quality. Recently, researchers have suggested enhancements to the B-C-G selection utility model designed to incorporate financial/economic considerations.

Variable costs, taxes and discounting. Boudreau (1983a) recognized that UA models addressed economic and financial consequences of HRM decisions, but failed to incorporate certain financial/economic considerations. He suggested that measuring utility with a payoff function reflecting sales revenue or "the value of output as sold", would probably overstate HRM program effects on discounted after-tax profit (the payoff scale used for financial investments). He showed how the utility formulas could easily be altered to account for three basic financial/economic concepts: variable costs, taxes and discounting.

First, Boudreau noted the difference between "sales (or service) value" (i.e., the value of the increase in sales revenue or output as sold), "service cost" (i.e., the change in organizational costs associated with the increased revenue), and "net benefits" (i.e., the difference between service value and service costs) produced by an HRM intervention. He suggested that productivity enhancements through improved HRM programs may require additional support costs (e.g., increased inventories to support higher sales, increased raw materials usage to support higher output volumes, increased salaries/benefits as incentives for improved performance). Moreover, many interventions operate not by increasing sales revenue or output levels, but by reducing costs (e.g., Florin-Thuma & Boudreau, 1987; Schmidt, et al. 1986). He suggested including the effects of HRM programs on service costs in the model in either of two ways: (1) by reflecting the change in costs through a correlation coefficient (between the predictor and service costs) and the dollar-valued standard deviation of service costs among applicants; or (2) by assuming service costs are proportional to service value increases and simply multiplying the incremental service value increase by a proportion (1-*V*) reflecting the change in net benefits. Greer & Cascio (1987) derived variable costs more precisely using accounting conventions for soft-drink route salesmen. Boudreau (1983a) showed how incorporating such considerations could increase utility values (if costs fall when productivity increases) or decrease utility values (if costs rise when productivity increases).

Second, Boudreau noted that most organizations do not keep the full value of increases in net benefits. Rather, they must pay taxes on increased income to Federal, State and Local governments. Thus, adjusting utility values from the B-C-G model to reflect increases in net benefits may still overstate the organizational payoff by failing to account for increased taxes. Boudreau proposed multiplying both the net benefits and the implementation costs (*C*) by one minus the applicable tax rate (i.e., 1-*TAX*) to

reflect after-tax effects. He speculated that *TAX* levels might be as low as zero (for organizations reporting losses) and as high as .55 (for organizations subject to multiple income tax obligations).

Third, Boudreau observed that UA models typically focused upon benefits from interventions lasting into future time periods (the second column of Table 3 indicates the number of future time periods analyzed in empirical applications). UA models had treated such year-to-year effects as equivalent to each other. Returns derived in future years were simply added to the returns from initial years. He noted that such treatment ignored a fundamental fact of organizational management--money can be invested to earn interest. When interest can be earned, accelerated program returns and postponed program costs can be invested to earn interest for a longer period of time. Therefore, financial analysis "discounts" future earnings (and costs) to reflect these potential investment returns. Boudreau demonstrated how the interest rate earned on program returns (i) could be incorporated into the UA model, producing a "discount factor" (i.e., *DF*, the summed effects of discounting over a number of future periods). He demonstrated that discounting reduced utility value levels, with the most substantial reductions occurring when program returns occur farther into the future, and when the discount rate is high.

Boudreau (1983a) incorporated these factors into the selection utility formula and derived the combined effects of hypothetical levels of the parameters on reported utility values. His derivations suggested when HRM programs face zero tax and interest rates, and variable costs are reduced with productivity increases, B-C-G utility values might understate actual discounted, after-tax net benefits by as much as 33%. However, when HRM programs face positive tax and interest rates and costs rise with productivity, reported values might be overstated by as much as 84%. Studies applying financial/economic considerations suggest that unadjusted utility values commonly exceed adjusted values. As shown in Table 3, Mathieu and Leonard (1987) found TAX=.46, i=.15, and V=-.07; Burke and Frederick (1986) found TAX=.49; i=.18 and V=-.05; Rich and Boudreau (1987b) found TAX=.39; i=.15; and V=0.

Table 5 extends the example begun in Table 2, incorporating the financial/economic considerations noted by Boudreau (1983b), assuming the variable cost proportion equals 5%, the tax rate equals 45%, and the interest rate equals 10%. Program costs are assumed to occur at the beginning of the analysis, so they are not discounted but are adjusted only for their effect on taxes. Assuming a 10-year analysis period, unadjusted quantity equals 6,180 person-years. Unadjusted quality per person-year is $5,331 per person-year. Thus, the unadjusted product of quantity and quality is $39.125 million. This is adjusted to reflect the 5% variable costs and 45% taxes. Finally, to reflect discounting at a 10% rate, the 10-year quality effect is multiplied by .614,producing an after-cost, after-tax discounted net program benefit of $12.5519 million as shown in Table 5. Subtracting the after-tax testing cost of $6,698 produces the after-cost, after-tax, discounted net utility of $12.55 million as shown. Though substantially smaller than the $37.9 million reported by Schmidt, et al. (1979), derived in Table 2, this return remains substantial.

Insert Table 5 Here

Boudreau (1983a) stated that utility values incorporating these financial/economic considerations would better reflect the decision context of organizations that compute such investment values for

programs in other management functions, and might be more credible to managers accustomed to working with financial analysis. He also noted a number of theoretical implications. First, employee wages and salaries are a different concept from their productive value, with wages and salaries reflecting resource costs, while productive value reflects the output of applying human resources to a production process. Equating compensation with productive value will usually understate value, but in some cases will overstate it because wages may exceed production value for some jobs or individuals. As we shall see, the relationship between improved labor quality and compensation is central to labor economic theory, and this model provides a framework in which to integrate them. Second, utility analysis reflects temporal effects that may not remain constant over time, as the B-C-G model assumes, and might lead to biased utility value estimates. Although Schmidt, Hunter, Outerbridge & Goff (1988) found that validities and performance differences remained constant over time, temporal instability has been incorporated into utility models for training (Mathieu & Leonard, 1987; Schmidt, et al., 1982). Third, the enhanced financial/economic utility model might partially explain the unreliability observed in managerial SD_y estimates when managers are asked to use two conceptually different anchors--the value of output and the cost of contracting for that output--to derive one value (Day & Edwards, 1987; Reilly & Smither, 1985).

Applying Capital Budgeting Indices to Utility Analysis. Cronshaw and Alexander (1985) suggested that "a major reason for the differential success of human resource and financial managers in implementing their respective evaluation models is the greater rapprochement of capital budgeting with the everyday language of line managers and with the financial planning needs of the organization" (p. 102). They speculated that by integrating UA results into the financial decision making context, personnel managers would better communicate the impact of their programs on the "value of the firm" as opposed to "increased productivity" or "operating costs."

Cronshaw and Alexander (1985) separated the cost component of the selection utility model into two components, C_o, the original one-time costs of developing and validating a selection instrument, and C_i, the implementation costs incurred each time the instrument is used. The "return" (i.e, R) of the program was the one-year, one-cohort productivity increase from a selection device (i.e., the product of N_s, SD_y, $r_{x,y}$, and $\overline{Z}_x$). They explained the analogies between the selection utility model and five standard capital budgeting indices often discussed in financial investment textbooks. First, the pay-back period (PP) or "number of years a firm requires to recover its original investment from net returns" was formulated as the sum of C_i and C_o divided by R (a more consistent formulation would be C_o divided by the difference between R and C_i). The authors note that this index is deficient because it ignores interest earned on returns over time, and it ignores returns that occur subsequent to the payback period.

Second, they defined return on investment (ROI) as the ratio of "annual cash returns to original cost" and formulated it as the ratio of R to the sum of C_i plus C_o (a more consistent formula might be the difference between R and C_i divided by C_o). They noted that this index ignores interest returns, but it also ignores any multiple-year returns because it reflects only the one-year return from selection divided by the entire implementation cost.

Third, they defined "net present value" (NPV) as the difference between the discounted sum of returns over time (where the discount rate is the expected rate of return earned by the firm on contributed capital) and the original and implementation costs. This formulation is virtually identical to Boudreau's

(1983a) formula, but Cronshaw and Alexander do not account for variable production costs, multi-year implementation costs, and taxes.

Fourth, they defined the "profitability index" (PI) as the "ratio of the present value of net cash inflows to cash outflows", and formulate it as the discounted sum of returns (R) divided by the sum of implementation and original costs (a more consistent formulation might be the discounted sum of R minus C_i divided by the original costs). The authors note that a PI greater than 1.0 suggests a payoff exceeding costs as well as meeting the discount rate. They note that such an index fails to take into account the relative size of investments.

Fifth, they defined the "internal rate of return" (IRR) as "the rate which equates the NPV of cash inflows with cash outflows" and formulate it as the value of the discount rate that equates the discounted sum of the returns with the sum of the one-year implementation costs and the original costs (a more correct formulation might equate original costs with the discounted sum of the difference between the returns and implementation costs). The authors noted that the derived rate of return is then compared to the organization's required rate of return to determine project acceptability. An additional important limitation to this index is it's assumption that each project's returns would earn interest at that project's IRR, thus incorrectly implying different interest rates on the returns from different projects.

Cronshaw and Alexander briefly discuss the issues of taxation, application to non-selection programs, multi-year benefits and the flow of employees through the work force over time, which would make their financial models more compatible with Boudreau's (1983a; 1983b) derivations, and could address some of the inconsistencies noted above. They also provide a useful distinction between viewing HRM program expenditures as "operating costs" written off in the current period (presumably implying that program returns occur only in that period), and as "capital investments" (presumably implying multi-year future returns). They speculate that the reason for the low credibility and the presumed expendability of HRM programs may be that HRM managers fail to adequately communicate the multi-year benefits accruing from such programs. This point is analogous to an earlier observation made by Boudreau (1984a) which showed that break-even values suggested high HRM payoffs, as well as suggesting the even large cost outlays were justified by HRM program returns. However, two of Cronshaw and Alexander's financial indexes (payback period, one-year return on investment) will also understate multi-year benefits.

In fact, only two indices (net present value and profitability index) accurately reflect the relative discounted multi-year payoffs from competing financial investments. The profitability index is especially intriguing because it suggests considering payoff in terms of the benefits *per dollar expended*, rather than the benefits *minus dollars expended*, as is the traditional utility focus. It is straightforward to re-formulate utility equations to reflect this alternative perspective. It would be interesting to learn whether such reformulations would affect decision processes.

"Pitfalls" in Using Financial/Economic Considerations. Hunter, Schmidt & Coggin (1988, p. 522) proposed that financial accounting methods are "frequently inapplicable to human resource programs and, in addition, may sometimes have negative consequences even when they are applicable on a purely logical basis." First, they noted that except for discounted present value, the financial indexes discussed by Cronshaw and Alexander (1985) require that a portion of the costs be designated as the "investment" (e.g., Co), and they speculate that many improved selection systems may actually involve no original costs and/or may actually reduce ongoing testing costs. They correctly note that under such conditions,

one can compute discounted net present value (as described by Boudreau, 1983a), but not the other indexes. However, assuming costless HRM programs reduces the justification for *any* dollar-valued utility analysis, because any non-negative change in validity must produce increased value for the organization, as we have seen. Thus, this argument is less an indictment of capital budgeting than a recognition that UA models are best applied to decisions where programs compete for resources, and those resources are obtained at some cost.

Second, Hunter, et al. (1988, p. 524) argued that "discounting is meaningful only if there is in fact a delay in receiving the benefit." They correctly recognize that the B-C-G model reflects hiring only one cohort of employees, but when HRM programs are applied repeatedly their effects build as the work force becomes progressively more saturated with better-selected individuals (Boudreau, 1983b; 1987, 1988, in press; Boudreau & Berger, 1985). They observed that once the saturation point is reached (after Year 4 in their example), better-selected new hires simply replace departing better-selected employees, so the total work force value remains the same as long as the program is re-applied (we will discuss employee flows subsequently). They stated that "there is no such time delay in receiving the utility benefits once the program attains it equilibrium utility level" (p. 524) and that "if the program were used indefinitely, as is typically the case" the equilibrium value is constant for every future year. Technically, discounted and un-discounted utility values are equal for infinitely-long investments with constant returns. However, the investment *decision* must be made at the beginning of the program, not in the middle after it has reached equilibrium. Thus, the time delay *is* relevant. Moreover, the typical decision involves not one, but two or more competing alternatives, and different programs often reach equilibrium at different times, so that utility differences between programs will be affected by discounting. Hunter, et al. (1988, p. 593, Footnote 1) also noted that when businesses typically omit discounting from their investment evaluation methods, they may find that discounted utility values do not enhance credibility. This remains an empirical question, but considering the logically sound basis for discounting, it seems unlikely that discounting will actual detract from credibility.

Hunter, et al. (1988, p. 524) further proposed that even when programs involve investments, "return on investment and other capital budgeting figures for personnel programs, even when correctly calculated, will often appear to be extreme compared to other investment opportunities, and that, as a result, they could appear to management as lacking credibility." Their example, like the results in Table 3, demonstrate that the return on investment, payback period, profitability index and internal rate of return calculated for many HRM programs will be quite high. Yet, the adjustments to compute after-tax, discounted net benefits will frequently produce *lower* utility values than those derived from simpler models. For example, Boudreau (1983a, p. 569) noted that such adjustments would produce values that were 33% as large as the original values reported by Schmidt, et al. (1979), as demonstrated in Table 5. Thus, rather than threatening credibility, financial/economic factors can reduce utility values to potentially more credible levels, when compared to other financial investments. If they remain extreme after such adjustments, it is difficult to see how the unadjusted (i.e., more extreme) values are more credible than adjusted values.

As noted earlier in discussing the payoff function for SD_y, Hunter, et al. (1988, p. 526) correctly noted that adjustments for variable costs are often appropriate and consistent with utility definitions that emphasize cost reduction or profit contribution, and that the "value of output as sold" is appropriate only

when concern focuses on increases in revenue. They also acknowledged that "contribution to after tax profits ... might also have informational value for some purposes" (p. 527).

Financial accounting methods are not "frequently inapplicable" to utility analysis. The discounted present value model (Boudreau, 1983a) is inapplicable only for costless investments and/or investments with an infinite time horizon and equal temporal returns. Under such conditions, any dollar-valued utility model is of little value, and programs could be chosen based on their effect size alone. However, where programs require investments and where competing investments may produce different temporal returns (i.e., where dollar-valued utility analysis is applicable), the financial/economic utility model (Boudreau, 1983a; 1983b) integrates potentially important factors so that they can be explicitly considered.

Hunter, et al. (1988, p. 527) correctly concluded that "there is no single correct index of utility", and that "industrial/organizational psychologists and other human resource specialists should maintain the flexibility to match the information presented with the information needed for the purposes at hand," a position completely consistent with the arguments presented here and elsewhere (Boudreau, 1983a, 1983b, 1984a, 1987, 1988, in press; Boudreau & Berger, 1985; Boudreau & Rynes, 1985; Florin-Thuma & Boudreau, 1987; Rich & Boudreau, 1987). As we have seen, every utility model is deficient, including models that fail to adjust for financial/economic factors. However, this is not synonymous with saying that the financially/economically adjusted model is "the only legitimate definition of utility" (Hunter, et al., 1988, p. 526). Indeed, as we have seen, the model proposed by Boudreau (1983a; 1983b) encompasses both the unadjusted and adjusted utility values. As Boudreau (1983a) and Table 5 demonstrated, by setting the tax rate, discount rate and variable cost proportion to zero, one obtains utility values identical to those of the unadjusted B-C-G model. Explicitly recognizing such factors, encourages and enables managers and researchers to consider each factor's relevance to their decision, and then (if appropriate) focus on those that are most important. Hunter, et al. (1988, p. 527) argued that such a general model would produce an equation that is "so long and complex as to be daunting ... and as such is very difficult for personnel psychologists and human resource managers to understand." They claimed that "most managers and human resource personnel would rather deal with different models and different computational procedures for different cases." Research has demonstrated the financial/economic model's appropriateness (Boudreau, 1983b; Boudreau & Berger, 1985, Boudreau & Rynes, 1985; Burke & Frederick, 1986; Cronshaw, et al., 1986; Florin-Thuma & Boudreau, 1987; Mathieu & Leonard, 1987; Rich & Boudreau, 1987). Managerial preferences for general models versus inventing new models for every situation remain an empirical question, but this question can only be tested if such general models exist. Rauschenberger & Schmidt (1987, p. 57) correctly noted that it would be "presumptuous" to contend that research building such models should be "scaled down in lieu of an emphasis on better communicating existing utility methods to organizational decision makers."

Human Resource Accounting. Human Resource Accounting (HRA) models enjoyed substantial attention during the 1960's and early 1970's. These models were derived by researchers with accounting expertise, concerned with the fact that standard accounting reports provided no explicit mechanism for recognizing the contributions of human resources, in the same manner as capital or land resources. Flamholtz (1974, 1985) provided the most widely-cited treatment of the HRA model. HRA models arose out of a desire to provide accounting data for managers to use in HRM decisions. They were motivated by desires to put people on the balance sheet, to measure the value of human resources as assets to the

organization, and to better reflect the accounting consequences of managerial decisions, such as the possibility that a manager might achieve apparent short-term salary savings at the expense of hidden long-term productivity detriments by allowing high turnover to "liquidate" human assets.

HRA models address the "value" of human resources in two basic ways. First, the "cost" method (embodied in Flamholtz's "original cost" and "replacement cost" notions) measures or estimates what the cost would be of replacing a person or group of persons with another group capable of rendering the same value to the organization. This approach provides methods for measuring costs of separation, hiring and development, under the notion that these costs represent investments that should not be charged as expenses in the year they are incurred, but rather should be allocated over the tenure of individuals. Individuals have value to the extent that their investment has not been "amortized" (Flamholtz, 1974, p. 20). Such an approach has some merit as a mechanism for helping managers appreciate the magnitude of the costs associated with human resources. In situations where managers are inclined to allow excessive turnover because they do not appreciate its effect on costs, such a model might help decision makers realize that keeping experienced individuals has value. However, such a model ignores individual differences because it uses "typical" hiring, separation and development costs. Moreover, it fails to recognize job performance (implicitly equating the costs of replacing an individual with their value), and could lead managers to incorrectly attempt to reduce replacement costs by reducing turnover when in fact the opposite strategy might better enhance productivity.

Second, the "asset" method (embodied in Flamholtz's "human resource asset value") measures or estimates "the present worth of the set of future services the person is expected to provide during the period he or she is expected to remain in the organization" (Flamholtz, 1985, p. 173). Such a model considers the discounted stream of productive value and the costs incurred to maintain and improve that productive value over an individual's useful life with the organization. The aggregate value of human resources is the total productive value generated by the employee group over time. Such a concept is logically consistent with the notion of human resources as assets, and it effectively focuses attention away from acquisition costs and toward net productive value. Such a model rapidly becomes highly complex (requiring estimates of future productivity, future career progressions, probabilities of turnover, probabilities of dismissal, probabilities of death, etc.). Like the "cost" approach, it assumes all job incumbents achieve typical or average productivity in a job, ignoring individual differences.

Both the "cost" and "asset" models focus on measurement rather than decision making. Neither model explicitly proposes how HRA information should be used to make decisions concerning HRM programs. Conceivably, one could compute the replacement cost consequences of different managerial decisions causing different levels of turnover, and use that information to compare the decision options. One could also conceivably compute the "asset value" of a work force likely to result from different HRM program decisions, to determine the program producing the greatest asset value. However, both methods imply ongoing, detailed and complex measurement tasks and to not emphasize individual responses to HRM programs. Thus, they fail to address the program decisions fundamental to HR management.

Traditional investment decisions for new plant and equipment must translate the values on a balance sheet or income statement to reflect the discounted cash inflows and outflows of that particular investment decision. Similarly, even if HRA data were available, specific HRM program decisions must focus on

the relative investment value of different options, so the ongoing and detailed measurement tasks implied by HRA models may not be necessary for many HRM decisions.

Still, HRA research provides extremely useful measures for HRM program costs. Also, HRA models focus attention on the long-term impact of HRM decisions on work force productivity, suggesting that HR managers avoid considering programs in isolation or considering only the short-term impact of HRM programs. As will be discussed subsequently, it is possible for UA models to incorporate this focus in a manner that emphasizes decision making as opposed to measurement.

Behavioral Costing. Recently, Cascio (1982; 1987) proposed integrating the cost focus of HRA with behaviors. His behavioral costing approach consists of the "quantification in financial terms of a set of common behavioral and performance outcomes. Standard cost accounting procedures are applied to employee behavior" (Cascio, 1987, p. 7). This approach measures the cost consequences of behaviors such as turnover, absenteeism, and smoking to quantify the effects of HRM programs reducing those behaviors. While such an approach links the costing methodology of HRA with a more behavioral focus, it may also cause decision makers to overemphasize cost reduction as opposed to performance enhancement. This can lead to incorrect decisions when cost reductions are gained at the expense of performance (e.g., where banning smoking causes better performers to leave or be absent; or where reductions in turnover reduce opportunities to replace poor performers with better performers). Cascio (1987) discusses some of these limitations, but does not integrate these behavioral costing methodologies with selection utility. Moreover, the behavioral costing methodology suggests ongoing behavioral measurement which may not be necessary for all HRM program decisions. Still, Cascio's (1987) behavioral costing treatment provides several costing systems for employee behaviors and suggests one link between HRA costing procedures and HRM decisions.

Equal Employment Opportunity

Equal Employment Opportunity (EEO), and Affirmative Action (AA) programs are often adopted or mandated by Courts. In the U.S. discrimination against protected groups is illegal, and organizations are held responsible for guarding against such discrimination. All HRM programs are subject to examination for their discriminatory effects, though most attention has centered on staffing decisions (usually the use of selection predictors). Unlike the factors discussed above, EEO and AA reflect "equity" rather than "efficiency" (Milkovich & Boudreau, 1988). However, in view of the importance of these issues to organizations, it seems quite likely that decision makers take them into account when considering HRM programs, especially selection. Such equity considerations may help to explain decisions that appear at variance with the UA model prescriptions. For example simple top-down hiring may produce the highest utility, but if minorities score lower the test can reject a substantially higher proportion of minority group members (Wigdor & Hartigan, 1988). Such selection devices are likely to be more closely scrutinized by Government agencies, so decision makers may reject them in favor of methods presumed to be legally "safer", such as setting low cutoff scores, or using less valid procedures that produce less adverse impact. Utility analysis can help to quantify the economic impact of such decisions. Although existing legislation appears to allow more valid selection devices even if they reject more minority applicants, it is up to the organization to demonstrate that there is clear justification for the devices--that the devices are a "business

necessity". Dunnette, et al. (1982, p. 26) noted that UA models can provide evidence of "business necessity" (defined as the need for improved selection in order to reduce instances of ineffective and costly job performance).

No research has examined whether EEO considerations explain the failure to adopt more-valid selection devices (or other HRM programs to enhance productivity). However, several authors have recognized EEO in their utility analyses, providing mechanisms for computing the effects of different EEO or AA strategies on selection utility. Kroeck, Barrett & Alexander (1983) developed a simulation depicting the effects of different affirmative action policies on minority hiring and performance levels. Schmidt Mack and Hunter (1984) used utility data for U.S. Park rangers to examine the effects of setting minimum cutoff scores at different points in the applicant population distribution (such minimum cutoffs might be adopted in an effort to allow less-qualified minorities to meet hiring standards). They found setting the cutoff at the mean reduced utility values to 45% of the utility of top-down hiring, and setting the cutoff one standard deviation below the mean reduced them to 16% of top-down hiring. Finally, Steffy and Ledvinka (1986) developed a simulation designed to examine the utility consequences of selection strategies based on three different definitions of "fairness". As Schmidt, Mack and Hunter (1984, p. 496) conclude, "the question can be raised whether those employers currently using the low cutoff method of employment test use are aware of the large price in productivity that they are paying." Wigdor & Hartigan (1988) also showed how alternative race-conscious scoring systems can enhance minority representation with smaller reductions in the standardized productivity level of selectees. It remains for future research to determine whether such estimates actually affect decisions.

Integrating Labor Economics and I-O Psychology Through Utility Analysis

The hypotheses and findings of labor economics seem well-suited to provide some insights into future UA research. Moreover, the UA framework suggests that theories of labor economics can be tested more rigorously at the level of individual decisions, in contrast with the more traditional focus of labor economics at the level of the firm or the economy. We can identify a number of issues suggesting a bridge between labor economics, personnel management and I-O psychology.

Labor economics has traditionally dealt with the implications of economic theory for the behavior of organizations and individuals in a labor market. A labor market is the arena in which individuals provide a *supply* of limited labor resources to organizations, with the characteristics of the labor resources depending on individual/organizational characteristics and decisions regarding the "human capital" investments they make (e.g., education, training, willingness to re-locate, etc.). Organizations *demand* labor resources from individuals, depending on the needs of the organization's services or production processes. Thus, a market of labor supply and demand exists, with both parties bargaining to identify the prices at which certain labor resources will be supplied to fill demand. Labor economic investigations often focus on the behavior of wages and the quantity of labor employed as a function of various individual and organizational characteristics, to determine whether such behavior can be explained using the labor market model.

Although several I-O psychology research areas are related to labor market behavior (e.g., motivation theory is often studied within the context of organizational reward systems), seldom are the two disciplines explicitly related. Labor economics tends to presume a market-based pattern of choices and

behaviors at the individual or firm level (or at least presumes that useful predictions can be made by assuming such patterns), and focuses on national or industry-wide trends. I-O psychology often focus on the individual behaviors and choices related to organizational attributes (e.g., predicting the likelihood of separation as a function of satisfaction with various organizational attributes), but devotes less attention to the implications for aggregate behavior among those supplying and demanding labor resources. Clearly both perspectives are important to fully understand the implications of HRM program decisions. UA research will inevitably lead to greater concern with labor-market-related concerns and implications, because it inevitably directs attention to the dollar-valued market performance of the organization (Boudreau, 1988).

Labor Markets Require a Price for Labor Force Quality. UA models generally ignore labor market reactions to improved selection (or other HRM programs), reflecting an assumption that an organization adopting more valid selection is the only firm affected by that decision. However, if other firms become aware of the more valid procedure, competitive pressures can provide an incentive for them to adopt it as well, increasing competition for higher-quality applicants, and initiating bidding for these applicants through increased wages or other incentives (Boudreau, 1983b, p. 404). Becker (1988) noted that this can occur even if firms are unaware of their competitors' improved selection, because those without such selection will observe the decrease in applicant quality as their competitors pull the best qualified applicants out of their applicant pool. Such effects become more likely as the duration of the selection program increases, and under conditions where organizations adopt selection procedures with unequal validities. Bishop's (1987, p. 239) analysis of the U.S. Bureau of Labor Statistics data also suggests that "workers' background characteristics have large and significant effects on both starting and latest relative wage rates" and "realized productivity has almost no effect on the starting wage when background is controlled but large and significant effects on wage rates after a year or so at the firm." Thus, firms apparently pay more for better-qualified workers, both initially and throughout their employment. As the financial/economic model showed, higher wages represent variable costs that rise with productivity, so realized utility values are likely to be lower than those computed using a simple revenue-focused model.

Recognizing that compensation responds to labor quality leads to concern with the "incentive effects of screening" (Bishop, 1988; Mueser & Maloney, 1987). Mueser and Maloney (1987) argue that if more valid selection devices measure stable abilities that are unaffected by individuals' efforts to become more productive, then using such devices may be unwise because they reduce the incentive for applicants to invest in self improvement. Why should applicants work hard in school or take technical training if their job prospects depend on an innate general trait that cannot be affected by such activities? However, Bishop (1988, p. 2) argues that "greater use by employers of tests measuring competence in reading, writing, mathematics and problem solving will increase the supply of these competencies." He presents secular, cross-sectional and longitudinal data on IQ-type tests suggesting that traits measured by these tests are indeed malleable. Thus, increased test use associated with greater incentives (e.g., wages) for better-qualified workers may increase the average quality of applicants, though utility will reflect not only greater productivity but higher variable costs as well.

Raising wages is only one of many ways to improve applicant inducements (others include more intensive recruiting, job re-design, improved career opportunities), but virtually all inducement

improvements are likely to increase variable costs in utility estimates. UA models that systematically incorporate this variable offer opportunities for needed integration between I-O psychology and labor economic theories.

Labor Markets Determine Applicant Population Characteristics. Incentive effects are one example of how selection programs affect not only the incremental value compared to random selection, but the average and distribution of value in the applicant population itself. For example, it is widely recognized that higher selection ratios (less choosiness) will lower selection utility levels. As labor demand rises (unemployment falls), the number of job applicants may also fall, leading to higher selection ratios.

Labor economic theory offers additional insights about how changing labor market conditions determine characteristics of the applicant population even if firms endeavor to maintain the same number of applicants through increased recruiting efforts. We have seen how firms demanding highly-qualified workers pay higher rewards to attract them, producing a hierarchy of firms with applicant pools that reflect the average and variability in qualifications available at each firm's pay level. Becker (1988) noted that increased labor demand (i.e., lower unemployment) can reduce the supply of higher-qualified candidates for lower-paying jobs. Highly-qualified applicants, who originally represented the upper tail of the applicant distribution for lower-paying jobs, now can qualify for higher-paying jobs, and are scarcer in the pool for lower-paying jobs. Fewer better-qualified applicants suggests a lower average applicant qualification level.

Becker (1988) suggests this may also lower the effective validity coefficient in the new applicant pool, due to truncation at the upper end of the predictor-criterion distribution. However, it seems possible that validity may not fall if the bottom tail of the distribution is increased as recruiting efforts are stepped up. Thus, incremental utility (i.e., the difference between random and more-valid selection) may not fall as much, but the value of selectees will fall because the average applicant value is lower. Utility estimates originally made under conditions of high unemployment and plentiful labor supplies will overstate actual utility values when the opposite conditions occur. While the opposite phenomenon is likely to occur in times of rising unemployment (i.e., more highly-qualified applicants will be forced to enter the pool for lower-paying jobs), the short-run hiring advantages may be subsequently negated because these individuals may leave the organization to take jobs in higher-paying organizations when labor demand increases again (Becker, 1988).

Thus, labor economics predicts that selection decisions are affected by the quality of the applicant pool, which may change with changing market conditions. Moreover, these changes are not only reflected in the parameters of the B-C-G selection utility model, but also through variable costs, average applicant value/costs, and future separation rates and patterns.

Labor Markets Determine Training Consequences. The B-C-G utility model has been extended beyond selection to training activities (Schmidt, Hunter & Pearlman, 1982). As with selection, however, labor economic theory demands that utility models acknowledge the effects of such programs on service costs and separation patterns. Becker (1975) distinguishes "general" versus "specific" training. General training refers to training that can be readily used by competing firms (e.g., word processing, computer programming). Specific training refers to training that is useful only to the organization providing it (e.g., operating a patented production process unique to the organization). Becker (1988) proposes that if

training is general, other organizations will be willing to pay the firm's trained employees to leave and work for them (essentially buying the value of training by paying higher wages), rather than provide the training themselves. Thus, general training will either increase employee separations (among the most valuable trained employees) or increase variable costs of wages or other rewards to induce trained employees to stay. In the case of specific training, employers can profit from training if the wages they pay following training are below the increase in productivity (but still high enough to induce trained employees to stay, because no competing employer has an incentive to pay more), or if employees are willing to accept lower wages during training (but not so low that the entire training cost is paid by the individual because then the firm could lay off the employee and train someone else at no loss). Specific training costs are predicted to be "shared" through some combination of lower wages during training and higher (but below-productivity) wages after training. Thus, with both specific and general training labor economic theory would predict that variable costs and separation patterns are likely to substantially affect training utility, and that omitting such factors may lead to overstated training effects.

Summary. Labor economic theory (proceeding from a basic premise of competitive labor markets and price-based decisions and behaviors) suggests several intriguing integrative research issues. Investigating these issues requires utility models incorporating recruitment, variable costs, employee separations, and internal employee movement (promotions, transfers, and demotions). Moreover, they demand explicit recognition that programs such as selection, affecting the flow of employees into, through and out of the work force interact with HRM programs such as training, affecting the existing stock of employees.

Utility models recognizing these issues would permit integrative research drawing on insights from both I/O psychology about the actual behaviors and decisions of employers and employees, and labor economics about market-based predictions concerning labor supply and demand. UA models offer a bridge that can help integrate the two disciplines. We currently have little information on whether HRM programs truly produce the turnover and reward consequences suggested by labor economic theory. At the same time, most I-O psychology research is notably deficient in addressing individual behaviors within the context of alternative opportunities and labor market variables.

As shown in Table 3, the vast majority of UA research focuses on employee selection. The B-C-G utility model emphasizes translating validity coefficients into terms more relevant to organizational decisions. However, as we have seen, HRM program consequences are certainly not limited to selection programs. UA models have the potential to integrate virtually all HRM program decisions, but that requires a UA theory encompassing HRM program consequences that affect employee flows as well as the existing stock of employees (Boudreau, 1987, 1988, in press). We now develop such a framework, categorizing HRM program decisions and their consequences into two types: (1) Programs/Consequences affecting employee "stocks" by changing characteristics of the existing work force or work situation (e.g., training, performance feedback, compensation); and (2) Programs affecting employee "flows" by changing the composition of the work force by adding, removing, or re-assigning employees (e.g., recruitment, selection, turnover, internal staffing).

Changing the Characteristics of the "Stock" of Employees

"Programs affecting employee stocks (such as training, compensation, performance feedback and employee involvement) aim to increase valuable characteristics (such as skills, abilities , or motivation) among existing employees to improve their current job performance. In terms of quantity, quality, and cost, decisions affecting employee stocks enhance productivity more when they affect a broad range of employees and time periods, cause large average increases in the value of employee job behaviors, and achieve both effects at minimum cost. Thus, decisions affecting employee stocks 'work' by improving employee behaviors in their existing assignments." Boudreau (1988, p. 1-134).

The Characteristic-Changing Utility Model

Landy, et al. (1982) and Schmidt, et al. (1982) first applied UA concepts to employee stocks. Both studies reformulated the utility model, with Schmidt, et al. (1982) focusing on training programs, and Landy, et al. (1982) focusing on performance appraisal and feedback programs. They recognized that in the B-C-G utility model, the product of the validity coefficient and the standardized predictor score of selectees represented the estimated difference between the average standardized criterion score of randomly-selected applicants and the average standardized criterion score of better-selected applicants. One could consider this standardized difference the effect of selecting "treated" (i.e., better-selected) applicants as opposed to "untreated" (i.e., randomly-selected) applicants. Other HRM programs could also be considered "treatments" (i.e., training, compensation, performance feedback), so it was logical to extend the utility concept to encompass them. The problem with a direct extension, however, is that characteristic-changing programs do not operate by choosing which employees to add to (or remove from) the work force. There is no predictor score, so there is no validity coefficient, no selection ratio, and no estimate of the average standardized predictor score of selectees.

However, characteristic-changing program effects are often reported in statistical terms (just as selection program effects are often reported in terms of validity coefficients). Landy, et al. (1982) and Schmidt, et al. (1982) noted the direct relationship between standard effect size measures (i.e., t and F statistics) and the correlation coefficient, and proposed transforming such statistics into d_t, the true difference in job performance between the treated and untreated groups, in standard deviation units. This concept is similar to the product of the validity coefficient and the standardized predictor score of selectees in the selection utility model. Thus, just as the B-C-G selection utility model provided a framework for placing the correlation coefficient into a more managerial perspective, so the Landy, et al. (1982) and Schmidt, et al. (1982) enhancements of the utility model place statistical findings from other personnel program interventions into a managerial perspective. Moreover, as is true for the validity coefficient, recent advances in cumulating findings from many studies to better identify generalizable effects (i.e., meta-analysis, Hunter, et al., 1982) can also be used to examine studies of characteristic-changing programs (e.g., Hunter & Hunter, 1984; Locke, Farren, McCaleb, Shaw & Denny, 1980; Burke & Frederick, 1985), producing d_t estimates even when experimental study is inappropriate or impossible in the decision situation.

The characteristic-changing model is analogous to the selection utility model in that it also reflects the three fundamental utility variables--quantity, quality and cost. The quantity is the number of person-years affected by the program; The quality change is the product of the effect size (i.e., d_t) and a scaling factor translating this standardized value into dollars (usually SD_y); the costs of developing, implementing

and maintaining the program are defined similarly to the costs of employee movement programs (though the actual cost components will differ).

The employee stock utility model can be reformulated to reflect financial/economic considerations (Boudreau 1983a), repeated program applications over time (Boudreau, 1983b), and can be analyzed using a break-even framework (Boudreau, 1984a; 1988). Expressed in terms of quantity, quality and cost, the employee stock utility model can be integrated with utility models for employee flows (discussed next). In addition to the Schmidt, et al. (1982) and Landy, et al. (1982) treatments, Mathieu & Leonard (1987) and Florin-Thuma and Boudreau (1987b) applied UA to a training program and a performance feedback program, respectively. As noted in Table 3, both studies' findings suggested substantial program benefits and low break-even values.

The Scaling Factor

In selection utility, there is a conceptually clear population upon which the utility parameters are based--the applicant population. While identifying characteristics of this population may be difficult, the model is nonetheless internally consistent in that all utility parameters refer to this population. However, in the characteristic-changing model the focus population is not as clearly defined. The treatment is given to the entire employee group, so two populations exist during the intervention--the pre-treatment population and the post-treatment population. Consistency would suggest that all utility parameters be based upon the same population, but applications of the model generally derive the d_t statistic from t or F statistics based upon the estimated standard deviation of the *pooled* samples, and derive the scaling factor (i.e., SD_y) based upon the pre-treatment group (Landy, et al., 1982; Schmidt, et al., 1982) because organizational experts can seldom estimate SD_y among the pooled group. Yet, the pre-treatment and pooled SD_y values may differ if the program alters within-group performance. For example, if training alleviates severe performance problems by moving the low performers closer to the mean, then the pre-treatment standard deviation will exceed the pooled standard deviation, biasing utility estimates upward.

At least two approaches might resolve this dilemma. First, when estimating d_t, researchers might re-scale it in terms of the pre-treatment group variability. Second, researchers might focus on the performance difference induced by the program *in actual production units*. Florin-Thuma and Boudreau (1987b) found that a high-quality measure of employee performance consequences (i.e., the level of inventory needed to support production) demonstrated substantial change associated with a performance feedback program. This production measure reflected the total group's performance in each period, rather than individual performance differences. Thus, it was not necessary to measure the program's effect on a per-person, per-period basis because the performance measure already reflected the per-period effect for the entire treated group. Moreover, this performance index was easily translated into dollars using standard inventory cost figures, thus circumventing the need to measure SD_y or to express the experimental results in standardized form. This second strategy focuses the utility analysis more directly on the program's consequences and their value to the organization, rather than on deriving a scaling factor (i.e., SD_y) to translate it into dollars. However, this process is very situation-specific, and unlikely to produce results that are as easily cumulated across studies as d_t values. Measuring performance effects for entire employee groups might appear to move away from "psychological" variables (because it moves away from measuring individual behaviors), but such an approach is necessary to reflect the concept of program utility, and may often be a more accurate representation of decision maker's objectives for HRM

programs.

Conclusion

Extending UA to encompass "characteristic-changing" programs allows increased application of UA models, as well as integration between HRM program planning in different functional areas. So far, this UA model has been applied only to decisions concerning whether to adopt a program or not (similar to selection utility applications focusing on whether to replace a relatively low-validity predictor with a highly-valid one), but the potential exists for a much broader integrative perspective. HRM decision makers need not consider their programs as competitors, when program combinations may produce higher productivity enhancements than individually-designed and evaluated programs. A program to enhance performance feedback and a program to improve training could be addressed separately by estimating which program provides the highest utility if used individually, but such a strategy ignores the fact that there are actually four options: (1) do neither program; (2) do training alone, (3) do performance feedback alone; and (4) do both programs. UA models can be applied to all four options, and may demonstrate that a version of the fourth option is superior to the others (Boudreau, 1984a, p. 213). Thus, this UA model provides the potential for integrative HRM strategies that draw on program interactions. This potential for integration is even more apparent when one realizes that UA models can apply not only to characteristic-changing programs and selection programs, but also to virtually any HRM program whose consequences alter the pattern of employee movement into, out of and through the organization. The next section develops such an employee movement utility model--the utility model for employee flows.

Changing the "Flow" of Employees

> "Employee flows occur when employees move into, through, and out of an organization through selection, promotion, demotion, transfer and separation. In terms of quantity, quality and cost, decisions affecting employee flows enhance productivity more when they impact large numbers of employee flows and time periods, greatly increase the value of job behaviors through better person-job matches, and achieve these goals at minimum cost. ... Programs affecting employee flows 'work' by improving the pattern of movements into, through, and out of the organization so that more valuable employees are placed in jobs or work roles." (Boudreau, 1988, p. 1-134).

The B-C-G selection utility model reflects the consequences of HRM programs that add one group of better-selected employees to the work force. However, actual managerial decisions seldom hinge on the consequences of hiring only one cohort of new employees. The selection program decision usually occurs in an environment where employees are dismissed or choose to leave, where they are selected or choose to move to other positions within the organization, and where new groups of employees will flow into, through and out of the organization over time. Decision makers will often wish to consider the effects of different selection systems on these other movement patterns. A selection program could appear to produce high utility values when considering only its effects on the first cohort hired, but may produce changes in other movement patterns over time that might negate such high utility values (e.g., where turnover is increased because better-qualified employees have more external opportunities). Moreover, HRM decisions contain opportunities for substantial synergy between programs affecting employee movement. For example, the benefits of improved selection may be enhanced by retaining the best-performing employees or moving them into higher-level jobs. The one-cohort selection utility model can be enhanced to reflect HRM decisions and consequences affecting employee movements between

positions.

A Definition of Employee Movement

Employee movement may be defined as *"the establishment, alteration, or termination of the employment contract between an individual and an organization"* (Boudreau & Berger, 1985b, p. 33). Jacques (1961) defined the employment contract as "an implicit or explicit agreement between employees and employers in which the employee carries out designated tasks or objectives in return for payments over a specified or (usually) unspecified time period" (Boudreau & Berger, 1985b, p. 34). This definition is consistent with previous employee movement definitions and taxonomies (though previous research had focused on one type of movement at a time, usually turnover).

Boudreau and Berger (1985b) proposed a taxonomy of employee movement in which each movement type could be characterized by four attributes: (1) whether the employee was previously a member of the organization; (2) The direction of movement (inward, outward, upward, downward, laterally); (3) the permanence of the movement (i.e., whether the duration of the movement is specified in advance); and (4) the decision maker (i.e., employee, employer, or both). They demonstrated how this taxonomy distinguished employee movements based on the degree of discretion allowed the decision maker, the types of information available, and the certainty with which outcomes can be predicted, among other variables. They distinguished between *external* employee movement that involves crossing the organizational boundary by initiating or terminating an employment contract, and *internal* employee movement involves altering an employment contract but not terminating it.

Like other UA models, movement utility models focus on three variables: (1) the quantity of movers, (2) the quality of the movers; and (3) the costs incurred to produce the movement (Boudreau, 1984c; 1987; 1988; in press; Boudreau & Berger, 1985a; 1985b). Because these three basic variables are common to all employee movements, we can derive an extended utility model that simultaneously encompasses the consequences of decisions affecting not only selection, but decisions affecting other types of external employee movement and internal movement as well.

The Employee Flows Utility Model

Organizations seldom invest in a selection program to use it once and then stop, but continuously reapply the program as new members enter the work force. To analyze only the first-cohort effects is tantamount to a financial analyst attempting to analyze an investment in new manufacturing facilities by assuming the facilities will only be used for one production run. Clearly, such a focus omits a large part of the decision's effects. Boudreau (1983b) redefined the one-cohort selection utility model to encompass the flow of employees into and out of the work force over time.

Boudreau's (1983b) "employee flows" utility model was derived by changing the quantity of employee-years of production to reflect employee inflows and outflows over a given evaluation period. Previous models reflected the quantity of employment-years as the product of the number hired times average tenure. Boudreau's model used the number of "treated" (i.e., better-selected) employees entering the work force in each future period (i.e., N_{a_t}) and the number of "treated" employees separating from the work force in each future period (i.e., N_{s_t}) to compute the number of treated employees in the work force

in each future time period (i.e., N_k). The utility in each future period was the product of N_k times the incremental quality per person-year (i.e., the product of $r_{x,y}$, $\overline{Z}_x$, and SD_y), minus the costs incurred to select the employees joining the work force in that future time period (i.e., C_k) Summing these values over all future time periods of analysis (k=1...F) produced the total utility of the selection program. Boudreau also incorporated the effects of discounting, variable costs and taxes into his formulation (Boudreau, 1983b, p. 399, Equation 6). Though this formula reflected constant utility parameters over time, Boudreau showed how to incorporate temporal changes.

Insert Table 6 Here

Table 6 continues the earlier example, incorporating the effects of selecting multiple cohorts over time. The number of incumbents, separations and acquisitions, test information, and financial information remain the same. New parameters now reflect the analysis period and test application period, which combine to produce the *leverage*, or number of person-years affected by the selection program: The analysis period is 10 years by convention (Boudreau & Berger, 1985a), and the test is assumed to be re-applied for seven years.

In each of the first seven years, 618 better-selected new hires are added to the work force, replacing 618 employees selected without the PAT. Because each cohort stays for 10 years, the number of better-selected programmers in the work force steadily increases by 618 employees in each of the seven years until (in Year 7) the work force is virtually saturated with better selected workers. In Year 7, 4,326 (i.e., 7 X 618) out of 4,404 programmers have been selected using the PAT. All 4,326 programmers stay for the remaining three years of the ten-year analysis. Thus, the total leverage is 31,282 person-years (618 + 1,236 + 1,854, and so on).[2]

The calculation at the bottom of Table 6 reveals the effects of multiple-cohort selection. After-tax, discounted selection costs increase substantially (from $6,798 in Table 5 to $0.04 million in Table 6), but selection program returns also increase substantially (from $12.55 million to $54.32 million), producing a total multiple-cohort utility value of $54.28 million. Repeatedly applying improved selection programs can have massive potential productivity effects because of their huge *leverage*. Just as one would not attempt to justify a million-dollar investment in a new manufacturing plant based only on the first production run, HRM decision makers should not evaluate HRM programs based only on the first cohort affected.

---

[2] Boudreau (1983b) assumed the PAT would be re-applied for 15 years, that 6,180 job vacancies existed, and analyzed effects for 25 years. The effect of these assumptions was that the number of better-selected employees in the work force steadily rose (by 618 per year) during the first 10 years until it reached 6,180. Then, in Years 11 through 15, vacancies were filled with better-selected employees, so the number of treated employees in the work force remained at 6,180. When the program was terminated, the number of treated employees in the work force slowly diminished (by 618 each year) until it reached zero in Year 25. This produced higher costs, leverage and utility values. Table 6 adopts assumptions more consistent with financial convention and the size of the reported computer-programmer work force.

Mathieu and Leonard (1987) and Rich and Boudreau (1987) incorporated the concept of cohort flows through the work force into their analyses. Rich and Boudreau employed a methodology that incorporated period-to-period differences in turnover due to group tenure. Mathieu and Leonard (1987) incorporated period-to-period differences due to training effect dissipation.

Integrating Recruitment Into Selection Utility Analysis

Boudreau and Rynes (1985) noted that while the early Taylor-Russell selection utility model explicitly included the "base rate" (i.e., the proportion of applicants whose performance would exceed minimally acceptable levels if randomly selected), subsequent utility models did not reflect this factor. Indeed, the majority of selection utility research was conducted under the implicit or explicit assumption that all selection options would be implemented within the same applicant population. Such assumptions probably simplify organizational reality. We have already how labor economic theory suggests that applicant populations change both independently and as a result of selection decisions. Boudreau and Rynes (1985) noted the common belief that more rigorous or intrusive selection methods may affect the size and/or characteristics of applicant pools, and that recruitment strategies (e.g., personalized follow-ups, realistic job previews, choices of recruitment sources) are explicitly designed to alter applicant population characteristics, presumably to enhance organizational outcomes. Boudreau and Rynes (1985) derived a selection utility model that can explicitly incorporate the effects of recruitment, reflecting financial/economic factors and employee flows.

Every parameter of the B-C-G selection utility model could be affected by applicant reactions to recruitment/selection strategies. For example, applicant populations might become more homogeneous (reducing both SD_y and the correlation coefficient) if more stringent recruitment standards were used. Or, higher salary offers might increase the size and perhaps the qualifications of the applicant pool, affecting both the selection ratio and the average qualification level of the population.

Although the Taylor-Russell model had explicitly incorporated a variable reflecting the average level of applicant qualifications (e.g., the base rate), this variable was removed from later utility models, which adopted an "incremental" focus, computing utility values by comparing selection strategies to random selection. In the Boudreau-Rynes model, utility values are represented on an absolute scale, reflecting both the average and the incremental value of the selectees. The Boudreau-Rynes model encompassed the observation by Alexander, Barrett and Doverspike (1983) that self-selection and initial organizational screening might cause "examinees" to be a non-random sample from the applicant population, and reflected the observations by Hogarth and Einhorn (1976) and Murphy (1986) that job offer rejections can affect selection utility.

Boudreau and Rynes showed how recruitment effects might alter the conclusions of utility models reflecting only selection. Improved selection may offer less *incremental* utility (compared to random selection) when recruitment practices produce a more qualified but less diverse applicant pool. However, an integrated recruitment-selection strategy applied to such an applicant pool can produce the greatest total value despite reducing *incremental* selection utility. The key is whether the increased average applicant value is great enough to offset the reduction in incremental selection utility, so that the combined strategy becomes more economically attractive. Utility models reflecting only incremental selection utility cannot

reflect this possibility because they assume constant applicant population parameters, and they omit the average applicant qualifications.

Insert Table 7 Here

Table 7 continues the running example of computer-programmer selection, integrating recruitment strategy options. It estimates the returns from selection when combined with two competing recruitment strategies--recruitment advertising or a recruitment agency. Recruitment advertising produces an applicant pool with diverse qualifications but a moderate average applicant value, because advertising reaches a wide audience but provides little pre-screening. The recruitment agency generates a less diverse applicant pool with a higher average applicant value due to the agency's screening of applicants before referral. The upper part of Table 7 shows the variables that are assumed not to change as a result of recruitment. Utility is assessed using the same 7-year application of the staffing program, and 10-year analysis period.

The variables assumed changed by recruitment are shown in the bottom of Table 7. Recruitment advertising costs $2,500 per hire, while the recruitment agency costs $4,450 per hire (American Management Association, 1986). Recruitment advertising is expected to produce an applicant pool similar to the present one, so validity is .76 and SD_y is $10,413 as before. Through pre-screening, the recruitment agency generates applicants with less variability, reducing validity to .60 and SD_y to $8,500 per person-year. Net applicant value for advertising-recruited applicants is $15,620, reflecting an average service value of $52,065 and average service costs of $36,445 per person-year. Agency pre-screening is presumed to identify higher-quality applicants, with a net value of $20,000, reflecting higher average service value of $60,000 per person-year, partially offset by higher service costs of $40,000 per person-year (including higher salaries/benefits to attract and retain these applicants).

The expected value of each new hire is the sum of two values: The value produced by random selection--hiring average-value applicants from a particular applicant pool, plus the incremental value produced by systematic selection from that applicant pool. Thus, the expected value of those hired from the agency-generated pool is the average value of the pool (i.e., $20,000) plus the incremental value added by selection (i.e., .60 X .80 X $8,500, or $4,080) or $24,080 per person-year. Similarly, the expected value of those hired from the agency-generated pool is $15,620 plus $6,331, or $21,951 per person-year. As before, these quality levels are multiplied by the quantity of person-years affected by the selection program (i.e., 31,282) and adjusted to reflect financial/economic considerations, recruitment costs and selection costs. The bottom section of Table 7 shows the total, after-tax, discounted value of the employee flows. It separates the effects into the work force value and movement costs if only Random Selection were used (i.e., the only quality difference is created by recruitment), and the incremental work force value and movement costs added by Testing.

If only selection utility is considered, the testing pays off under either recruiting strategy, although its payoff is smaller when applied to agency-generated applicants rather than advertising-generated applicants ($34.96 million versus $54.28 million). However, the agency-generated applicant pool produces a much higher average value than the advertising-generated pool ($180.5 million versus $141.04 million). Integrating the effects of recruitment and testing demonstrates the advantage of combining agency recruiting with testing (i.e., $207.45 million versus $190.76 million). Sacrificing some testing

effectiveness for an increase in average applicant quality makes sense.

HRM decision makers may often face opportunities to integrate recruitment with selection. A recruitment-selection utility model is necessary to accurately reflect such situations, and provide a framework for explaining and enhancing managerial decisions. For example, in spite of its low validity, the recruitment interview still enjoys widespread application while more valid cognitive ability tests do not. One explanation for such decisions may be that decision makers feel accurate selection is not necessary because their recruitment activities have already identified such a well-qualified pool of applicants. Decision makers may feel that virtually any applicant from such a pool will be an acceptable performer, using the interview primarily to attract enough applicants to fill existing job openings. Indeed, subsequent research (Rynes and Boudreau, 1986) found that filling job openings was the primary measure of recruitment effectiveness, and that variables related to subsequent job performance (i.e., performance ratings, turnover) were seldom even recorded. Thus, the enhanced recruitment-selection utility integrates two highly complementary organizational processes, and encourages future recruitment and selection research.

The External Employee Movement Utility Model

Drawing on the substantial similarities between employee acquisitions and employee separations in their effects on work force value (Boudreau, 1983b; 1984b; 1984c; 1987; 1988; in press; Boudreau and Berger 1985a; 1985b; Milkovich & Boudreau, 1988), Boudreau & Berger (1985b) developed a utility model that could encompass not only the effects of employee acquisitions but also employee separations. UA research typically ignores the potential effects of HRM programs on the quantity and pattern of employee separations (e.g., quits, layoffs, dismissals). Similarly, employee turnover research usually focuses on describing the cognitive processes leading to turnover, but not on the cost and benefit consequences of turnover. Boudreau & Berger (1985b) proposed to integrate and enhance both utility analysis and turnover research with a utility reflecting employee acquisitions and separations. Because both acquisitions and separations involve crossing the organizational boundary by initiating or terminating an employment contract, this is an "external employee movement" model.

The external movement utility model was developed to encompass three related phenomena: (1) repeated acquisitions without separations over time (where the work force is increased through selection); (2) repeated unreplaced separations over time (where the work force is reduced in future time periods); and (3) repeated separations over time that are replaced with acquisitions. The first case reflects the focus of most selection utility models; the second case reflects decisions about using layoffs, attrition or dismissals to reduce a work force; and the third situation is the most general case, capable of encompassing both of the other two. Figure 1 depicts the concepts underlying the Boudreau-Berger external movement utility model. Each box describes a component of the utility model. Figure 1 presents two periods of employee acquisitions and separations with the work force value in each prior time period serving as the starting point for the utility effects on the subsequent time period. Box A represents the work force utility at the beginning of the analysis (i.e., in the period prior to implementing programs to change the quantity or quality of employee movement). Box C represents the work force

utility at the end of the first period, serving as the starting point for the next period (Box E), as indicated by the line connecting the two boxes.

Insert Figure 1 Here

In each time period, two processes may occur to change work force utility. First, employees may be added. The utility of acquisitions in the first time period (t=1) is represented by Box B. The utility of the acquisitions becomes part of the utility of the work force following acquisitions, as indicated by the arrow from Box B to Box C, and by the description within Box C. Second employees may separate. In the first period, this is shown in Box D. These separations will affect the quantity and/or quality of those retained from the beginning work force, as indicated by the arrow between Boxes A and C and by the description within Box C. In the second period (shown in Boxes E through H), the same process occurs, but the beginning work force utility reflects the work force at the end of the first time period, so the quantity, quality and costs of acquisitions and retentions may differ from the first period. Finally, as indicated at the bottom of Figure 1, the process is assumed to continue for the duration of the utility analysis (Time Periods 3 through F). The utility values produced by this model reflect the sum of the discounted, after-tax net benefits of the work force in each period of analysis (e.g., the sum of Boxes A, C and E for the two-period illustration in Figure 1).

Figure 1 demonstrates the close analogies between selection and retention utility. Selection utility involves choosing a subset of employees to join the work force from a pool of applicants. Retention utility involves a subset of the previous-period's incumbent work force choosing or being chosen to remain with the organization. Though retentions are somewhat more bilaterally chosen than acquisitions (Boudreau & Berger, 1985a; 1985b) the analogy is still correct. The utility of both acquisitions and retentions depends on the *quantity*, or number of employees hired and retained; the *quality*, or per-person, per-time period effects of selection and retention patterns; and the *costs* incurred to implement or accommodate the movements, such as selection device development/implementation, severance pay, and relocation assistance (Boudreau, 1987; in press, 1988; Boudreau & Berger, 1985a; 1985b; Milkovich & Boudreau, 1988).

Building on these analogies, Boudreau and Berger (1985a) derived an algebraic utility model encompassing both acquisitions and retentions. This utility model represented a multiple-cohort acquisition and retention utility model, and was capable of reflecting the effects of both types of movement on organizational outcomes. Boudreau and Berger summarized the derivation as both a formula (Equation 14, p. 594) and as shown in Figure 1. The Boudreau-Berger model is expressed in terms of the absolute value of the work force, rather than the incremental value added by improved selection.

Boudreau & Berger (1985b, pp. 598-599) demonstrated that their model was a more general case of both the one-cohort selection utility model and the employee flows utility model. They described the assumptions necessary to produce the two previous models from the more general external movement utility model, and discussed conditions under which such assumptions might be appropriate, concluding that a utility analysis based only upon selection consequences often risks producing not only deficient utility values, but values that could lead to faulty decision making.

Insert Table 8 Here

Table 8 continues the example of computer-programmer selection using the external movement utility model. The size, number selected and number separating per year, financial/economic considerations and 10-year analysis period are all the same as before. We assume recruitment through advertising, so the selection and recruitment parameters correspond to the Recruitment Advertising portion of Table 7. The test validity, number of applicants, standard test score, testing cost, average applicant service value, average applicant service cost, and SD_y remain the same as before, with each group of 618 acquisitions producing an average value of $21,951 per person-year.

Each acquisition is assumed to carry costs of $7,000 per year, reflecting the $2,500 recruitment cost, relocation, orientation and other administrative activity. Each separation likewise carries a cost of $7,000, reflecting administrative activity, outplacement assistance exit interviews, severance pay and other activities undertaken when separations occur (Boudreau & Berger, 1985a). Such costs are incurred regardless of the quality of the person joining or leaving. Because the external movement utility model does not assume hired cohorts stay intact for the duration of average tenure, we can now drop the assumption of a 7-year test application period, and adopt the more realistic assumption that the test is applied throughout the 10-year analysis.

Because the analysis now focuses on the total work force value, rather than on simply the value of those added and retained, we assume that at the beginning of the analysis the work force resembles the average of the applicant population. Average incumbent service value per person-year is $52,065, and average incumbent service cost per person-year is $36,445, for a net value of $15,620 per person-year.

The Work Force Utility Results reflect two levels of test validity--random selection and validity of .76, contrasted with three separation pattern effects. The separation effect reflects whether the organization retains its better or poorer performers, and is the average performance difference between those retained and the pre-separation work force. Boudreau & Berger (1985a) noted that this parameter would usually be directly observed, but estimated as the product of the standard deviation of service value in the pre-separation work force and the standardized difference in average service value between the retained and pre-separation work force (i.e., $d_{sv_{r_t}}$). Under assumptions similar to those used to estimate the average standardized test score from the selection ratio, $d_{sv_{r_t}}$ can be estimated from the "retention ratio"--proportion of the work force retained. Boudreau & Berger (1985a) assumed values for $d_{sv_{r_t}}$ ranging from -0.26 to +0.26, and an incumbent service value standard deviation of $10,413, producing a range of separation effects from -$2,707 to $0, to $2,707 per person-year. The negative value reflects the assumption that the organization retains its worst employees, with those retained producing an average service value $2,707 less per person-year than the pre-separation work force. Zero assumes retentions are random with respect to service value, so those retained produce an average service value equal to the pre-separation work force. The positive value reflects the assumption that the organization retains its best employees, with those retained producing an average service value $2,707 more per person-year than the pre-separation work force.

A specially-designed LOTUS 1-2-3 (R) personal computer program (Boudreau, 1984b) simulated various acquisition and retention strategies (Boudreau & Berger, 1985a). The results of four of the

strategies are shown in Table 8. Under Option 1, the organization selects and retains randomly, attaining a 10-year work force value of $200.31 million. Under Option 2, valid selection is introduced, but retentions remain random, increasing the work force value to $242.10 million. Option 3 illustrates the best of worlds--highly-valid selection combined with retaining the best employees, producing the highest work force value of $351.69 million. Option 4 acknowledges that better-performing employees may have more opportunities and separate more often, in this case causing the worst employees to be retained, producing a total work force value of $132.50 million.

The interaction between separation and selection patterns is obvious. Considering only the incremental value of improved selection suggests a $41.79 million increase in work force value (i.e., $242.10 million - $200.31 million). However, combining improved selection with improved retention can provide an additional $109.59 million (i.e., $351.69 million - $242.10 million). Conversely, dysfunctional retention patterns can disrupt the effects of improved selection, as illustrated by Option 4, where the work force value is $67.81 million *lower* than even random selection and retention. If more-valid selection acquires high-quality employees who leave in response to better opportunities, projected selection utility can be substantially reduced. While these effects are based on a specific set of assumptions, the model allows them to be explicitly manipulated to examine their relative effects (Boudreau, 1988).

Though the simulation was intricate, the results could be expressed as a linear Equation (Boudreau & Berger, 1985a, Table 30), as shown in Equation 11, with dollar values expressed in millions.

$$U_w = \$200.31 + \$54.95\ (r_{x,sv_{a_t}}) + \$421.53\ (d_{sv_{r_t}}) \qquad (11)$$

Where:

U_w = the total discounted, after-tax, after-cost work force utility summed over 10 future analysis periods,

$r_{x,sv_{a_t}}$ = the correlation between the selection device score and service value among job applicants in Future Time Period t, and

$d_{sv_{r_t}}$ = the standardized service-value difference between those retained and the pre-retention work force in Future Time Period t.

The substantially larger coefficient on the retention utility parameter (i.e., $d_{sv_{r_t}}$) relative to the correlation coefficient, suggested that the utility effects of HRM decisions on retention patterns could be substantial, and that models failing to acknowledge these retention effects risk ignoring important organizational outcomes. Omitting retention considerations can severely bias selection utility estimates when improved selection either causes the retention pattern to become less optimal or the retention pattern causes the value of improved selection to be quickly lost. Turnover and selection research can be better integrated, with both areas attending to the effects of the other.

The Boudreau-Berger model demonstrates the danger in focusing only on the separation (or turnover) *rate*. HRM program utility should reflect not just the quantity of employee acquisitions and separations, but also the pattern of separations/retentions relative to employee value. The costs of separations or the characteristics of those who leave and stay (e.g., Cascio & McEvoy, 1985; Dalton, Krackhardt, & Porter, 1981; McEvoy & Cascio, 1985) must be considered in light of the effects of those acquired to replace the separations. Thus, the external employee movement utility model integrates and expands employee

movement research. Moreover, such a model encourages integration between I/O psychological theory and labor economic theory, where employee mobility is a central concept (Bishop, 1987, p. 240; Gerhart, 1987). Still, even this integrated model focuses only on the utility consequences for one job in the organization. A more complete perspective would encompass movement between jobs within the organization as well (Boudreau & Berger, 1985a; 1985b).

Integrating Internal and External Employee Movement Utility

Boudreau and Berger (1985b) described similarities and distinctions between *internal employee movement* that alters an employment contract but does not involve crossing the organizational boundary (such as promotions, demotions and transfers) and *external employee movement* that does involve crossing the organizational boundary. Internal staffing research usually describes internal movement patterns or examines the effects of career processes on individuals, but less frequently addresses the effects of career systems on organizational performance and the reasons for using different internal staffing arrangements in organizing the employment relationship (cf. Milkovich and Andersen, 1982, p. 382; Pfeffer and Cohen, 1984, p. 550). The effects of *external* movement on organizational performance have received more attention, but such effects interact with internal employee movement, so a full analyses demands an integrated framework.

Selection/retention programs that appear optimum for a single job may have substantial consequences for internal movement. For example, if improved selection for lower-level jobs also identifies skills and abilities useful in upper-level jobs, then more-valid external acquisition strategies may produce substantially higher benefits than the simple selection utility model, or even the external movement utility model, can recognize. Conversely, selection devices targeted exclusively to skills applicable only in the entry-level job may appear valuable in a single-job model, but if employees routinely move from that job to upper-level jobs using other skills, then maximizing entry-level selection utility may simultaneously reduce utility in the upper-level job.

Finally, evaluating internal selection devices based only on their validity for jobs receiving employees (e.g., Cascio & Silbey, 1979) may miss important negative consequences for the lower-level jobs that lose the internal movement candidates. High-performing engineers are commonly promoted from technical engineering positions into managerial positions. Such a strategy may produce acceptable managerial job performance but, if it removes the best engineering talent from the technical jobs, may actually decrease organizational effectiveness. All of these phenomena require integrating the consequences of internal and external employee movement and identifying variables likely to determine the utility of such movements--an integrated external/internal movement utility model.

Boudreau's (1986b) utility model draws upon the analogies between internal and external employee movement, proposing that each internal employee movement involves a separation from one organizational job and an acquisition by another. Thus, the pattern of internal employee movement can be analyzed using the concepts of selection and retention utility, but must recognize that both types of utility are affected by the same movement (Milkovich & Boudreau, 1988, Chapter 13).

Boudreau (1986b) proposed that modelling the relationships between internal and external employee movement would reveal substantial opportunities for optimizing employee movement decisions. Therefore,

the internal/external employee movement utility model encompasses the Boudreau-Berger (1985a) external employee movement utility model while recognizing internal movement utility consequences. Boudreau (1986b) illustrated the internal/external movement utility model using the simplified hypothetical external and internal movement system depicted in Figure 2.

Insert Figure 2 Here

In this movement system, external selections and internal movements fill vacancies created by external separations. Job B represents an upper-level job that experiences *external separations* (those that leave the organization). To fill these vacancies, individuals are moved through *internal selection* from a lower-level position (Job A) to Job B through a promotion-from-within policy. Job A also experiences external separations. Thus, the organization must make *external acquisitions* into Job A to fill the vacancies created by both internal and external separations. It is the quality, quantity and cost of these four movement types that determines total work force utility over the analysis period. In each analysis period the total work force utility is the sum of the work force value in the two jobs, minus the costs of accommodating internal/external movements occurring during that period. Figure 2 depicts the work force value in the two jobs initially (Boxes A and B), and following the movements occurring in the first analysis period (Boxes G and H). The full utility model tracks these effects and calculates the discounted, after-tax, net work force value in the two jobs over the period of analysis (t=1...F).

Three types of movement affect Job A. First, employees separate from Job A and leave the organization (depicted in Box C of Figure 2), so the utility of job A's work force (Box G) reflects the quality and quantity of employees retained after these *external separations*. Second, employees move from Job A to Job B (depicted in Box E of Figure 2), so the utility of Job A's work force (Box G) also reflects the quality and quantity of employees retained in Job A after these *internal separations*. Third, after external and internal separations have reduced Job A's work force, *external acquisitions* occur to bring the work force back to its original level (depicted in Box F of Figure 2), so Box G reflects the quality and quantity of these external acquisitions.

Two types of movement affect Job B in Figure 2. First, employees separate from Job B and leave the organization (depicted in Box D), so the utility of Job B's work force (Box H) reflects the quality and quantity of the employees retained as a result of these *external separations*. Second, to fill the vacancies, employees move to Job B from Job A (depicted in Box E), so the utility of Job B's work force (Box H) reflects the quality and quantity of the employees acquired through these *internal acquisitions*.

The total utility value is the discounted, after-tax, after-cost sum of the work force utilities in Jobs A and B over the period of analysis. In Figure 2, this sum would include the work force values represented by Boxes G and H (for Time Period t=1), plus any subsequent work force values affected by movement in Future Time Periods 2 through F. Though this movement system is simplified, the concepts of internal and external separations and acquisitions apply generally, even to more complex systems. The jobs serving as sources of employees represent applicant populations for internal acquisitions, just as external applicant populations are the source of external acquisitions. The model analyzes internal separations similarly to external separations by focusing on the quality and quantity of

those retained, recognizing that the pattern of internal separations (i.e., promotions or transfers out of a job) will probably differ from the pattern of external separations (i.e., separations from the organization).

The sequence of employee movements is important when considering internal and external movement utility (Boudreau, 1986b). If external separations occurred from Job A before internal promotion decisions were made, then the internal applicant pool for promotion into Job B would not include those who externally separated from Job A, and vice versa. In reality, internal/external movements do not all occur in a group, but occur throughout each time period. This model can encompass such phenomena simply by choosing time periods for analysis that are short enough to meaningfully capture the movement pattern (as has often been done in Markov analysis) or to adjust the initial value of employee movements to reflect that the movers only occupy the job for a partial period (see Rich & Boudreau, 1987).

This model reflects the effects of employee characteristics for both their current job and for jobs representing potential destinations of internal movement. Selection utility models recognize that skill/ability differences between job incumbents implies performance differences (e.g., SD_y) for their current job. The internal/external movement utility model recognizes that the same skill/ability differences may also affect performance differences in internal destination jobs. If the destination job is a higher-level job allowing more discretion, SD_y among job incumbents in their *current* job can be lower than the SD_y of the same employees when considered as the applicant pool for the destination job. Thus, the ability of promotion to enhance incumbent job performance differences is explicitly modelled. Similar relationships exist for other utility parameters, such as the validity, selection ratio and average predictor score.

The internal/external employee movement utility model encompasses the one-cohort selection utility model, the employee flows model, and the external employee movement model. Internal movements have measurable consequences not only for the jobs that internally acquire employees but for the jobs that internally separate employees as well. Such implications are seldom discussed, with virtually all research preferring to focus on the consequences of internal movements for the receiving job. However, whenever an internal acquisition takes place, it is associated with an internal separation that has consequences for the jobs that lose employees.

Insert Table 9 Here

Table 9 extends the computer-programmer staffing example to encompass internal employee movement consequences. The example now encompasses an upper-level job (Job B in Figure 2) of Data System Manager, presumed to serve as a destination for promotions among Computer Programmers (Job A in Figure 2). Internal selection is assumed to occur through an assessment center, and all separations among Data System Managers are assumed to be filled from the Computer Programmer job.

The external staffing variables for the programmer job are the same as before, except that instead of 618 new hires to replace external separations, this example requires 718 new hires to replace both the 618 external separations and the 100 promotions out of the programmer job. The financial/economic considerations and 10-year analysis period are the same for both jobs.

One-hundred external separations occur from the Manager job, replaced by 100 internal acquisitions. Each separation from the manager job costs $8,000, slightly higher than the $7,000 cost for programmers,

and these costs are incurred regardless of the quality of retentions. Each promotion from programmer to manager also costs $8,000 (including relocation, orientation, administration, etc.) regardless of promotion quality. Moreover, internal selection uses an assessment center which, at an average cost of $380 per tested applicant (Cascio & Silbey, 1979) produces a total cost off $1.44 million per year to assess all 3,786 promotion candidates.

Because the managerial job involves more discretion and responsibility, Average Applicant Service Value among programmers when considered as promotion candidates is assumed to be 10 percent higher than the average value of the same employees as programmers. The ratio of Average Applicant Service Value for the manager job divided by average incumbent service value in the programmer job is 1.10. Average Service Costs are also 10% higher when programmers are promoted to managers, reflecting higher average salaries, benefits and other employment costs. As the value of the programmer work force changes as a result of external selection and retention, programmers' value as promotion candidates also changes. Decisions that improve the programmer work force produce an added benefit by improving promotion candidates for manager jobs even when promotion candidates are selected randomly, and vice versa for decisions that worsen the quality of programmers. Stronger or weaker relationships between individual differences among programmers and managers could be modelled by changing this parameter.

The external staffing variables affecting the programmer job are analogous to the internal staffing variables affecting the Manager job. The applicant pool for promotions is the group of 3,786 programmers (4,404 incumbents - 618 separations) available each year. This assumes that all programmers are promotion candidates, but it could easily be adjusted for situations where only a limited number of programmers are eligible or tested. With 3,786 applicants for 100 job openings, the organization can be quite choosy, so the average standard assessment center score of those promoted is 2.32 standard deviations above average (using the Naylor-Shine table with a selection ratio of 100/3,786). Performance differences among programmers considered as managerial candidates are presumed to be about 10% larger (SD_y = $11,454) than among applicants for programmer jobs (SD_y=$10,413).

Although the algebraic model is intricate, it's explicitness permitted simulation using a LOTUS 1-2-3 (R) personal computer program (Boudreau ,1986a) that greatly simplified the analysis. The bottom of Table 9 shows the effects of four different internal/external staffing patterns, representing the after-tax, after-cost, discounted work force value in both jobs, summed over the 10-year analysis period. Option 1 depicts random external and internal staffing. Under such a system, the average value of each job's work force remains constant as internal and external movements occur, producing a total 10-year value of $249.86 million. Option 2 introduces valid external selection using the selection test with a validity of .76. This enhances the value of programmers, which in turn augments the value of the managerial work force when programmers are promoted, producing a total work force value of $296.90 million. Option 3 analyzes internal staffing in the typical manner, acknowledging the validity of the assessment center (presumed to equal .35, Cascio & Silbey, 1979) for internal acquisitions, but still assuming that promoting highly-qualified programmers has no effect on the quality of the programmer work force. Total work force value increases to $302.51 million. Finally, Option 4 considers the possibility that internal promotions will pull high performers from the programmer work force, reducing the average value of the

retained programmers by $625 per person-year.[3] This produces a total work force utility of $278.58 million. Although the assessment center validly predicts future job performance for managers, its negative impact on the programmer work force costs the organization $12.22 million compared to random internal staffing (Option 4 minus Option 2).

It would be inappropriate to conclude from this hypothetical analysis that assessment centers always represent poor investments, but it illustrates that internal selection programs which pull the best employees from lower-level jobs can have serious organizational consequences--consequences that are virtually ignored by simple selection-utility models. Typical internal staffing analyses that consider only the quantities or rates of movement between jobs will also omit the effects of such movements on the quality of the work forces in the internal staffing system. Substantial work force quality differences emerged in Table 9, despite the fact that the quantity of movements was held constant.

Modifying the model parameters allows extending these concepts to encompass other decisions affecting internal/external employee movements, such as "make-or-buy" decisions between internal and external selection, reductions in work force size, and internal staffing systems involving more than two jobs. Of course, such extensions require considering a larger number of parameters, and these parameters are likely to represent estimates under uncertainty. Computer-based analysis permits sensitivity analysis to explore the implications of such uncertainty. Boudreau (1986b) simulated combinations of different parameter values for seven effects, producing a linear equation showing the relative impact of seven model parameters. Boudreau (1988) demonstrated how break-even analysis could also be applied to the integrated internal/external movement utility model.

Summary. The actual decision situations facing human resource managers and program planners are seldom as simple as the choice between two selection devices, evaluated based upon the productivity of the first cohort selected. HRM decisions, even if they only involve new selection devices, are likely to affect employee separations and employee internal movement patterns (Boudreau, 1987; in press; 1988; Ledvinka & Ladd, 1987). The effects of these phenomena may act to enhance or reduce selection utility. Moreover, HR managers frequently face situations in which it is possible to optimize decisions affecting employee movement. Investments in improved selection may be combined with investments designed to improve recruitment, separation patterns or internal movement patterns. In situations where resources are limited, it may not be optimal to devote all resources to one task (e.g., improved selection), but to combine programs, producing synergistic effects that surpass those achievable by only one employee movement program. An employee movement utility model that encompasses both external and internal movement can accommodate such decision situations. Moreover, such a model extends the theoretical domain of utility analysis because the internal movement responses of individuals and organizations are often studied in sociology and labor economics. Integrating these theories with I/O psychology research can produce a broader understanding of the implications of HRM program decisions.

---

[3] The $625 value was derived based on the retention ratio for promotions (i.e., 3,686/3,786), the Naylor-Shine tables, and the presumed incumbent programmer standard deviation of $10,413 (Boudreau, 1986b; Boudreau & Berger, 1985a). In actual applications, this parameter could be estimated directly based on differences between the retained and pre-promotion work force.

Is The Complexity Really Worth It?

A common reaction from reviewers and commentators on the movement utility models is that they are complex and detailed. Such enhancements might be

> "expanding the models to a point where their practical application may be jeopardized. The complexity of these new and expanded models will make it very difficult for researchers, let alone practicing industrial psychologists, to fully comprehend their implications and communicate the models and their findings to organizational decision makers." Rauschenberger & Schmidt (1987, pp. 56-57)

From the standpoint of communicating and improving decisions, UA model enhancements should be evaluated according to the value of the information they add, as noted in earlier sections of this Chapter, and elsewhere (Boudreau, 1984a; 1987; 1988; in press; Boudreau & Berger, 1985a; Boudreau & Rynes, 1985; Florin-Thuma & Boudreau, 1987; Rich & Boudreau, 1987). The value of the added considerations for enhancing particular decisions will be situationally specific, depending on the probability that it would change decisions in important ways, and on the costs of incorporating it into the decision process (Boudreau, 1984a). *Measuring* the entire movement utility model will prove unnecessary when the enhanced model is unlikely to improve decisions based on a simpler framework. Simple one-cohort selection utility models derived from Equation 7 have proven quite popular with researchers (See Table 3) and have produced useful insights. Still, without enhanced models reflecting variables such as discount rates, variable costs, recruitment, turnover and internal movement, it becomes much more difficult to explicitly distinguish situations where simple models are sufficient from those where added complexity adds value. More complex models help explicitly identify when additional measurement is unnecessary, rather than simply ignoring these considerations. Moreover, when simple one-cohort selection models produce biased or misleading results by omitting these variables (Boudreau, 1983a, 1983b, 1984a; 1986; 1988; Boudreau & Berger, 1985a; Boudreau & Rynes, 1985), these models allow measuring variables that can enhance decisions (see Tables 5 through 9).

The employee movement utility model, integrated with the utility model for effects on employee stocks, and reflecting financial/economic considerations, represents progress toward a more general utility model. With such a model, utility analysis research can include or exclude variables as appropriate, acknowledging such practices explicitly. Hunter, et al. (1988, p. 527) speculated that "the resulting equation is so long and complex as to be daunting. Furthermore, for any particular application, it contains numerous irrelevant terms and as such is very difficult for personnel psychologists and human resource managers to understand." The accuracy of these speculations remains an empirical question, but simulation results (Boudreau, 1986; Boudreau & Berger, 1985a; Boudreau & Rynes, 1985) suggest that recruitment, separation and internal movement will often be relevant, and personnel psychologists and HR managers have long recognized interactions between these phenomena, but have had few integrative models to describe them. Moreover, with increasing computational power, such integrative models offer frameworks for developing computer-based models that greatly ease managerial effort required to apply them (Boudreau, 1984b; 1985; 1988; Ledvinka & Ladd, 1987).

Aside from their practical ability to enhance decisions, integrated utility models bridge research topics

in I/O psychology (e.g., test development and validation, career effects on individual behaviors and attitudes, cognitive processes affecting turnover) with topics often considered by other behavioral sciences such as sociology (e.g., the demographic patterns and causes of internal movement) and economics (e.g., the effect of HRM programs on employee qualifications, internal and external labor market behavior, and wages). The integrated utility analysis perspective offers a step toward forging an interdisciplinary approach to such important topics--an approach necessitated by the myriad of organizational consequences affected by HRM programs and decisions. Lacking such integrated frameworks, future research risks becoming parochial and narrow, vastly limiting its potential for describing, predicting and explaining decision processes. "It would be presumptuous, of course, to contend that research in utility analysis should be halted or scaled down in lieu of an emphasis on better communicating existing utility methods to organizational decision makers" (Rauschenberger & Schmidt, 1987, p. 57). In fact, research that extends utility analysis to encompass and integrate such important variables should be encouraged.

Future UA Applications and Research

Although the UA models were initially developed to address selection decisions, we have seen that the UA framework is really a special case of Multi-Attribute Utility models applied to HRM program decisions. Viewed in this way, the model has great potential for studying HRM program decisions in virtually every functional area of Personnel management. Our review of empirical research suggests that UA applications are embryonic, with selection utility demonstrations dominating reported utility values. Moreover, existing research seems fixated upon the measurement properties of one particular selection utility parameter (SD_y). While this state of affairs should not be surprising in view of the fairly recent resurgence of attention to UA issues, the potential integrative role of UA models remains untapped.

A Framework for UA Research

Insert Figure 3 Here

Figure 3 depicts a matrix of future UA research directions. The rows (A through J) of the matrix represent specific content areas of HR management that can serve as the focus of UA research. Except for the top and bottom rows, these areas fall generally into the categories used to describe HRM activities (Milkovich & Boudreau, 1988), but they could be expanded to include several additional research areas from Organizational Behavior or I-O Psychology. For example, the "Selection" area (Row D) includes issues of test theory and job analysis (Burke & Pearlman, 1988). Row A refers to research that develops general models or frameworks applicable across functional areas, such as financial and break-even utility models. Row J refers to research that examines decision making processes, also spanning one or several functional areas.

The columns of Figure 3 represent types of research activity that can contribute to each area identified by a row. These research types progress from developing a conceptual framework that includes the Row's content (Column V), to simulation analyses demonstrating the potential effects of the concepts on organizational outcomes (Column W), to empirical demonstrations actually measuring UA parameters

and deriving utility estimates for different functional programs in different settings (Column X). With sufficient simulation and empirical application, it becomes possible to infer the behavior and boundary conditions of UA results across settings and applications (Column Y). Finally, with well-developed and widely-applied utility models, it is possible to test general theories regarding those models (Column Z).

Existing Research

The asterisks in the cells of Figure 3 represent an admittedly crude index of existing research activity. Cells with asterisks have received attention, while those without have not. As we have seen, substantial progress had been made in the first columns, across several functional areas. Concepts extending Outcome Evaluation (Cell A-V) have been developed to include financial/economic considerations, Equal Employment Opportunity, Affirmative Action, risk, and uncertainty. Simulations (Cell A-W) have demonstrated substantial potential effects of these variables, and extended models been applied to actual program decisions (Cell A-X). Future valuable research could incorporate less quantitative HRM program consequences. Tsui (1984, 1987) and Tsui & Gomez-Mejia (1988) proposed a measure of personnel department effectiveness based upon the reputation of the Department among its "important" constituencies. Tsui's unit of analysis is the personnel *Department,* while UA focuses on HRM *program* decisions, but HRM program consequences will be evaluated by a wide group of constituents on non-financial as well as financial attributes. Factors outside the UA model, such as labor union pressures, public opinion, and organizational tradition may well determine program decisions in actual organizations. Can a personnel department increase its reputation for effectiveness if it communicates the consequences of its decisions using UA models? If UA models reveal suboptimal organizational traditions, such as selecting employees through unstructured interviews, would they actually be changed? UA research should examine these questions.

As in Row A of Figure 3, the functional areas of Recruitment, Selection, Training, Internal Movement, and Turnover/Layoff analysis (Rows C, D, E, G, H and I) also exhibit concept development, simulations and empirical demonstrations (Columns V, W and X).

In Data-Based Inference (Column Y), meta-analysis and validity generalization are producing findings about the distributions of program effects across settings and studies. Cell D-Y reflects substantial evidence of selection procedure validity generalization (e.g., Hunter & Hunter, 1984; Schmidt, et al., 1982). Indeed, Table 3 suggests tentative conclusions about boundary conditions on selection utility, such as that substituting highly-valid predictors for much less-valid ones apparently pays off unless the costs are extremely high (Burke & Pearlman, 1988, p. 125). Others have conducted meta-analyses on Training programs (Burke & Frederick, 1985), turnover programs (Cascio & McEvoy, 1985), and other HRM programs (Locke, et al., 1980) as reflected in Cells E-Y, H-Y and I-Y. Valuable future research will link these effect-size estimates into a decision context.

Developing Utility Models for Other HRM Functions

An obvious gap in Figure 3 is the absence of UA concepts in several functional and theoretical areas. Little research addresses how existing UA models apply to Compensation decisions (Row F) such as pay policies, reward structures, and benefits (Milkovich & Newman, 1987). Yet, substantial and identifiable organizational resources are constantly being invested here. It seems likely that UA concepts reflecting

characteristic-changing consequences on employee stocks can encompass compensation decisions. In this regard, compensation is similar to training or performance feedback. As Figure 3 illustrates, fruitful future research will identify the concepts, simulate and apply them, paving the way for inferences and theory testing (Row F).

Similarly, Human Resource Planning (Row B) has not been well addressed by UA research. This is surprising because the conceptual link between the two areas is so clear. Human resource planning "ensures that the human resource decisions that managers make are integrated and directed toward achieving objectives" (Milkovich & Boudreau, 1988, p. 270). The integrated movement utility model can reflect the productivity implications of human resource planning, and the synergy between human resource programs (Boudreau, 1986), as illustrated in Tables 5 through 9. Future research should apply the integrated utility model to examine the implications of various HR planning systems and decisions. It is necessary, but not sufficient, to "fill in" the matrix by developing models and demonstrations for each different type of HRM program. The ultimate contribution will be derived by adopting a synergistic and integrative perspective on UA research. Research might begin with the planning process, where strategic program decisions are made. Little information exists on how decision makers decide what program combinations to implement, whether they consider the interactions between programs in different functional areas and their effects on organizational outcomes. Simulations reflecting such an integrative model (Boudreau, 1986; Ledvinka & Ladd, 1987) represent a start (Cell B-W). Demonstrations including programs from several functional areas (Cell B-X), such as training and selection, are also promising. It seems likely that the classification and variable-treatment selection models developed by Cronbach and Gleser (1965) will be relevant here (Human Resources Research Organization and others, 1984).

Column Z of Figure 3 suggests theory-testing research addressing the functional issues in each row. This type of research is the logical step that can integrate the demonstrations and parameter-focused research of Columns X and Y. Undoubtedly these decision theories will have unique attributes depending on the functional area they address, it is certainly not currently possible to outline theoretical frameworks for each Cell of Column Z. However, it is important to recall the links between I-O psychology and labor economic theory discussed earlier, and the potential for enhanced UA models to support an integration between these social sciences. It seems likely that future research addressing Column Z will draw upon this and other related social science theories.

Decision Processes and Contexts

The bottom row (Row J) of Figure 3 reflects HRM Decision Processes, perhaps the most fundamental, important and complex issues facing future UA research. This Chapter has noted that while UA results are often presumed to influence decisions, enhance credibility, and encourage a broader decision focus, existing research has not empirically investigated these phenomena. Future research must examine whether the UA results affect managerial decisions, whether decision makers' reactions to UA results are affected by different parameter estimation techniques, and whether UA models accurately reflect decision makers' concerns. UA models serve to describe, predict, explain and enhance decision making, which requires attention to actual decision processes.

For some time, researchers (Boudreau, 1984a; 1984b; 1987; 1988; in press; Boudreau & Berger, 1985a, 1985b; Florin-Thuma & Boudreau, 1987; Rich & Boudreau, 1987) have called for studies linking

UA models and actual managerial decision processes. More recent appeals for greater attention to better "communicating" UA results to organizational decision makers also reflect this concern (Burke & Pearlman, 1988; Rauschenberger & Schmidt, 1987). Such research must transcend simply persuading decision makers to provide more resources and status to I/O psychology and HRM programs, and exploit the full potential of UA research.

Florin-Thuma and Boudreau (1987b) assessed performance feedback utility in a small organization that had decided against implementing a performance feedback intervention. The authors asked decision makers to explicate their own decision models, and had them estimate the parameters of the normative utility model. Though only three decision makers were available, making the results exploratory, the authors found that decision makers underestimated the magnitude of the performance problem and the intervention's effect. They considered factors not included in the UA model, and these factors worked against the intervention. Yet, when dollar values were attached to these factors and when the decision makers' assumptions were incorporated into the UA model, the results still suggested substantial payoffs. Informal discussions with the decision makers indicated that they had failed to implement the performance feedback intervention because they simply had never considered the problem serious enough to warrant systematic consideration. No one believed that altering employee performance could have such a profound effect. Notably, the entire study was conducted without measuring SD_y. These results suggest research questions and methodologies to be replicated in other settings to explore how UA information affects decisions.

Future research should draw on the substantial body of knowledge regarding irrationality in decision making (e.g., Kahneman & Tversky, 1972, 1973; March & Simon, 1958). Bobko, Karren and Kerkar (1986) suggested such research directed at estimating SD_y, but the broader focus of such research is the entire process of HRM decision making. UA models offer detailed, normative theories about the factors decision makers *should* consider in making HRM decisions, but actual decisions probably depart from UA prescriptions. Etzioni (1986) has suggested that rational decision making must be induced because it is contrary to natural inclinations. For UA models to serve as one such inducement, we must first understand how actual decisions depart from UA prescriptions, and focus our efforts to induce more rational decision making. In exchange, such research will probably discover how to enhance UA models by better reflecting actual organizational decisions.

Utility analysis offers vast research potential. Moreover, the results of such research are likely to have very important implications for the ways HR managers (and those who assist them) apply findings from I-O psychology and other social sciences. With attention to the research questions noted above, it seems likely that researchers and decision makers will soon have decision tools that truly reflect a partnership between applied social science research and managerial decisions regarding human work behavior.

References

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin, 97*, 129-133.

Alexander, R. A., Barrett, G. U., & Doverspike, D. (1983). An explication of the selection ratio and its relationship to hiring rate. *Journal of Applied Psychology, 68*, 342-344.

Alexander, R. A., & Barrick, M. R. (1987). Estimating the standard error of projected dollar gains in utility analysis. *Journal of Applied Psychology, 72*, 475-479.

Alf, E. F. & Dorfman, D. D. (1967). The classification of individuals into two criterion groups on the basis of a discontinuous payoff function. *Psychometrika, 32*, 115-123.

American Management Association (1986). *Hiring strategies and costs: The AMA report.* New York: American Management Association.

Arnold, J. D., Rauschenberger, J. M. Soubel, W. & Guion, R. M. (1983). Validation and utility of a strength test for selecting steel workers. *Journal of Applied Psychology, 67*, 588-604.

Becker, B. (1988). *The influence of labor markets on human resources 'utility' estimates.* Unpublished manuscript.

Becker, G. (1975). *Human capital*, 2nd ed. New York: National Bureau of Economic Research.

Bierman, H. Jr. and Smidt, S. (1975). *The capital budgeting decision.* New York: MacMIllan.

Bierman, H. Jr., Bonini, C. P., & Hausman, W. H. (1981). *Quantitative analysis for business decisions.* Homewood, Ill: Irwin.

Bishop, J. (1987). The recognition and reward of employee performance. *Journal of Labor Economics, 5*.

Bishop, J. (1988). *The economics of employee testing.* Working paper #88-14, Center for Advanced Human Resource Studies, Cornell University, Ithaca, New York.

Bobko, P., Karren, R. & Kerkar, S. P. (1987). Systematic research needs for understanding supervisory-based estimates of SD_y in utility analysis. *Organizational Behavior and Human Decision Processes, 40*, 69-95.

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.

Bolda, R. A. (1985a). Utility: A productivity planning tool. *Human Resource Planning, 8*, (3) 111-132.

Bolda, R. A. (1985b). Individual productivity: A sourcing analysis. Unpublished manuscript.

Boudreau , J. W. (1983a). Economic considerations in estimating the utility of human resource productivity improvement programs. *Personnel Psychology, 36*, 551-557.

Boudreau , J. W. (1983b). Effects of employee flows on utility analysis of human resource productivity improvement programs. *Journal of Applied Psychology, 68*, 396-407.

Boudreau , J. W. (1984a). Decision theory contributions to HRM research and practice. *Industrial Relations, 23*, 198-217.

Boudreau , J. W. (1984b). *EXTMOV: A spreadsheet program for quantifying external employee movement decisions.* Ithaca, NY: Author.

Boudreau, J. W. (1984c). *Utility analysis for productivity improvement programs affecting work group composition.* Presented at the American Psychological Association's 92nd Meeting, Toronto, August. (ERIC Document Reproduction Service temporary accession number CG017713).

Boudreau, J. W. (1986a). *MOVUTIL: A spreadsheet program for analyzing the utility of internal and external employee movement.* Ithaca, NY: Author.

Boudreau, J. W. (1986b). *Utility analysis applied to internal and external employee movement: An integrated theoretical perspective.* Ithaca, New York: Author.

Boudreau, J. W. (1987). *Utility analysis: A review and agenda for future research.* Paper presented at the International Conference on Advances in Selection and Assessment. University of Manchester Institute of Science and Technology, Buxton, U.K., May.

Boudreau, J. W. (1988). Utility analysis. In L. D. Dyer (Ed.), *Human resource management: Evolving roles & responsibilities*, 1-125--1-186. Washington, D.C.: Bureau of National Affairs.

Boudreau, J. W. (in press). Utility analysis: A review and agenda for future research. In M. Smith and I. Robertson (Eds.), *Advances in Personnel Selection & Assessment.* London: John Wiley & Sons.

Boudreau, J. W. & Berger, C. J. (1985a). Decision-theoretic utility analysis applied to external employee movement. *Journal of Applied Psychology* [Monograph], *70*, 581-612.

Boudreau, J. W. & Berger, C. J. (1985b). Toward a model of employee movement utility. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resource management* (pp. 31-53). Greenwich, Conn.: JAI Press.

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

Brogden, H. E. (1946a). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology, 37*, 65-76.

Brogden, H. E. (1946b). An approach to the problem of differential prediction. *Psychometrika, 14*, 169-182.

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

Brogden, H. E. & Taylor, E. K. (1950). The dollar criterion --Applying the cost accounting concept to criterion construction. *Personnel Psychology, 3*, 133-154.

Burke, M. J. (1985). An investigation of dimensions employed and percentile ordering effects in estimating performance standard deviations in dollars for clerical occupations. Unpublished manuscript.

Burke, M. J. & Frederick, J. T. (1984). Two modified procedures for estimating standard deviations in utility analyses. *Journal of Applied Psychology, 69*, 482-489.

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

Campbell, J. P. (1982). Editorial: some remarks from the outgoing editor. *Journal of Applied Psychology, 67*, 691-700.

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

Cascio, W. F. (1982). *Costing human resources: the financial impact of behavior in organizations.* Boston, MA: Kent.

Cascio, W. F. (1987). *Costing human resources: the financial impact of behavior in organizations* (2nd ed.). Boston, MA: Kent.

Cascio, W. F. & McEvoy, G. M. (1985). Strategies for reducing employee turnover: A meta-analysis. *Journal of Applied Psychology, 70*, 342-353.

Cascio, W. F. & Ramos, R. (1986). Development and application of a new method for assessing job performance in behavioral/economic terms. *Journal of Applied Psychology, 1*, 20-28.

Cascio W. F. & Silbey, V. (1979). Utility of the assessment center as a selection device. *Journal of Applied Psychology, 64*, 107-118.

Cascio, W. F. & Zedeck, S. (1983). Open a new window in rational research planning: Adjust alpha to maximize statistical power. *Personnel Psychology, 36*, 517-526.

Cohen, J. (1962). The statistical power of abnormal-social psychological research. *Journal of Abnormal and Social Psychology, 65*, 145-153.

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

Cronshaw, S. F., Alexander, R. A., Wiesner, W. H. & Barrick M. R. (1987). Incorporating risk into selection utility: Tow models for sensitivity analysis and risk simulation. *Organizational Behavior and Human Decision Processes, 40*, 270-286.

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

Curtis, E. W. (1967). *The application of decision theory and scaling methods to a selection test evaluation.* San Diego, Calif.: U.S. Naval Personnel Research Activity (Technical Bulletin, STB 67-18). February.

Dalton, D. R., Krackhardt, D. M., & Porter, L. W. (1981). Functional turnover: An empirical assessment. *Journal of Applied Psychology, 66*, 716-721.

Darlington, R. B., & Stauffer, G. F. (1966). Use and evaluation of discrete test information in decision making. *Journal of Applied Psychology, 50*, 125-129.

Day, R. D. & Edwards, J. E. (1987). A comparative study of multiple SDy estimation procedures. Paper presented at the National Meeting of the Academy of Management, New Orleans, August.

DeSimone, R. L., Alexander, R. A. & Cronshaw, S.F. (1986). Accuracy and reliability of SDy estimates in utility analysis. *Journal of Occupational Psychology, 59*, 93-102.

Doppelt, J. E., & Bennett, G. K. (1953). Reducing the cost of training satisfactory workers by using tests. *Personnel Psychology, 6*, 1-8.

Dunnette, M. D., Rosse, R. L., Houston, J. S., Hough, L. M., Toquam, J., Lammlein, S., King, K. W., Bosshardt, M. J., & Keyes, M. (1982). *Development and validation of an industry-wide electric power plant operator selection system.* Edison Electric Institute.

Eaton, N. K., Wing, H. & Lau, A. (1985). *Utility estimation in five enlisted occupations.* Paper presented at the 1985 Military Testing Association (MTA), San Diego, CA.

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

Edwards, W. (1977). Use of multiattribute utility measurement for social decision making. In D. E. Bell, R. L. Keeney, & H. Raiffa (Eds.), *Conflicting objectives in decisions*. New York: Wiley.

Edwards, J. E., Frederick, J. T. & Burke, M. J. (1988). Efficacy of modified CREPID, SDy's on the basis of archival organizational data. *Journal of Applied Psychology, 73*, 529-535.

Einhorn, H. J. & McCoach, W. P. (1977). A simple multiattribute utility procedure for evaluation. *Behavioral Science, 22*, 270-282.

Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review, 86*, 465-485.

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.

Etzioni, A. (1986). Rationality is anti-entropic. *Journal of Economic Psychology, 7*, 17-36.

Eulberg, J. R., O'Connor, E. J., & Peters, L. H. (1985). *Estimates of the standard deviation of performance in dollars: An investigation of the influence of alternative sources of information.* Paper presented at the 45th Annual Meeting of the Academy of Management, San Diego.

Fischer, G. W. (1976). Multidimensional utility models for risky and riskless choice. *Organizational Behavior and Human Performance, 17*, 127-146.

Flamholtz, E. G. (1974). *Human resource accounting*. Encino, CA: Dickenson.

Flamholtz, E. G. (1985). *Human resource accounting* (2nd ed.). San Francisco, CA: Jossey-Bass.

Florin-Thuma, B. C. & Boudreau, J. W. (1987). Performance feedback utility in a small organization: Effects on organizational outcomes and managerial decision processes. *Personnel Psychology, 40*, 693-713.

Fowler, R. L. (1985). Testing for substantive significance in applied research by specifying nonzero effect null hypotheses. *Journal of Applied Psychology, 70*, 215-218.

Frank, R. H. (1984). Are workers paid their marginal products? *The American Economic Review, 74*, 549-571.

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association, 55*, 708-713.

Gow, J. F. (1985). Human resource managers must remember the bottom line. *Personnel Journal*, April, 30-32.

Guttman, I. & Raju, N. S. (1965). A minimum loss function as determiner of optimal cutting scores. *Personnel Psychology, 18*, 179-185.

Hertz, D. B. (1964). Risk analysis in capital investment. *Harvard Business Review*, Jan-Feb., 95-106.

Hillier, F. S. (1963). The derivation of probabilistic information for the evaluation of risky investments. *Management Science, 9*, 443-457.

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

Hogarth, R. M., & Einhorn, H. J. (1976). Optimal strategies for personnel selection when candidates can reject offers. *Journal of Business, 49*, 478-495.

Huber, G. P. (1980). *Managerial decision making*. Glenview, Ill.: Scott, Foresman and Co.

Hull, C. L. (1928). *Aptitude testing*. Yonkers, New York: World Book.

Hull, J. C. (1980). *The evaluation of risk in business investment*. Oxford: Pergammon Press.

Hunter, J. E. (1980). Test validation for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB). Washington, D.C.: U.S. Employment Service.

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

Hunter, J. E., & Schmidt, F. L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In M. D. Dunnette & E. A. Fleishman (Eds.), *Human performance and productivity* (Vol. 1, pp. 233-284). Hillsdale, NJ: Erlbaum.

Hunter, J. E. & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee job performance and work force productivity. *American Psychologist, 38*, 473-478.

Hunter, J. E., Schmidt, F. L. & Coggin, T. D. (1988). Problems and pitfalls in using capital budgeting and financial accounting techniques in assessing the utility of personnel programs. *Journal of Applied Psychology, 73*, 522-528.

Hunter, J. E., Schmidt, F. L. & Jackson, G. B. (1982). *Metaanalysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Hunter, J. E., Schmidt, F. L. & Judiesch, M. K. (1988). *Individual differences in output as a function of job complexity*. Working paper.

Jain, H. & Murray, V. (1984). Why the human resources management function fails. *California Management Review, 26*, 95-110.

Janz, J. T. & Dunnette, M. D. (1977). An approach to selection decisions: Dollars and sense. In J. R. Hackman, et al. (Eds.), *Perspectives on performance in organizations* (pp. 119-126). New York: McGraw-Hill.

Jacques, E. (1961). *Equitable payment*. New York: Wiley.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Karren, R. J., NKomo, S. & Ramirez, D. (1985). *Improving personnel selection decisions: A field survey using decision theory and utility analysis*. Paper presented at the 45th annual meeting of the Academy of Management. San Diego, CA.

Keeney, R. L. & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.

Kelley, T. L. *Statistical method*. New York: MacMillan.

Kroeck, G., Barrett, G. V. & Alexander, R. A. (1983). Imposed quotas and personnel selection: A computer simulation study. *Journal of Applied Psychology, 68*, 123-136.

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

Ledvinka, J., Simonet, J. K., Neiner, A. G., & Kruse, B. (1983). The dollar value of JEPS at Life of Georgia. Unpublished technical report.

Lee, R. & Booth, J. M. (1974). A utility analysis of a weighted application blank designed to predict turnover for clerical employees. *Journal of Applied Psychology, 59*, 516-518.

Lee, R. & Foley, P. P. (1986). Is the validity of a test constant throughout the test score range? *Journal of Applied Psychology, 71*, 641-644.

Levin H. (1983). *Cost-effectiveness: A primer*. Beverly Hills, CA: Sage Publications.

Locke, E. S., Faren, D. B., McCaleb, V. M., Shaw, K. N., & Denny, A. T. (1980). The relative effectiveness of four methods of motivating employee performance. In K. Duncan, M. Gruneberg, & D. Wallis (Eds.), *Changes in working life*. New York: Wiley.

Mahoney, T. A. & England, G. W. (1965). Efficiency and Accuracy of employee selection decision rules. *Personnel Psychology, 18*, 361-377.

March, J. G. & Simon, H. G. (1958). *Organizations*, New York: Wiley.

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

McEvoy, G. M. & Cascio, W. F. (1985). *Turnover and employee performance: A meta-analytic review*. Presented at the 45th annual meeting of the National Academy of Management. San Diego, August.

Meehl, P. E. & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194-216.

Mueser, P. & Maloney, T. (1987). *Ability, human capital and employer screening: reconciling labor market behavior with studies of employee productivity*. Unpublished Manuscript.

Milkovich, G. T. & Andersen, J. C. (1982). Career planning and development systems. In K. M. Rowland & G. R. Ferris (Eds.), *Personnel management* (pp. 364-388). Boston: Allyn and Bacon.

Milkovich, G. T. & Boudreau, J. W. (1988). *Personnel/human resource management: A diagnostic approach*. Homewood, IL: Irwin.

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

Naylor, J. C. & Shine, L. C. (1965). A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology, 3*, 33-42.

Pfeffer, J. & Cohen, Y. (1984). Determinants of internal labor markets in organizations. *Administrative Science Quarterly, 29*, 550-572.

Raju, N. S., Burke, M. J. & Normand, J. (1987). *A new model for utility analysis*. Working paper, Department of Psychology, Illinois Institute of Technology, Chicago, IL.

Rauschenberger, J. M. & Schmidt, F. L. (1987). Measuring the economic impact of human resource programs. *Journal of Business and Psychology, 2*, 50-59.

Reilly, R. R. & Smither J. W. (1985). An examination of two alternative techniques to estimate the standard deviation of job performance in dollars. *Journal of Applied Psychology, 70*, 651-661.

Rich, J. R. (1986). *Variance and risk in utility analysis: A case study*. Unpublished Master's thesis, Cornell University.

Rich, J. R. & Boudreau, J. W. (1987). The effects of variability and risk on selection utility analysis: An empirical simulation and comparison. *Personnel Psychology, 40*, 55-84.

Roche, U. F. (1961). The Cronbach-Gleser utility function in fixed treatment employee selection. Unpublished doctoral dissertation. Southern Illinois University, 1961. (portions reproduced in L.J. Cronbach & G. C. Gleser (Eds.), *Psychological tests and personnel decisions*, Urbana. Ill., University of Illinois Press, 1965, pp. 254-266.

Rosenthal, R. & Rubin, D. B. (1985). Statistical analysis: Summarizing evidence versus establishing facts. *Psychological Bulletin, 97*, 527-529.

Rynes, S. L. & Boudreau, J. W. (1986). College recruiting in large organizations: Practice, evaluation and research implications. *Personnel Psychology, 39*, 729-757.

Rynes, S. L. & Milkovich, G. T. (1986). Wage surveys: Dispelling some myths about the 'market wage'. *Personnel Psychology, 39*, 71-90.

Sackett, P. R., Zedeck, S. & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73*, 482-486.

Sands, W. A. (1973). A method for evaluating alternative recruiting-selection strategies: The CAPER model. *Journal of Applied Psychology, 57*, 222-227.

Schmidt, F. L. (1974). Probability and utility assumptions underlying use of the Strong Vocational Interest Blank. *Journal of Applied Psychology*, 1974, *59*, 456-464.

Schmidt, F. L. & Hoffman, B. (1973). Empirical comparison of three methods of assessing utility of a selection device, *Journal of Industrial and Organizational Psychology, 1973*, 13-22.

Schmidt, F. L. & Hunter, J. E. (1983). Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. *Journal of Applied Psychology, 68*, 407-414.

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

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

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., Outerbridge, A. N. & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology, 73*, 46-57.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N. & Trattner, M. H. (1986). The economic impact of job selection methods on size, productivity, and payroll costs of the federal work force: An empirically based demonstration. *Personnel Psychology, 39*, 1-29.

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

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validity studies. *Journal of Applied Psychology, 61*, 473-485.

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.

Sheppeck, M. A. & Cohen, S. L. (1985). Put a dollar value on your training programs. *Training and Development Journal*, November, 59-62.

Smith, M. (1948). Cautions concerning the use of the Taylor-Russell tables in employee selection. *Journal of Applied Psychology, 32*, 595-600.

Smith, P. C. (1976). Behaviors, results and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.) *Handbook of Industrial and Organizational Psychology*.

Steffy, B. D. & Maurer, S. D. (1988). Conceptualizing and measuring the economic effectiveness of human resource activities. *Academy of Management Review, 13*, 271-286.

Stone, C. H. & Kendall, W. E. (1956). *Effective personnel selection procedures*. Englewood Cliffs, N. J.: Prentice-Hall.

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, 23*, 565-578.

Tsui, A. S. (1984). Personnel department effectiveness: A tripartite approach. *Industrial Relations, 23*, 184-197.

Tsui, A. S. (1987). Defining the activities and effectiveness of the human resource department: A multiple-constituency approach. *Human Resource Management, 26*, 35-69.

Tsui, A. S. & Gomez-Mejia, L. R. (1988). Evaluating human resource effectiveness. In L. D. Dyer (Ed.), *Human resource management: Evolving roles & responsibilities*, 1-187--1-227. Washington, D.C.: Bureau of National Affairs.

Van Naersson, R. F. (1963). *Selectie van chauffers*. Groningen: Wolters. (portions reproduced in L.J. Cronbach & G. C. Gleser (Eds.), *Psychological tests and personnel decisions*, Urbana. Ill., University of Illinois Press, 1965

Wagle, B. (1967). A statistical analysis of risk in investment decisions. *Operational Research Quarterly*, *18*, 13-33.

Weekley, J. A. & Gier, J. A. (1986). *An examination of the temporal consistency of rational SDy estimates*. Paper presented at the annual meeting of the Academy of Management, Chicago, August.

Weekley, J. A., O'Connor, E. J., Frank, B., & Peters, L. W. (1985). A comparison of three methods of estimating the standard deviation of performance in dollars. *Journal of Applied Psychology*, *79*, 122-126.

Wesman, A. G. (1953). Better than chance. *Test Service Bulletin*, *45*, 1-67.

Wigdor, A. K. & Hartigan, J. A. (1988). *Interim Report: Within-group scoring of the General Aptitude Test Battery*. Washington, D.C.: National Academy Press.

Wroten, S. P. (1984). *Overcoming the futilities of utility applications: Measures, models and management*. Presented at the American Psychological Association's 92nd Meeting, Toronto, August.

Table 1. Example of a Multiattribute Utility Matrix for a Training Decision

	Decision Options		
Attributes	Program A	Program B	*Attribute Weights*
(a) Sales Levels (Dollars per Year)	$100,000	$130,000	1
(b) Required resources (Total Dollar Value)	$10,000	$30,000	-1
(c) Job Satisfaction (1=low, 7=high)	6.0	2.0	3,000
Total Utility Value =	108,000	106,000	

Table 2. *One-Cohort Entry-Level Selection Utility Decision*

Cost-Benefit Information	*Entry-Level Computer Programmers*
Current Employment	4,404
Number Separating	618
Number Selected (N_s)	618
Average Tenure (T)	9.69 years
Test Information	
Number of Applicants (N_{app})	1,236
Testing Cost	$10/applicant
Total Test Cost (C)	$12,360
Average Test Score ($\bar{Z}_x$)	.80 SD
Validity ($r_{x,y}$)	.76
SD_y (per person-year)	$10,413

Utility Computation

Quantity = Average Tenure X Applicants Selected
= 9.69 years X 618 applicants
= 5,988 person-years

Quality = Average Test Score X Test Validity X SD_y
= .80 X .76 X $10,413
= $6,331 per person-year

Utility = (Quantity X Quality) - Cost
= (5,988 person-years X $6,331 per person-year) - $12,360
= $37.9 million

Adapted with permission from: J. W. Boudreau (1988), Table 4.

Table 3. Results of Studies Deriving Actual Program Utility Values

Reference	Setting	N_s	*T or F*	*SR*	$\bar{Z}_s$	$r_{x,y}$	SD_y	Cost	ΔU	Utility Formula	B-E SD_y
Doppelt & Bennett (1953)[a]	Examined selection utility for predicting training success of grocery clerks.	1		.10			$308		$197		
Doppelt & Bennett (1953)[a]	Examined selection utility for predicting training success of adding machine operators.	1		.13			$214		$180		
Doppelt & Bennett (1953)[a]	Examined selection utility for predicting training success of produce workers.	1		.07			$179		$116		
Russmore & Toorenaar (1956)[a]	Examined selection utility for predicting training success of telephone operators.		1						$28,000		
Roche (1961)	Examined selection utility of a test battery for predicting radial drill operators overall performance per hour.	1	1 hour	.33	1.11	.313	$0.585	$0.0006/hr.	$0.203/hr.	Δ U = (.348 SDy)-$0.0006	$0.0017
Van Naersson (1963)	Examined selection utility of a driving experience questionnaire for reducing training time for drivers in the Dutch Army.	4,392	1 yr.	.81	.334	.66	$77.00	$1,627	$73,049	Δ U = (970 SDy)-$1,627	$1.68
Schmidt & Hoffman (1973)	Examined weighted application blank selection utility for reducing separations among nurse's aides.	308	2 yrs.	.302	NR	.47	$1,652	$ 628 per separ.	$ 161,243	NA	NA
Lee & Booth (1974)[a]	Examined the selection utility of a weighted application blank for predicting turnover among clerical employees.	245	25 mo.	.17	1.47	.56	$1,238	$0	$249,900	Δ U = (202 SDy)-$0	$0

Table 3. Results of Studies Deriving Actual Program Utility Values (continued)

Reference	Setting	N_s	*T or F*	*SR*	$\bar{Z}_s$	$r_{x,y}$	SD_y	Cost	ΔU	Utility Formula	B-E SD_y
Cascio & Silbey (1979)	Examined assessment center selection utility for food and beverage sales managers.	50	5 yrs.	.50	.80	.35	$ 9,500	$73,928	$504,211	ΔU = (61.04 SDy)-$40,328	$660.70
Cascio & Silbey (1979)	Examined interview selection utility for food and beverage sales managers.	50	5 yrs.	.50	.80	.25	$ 9,500	$62,600	$350,357	Δ U = (43.59 SDy)-$29,000	$665.29
Cascio & Silbey (1979)	Examined assessment center selection utility minus interview selection utility for food and beverage sales managers.	50	5 yrs.	.50	.80	.10	$ 9,500	$11,328	$350,357	Δ U = (17.45 SDy)-$11,328	$649.16
Schmidt, et al. (1979)	Examined interview selection utility selection utility for U.S. Govt. computer programmers.	618	10 yrs.	.50	.80	.14[b]	$10,413	$358,440[c]	$ 6,849,022	Δ U = (692 SDy)-$358,440	$517.85
Schmidt, et al. (1979)	Examined PAT selection utility for U.S. Govt. computer programmers.	618	10 yrs.	.50	.80	.76	$10,413	$370,800[d]	$38,755,422	Δ U = (3,757 SDy)-$370,800	$98.70
Schmidt, et al. (1979)	Examined PAT selection utility minus interview selection utility for U.S. Govt. computer programmers.	618	10 yrs.	.50	.80	.76	$10,413	$ 12,360	$31,906,400	Δ U = (3,065 SDy)-$12,360	$ 4.03
Arnold, et al. (1982)	Examined selection utility of a strength test to select steelworkers.	1,853	1 yr.	.06	1.97	.84	$3,000	$0	$9,199,033	Δ U = (3,066 SDy)-$0	$0
Dunnette, et al. (1982)	Examined selection utility of a test battery to select hydroelectric power plant operators.	1	1 yr.	.50	.80	.28	$15,600[e]	$100	$ 3,295	Δ U = (.216 SDy) - $100	$46.30
Dunnette, et al. (1982)	Examined selection utility of a test battery to select fossil power plant operators.	1	1 yr.	.50	.80	.44	$21,400[e]	$100	$ 7,335	Δ U = (.347 SDy) - $100	$288.18

Table 3. Results of Studies Deriving Actual Program Utility Values (continued)

Reference	Setting	N_s	T or F	SR	$\bar{Z}_s$	r_{xy}	SD_y	Cost	ΔU	Utility Formula	B-E SD_y
Dunnette, et al. (1982)	Examined selection utility of a test battery to select fossil power plant control room operators (CRO).	1	1 yr.	.50	.80	.44	$72,400[e]	$100	$25,285	ΔU = (.351 SDy) - $100	$284.90
Dunnette, et al. (1982)	Examined selection utility of a test battery to select nuclear power plant operators.	1	1 yr.	.50	.80	.30	$23,500[e]	$100	$ 5,440	ΔU = (.236 SDy) - $100	$424
Dunnette, et al. (1982)	Examined selection utility of a test battery to select nuclear power plant control room operators.	1	1 yr.	.50	.80	.30	$134,800[e]	$100	$32,150	ΔU = (.239 SDy) - $100	$418
Ledvinka, et al. (1983)	Examined selection utility of the JEPS test for life insurance claim approvers.	10	1 yr.	.07	1.918	.36	$5,542	$1,104	$37,162	ΔU = (6.90 SDy)-$1,104	$160
Ledvinka, et al. (1983)	Examined selection utility of the interview for life insurance claim approvers.	10	1 yr.	.07	1.918	.14	$5,542	$0	$14,881	ΔU = (2.68 SDy)-$0	$0
Ledvinka, et al. (1983)	Examined selection utility of the JEPS test minus the selection utility of the interview for life insurance claim approvers.	10	1 yr.	.07	1.918	.22	$5,542	$1,104	$22,281	ΔU = (4.22 SDy)-$1,104	$262
Schmidt, Mack & Hunter (1984)	Examined selection utility of using the interview to select U.S. Park Rangers.	80	10 yrs.	.10	1.758	.14	$4,451	$232,000[e]	$ 644,384	ΔU = (197 SDy)-$232,000	$1,178
Schmidt, Mack & Hunter (1984)	Examined selection utility of using the PACE test to select U.S. Park Rangers.	80	10 yrs.	.10	1.758	.51	$4,451	$232,000[f]	$2,960,542	ΔU = (717 SDy)-$232,000	$323.57
Schmidt, Mack & Hunter (1984)	Examined selection utility of the PACE test minus the selection utility of the interview to select U.S. Park Rangers.	80	10 yrs.	.10	1.758	.37	$4,451	$0	$2,316,482	ΔU = (520 SDy)-$0	$0

Table 3. Results of Studies Deriving Actual Program Utility Values (continued)

Reference	Setting	N_s	T or F	SR	$\bar{Z}_x$	$r_{x,y}$	SD_y	Cost	ΔU	Utility Formula	B-E SD_y
Wroten (1984)	Calculated selection utility for using various selection tests compared to random selection for Head Operators.	1	10 yrs.	.15	1.55	.30	$29,472[a]	$ 1,000	$ 136,045	ΔU = (4.65 SDy)-$1,000	$216.00
Wroten (1984)	Calculated selection utility for using various selection tests compared to random selection for Outside Operators.	1	5 yrs.	.15	1.55	.30	$21,591[a]	$ 1,000	$ 49,199	ΔU = (2.33 SDy)-$1,000	$429.18
Wroten (1984)	Calculated selection utility for using various selection tests compared to random selection for Pump Operators.	1	3 yrs.	.15	1.55	.30	$14,205[a]	$ 1,000	$ 18,816	ΔU = (1.40 SDy)-$1,000	$714.28
Wroten (1984)	Calculated selection utility for using various selection tests compared to random selection for Instrument Technicians.	1	15 yrs.	.15	1.55	.30	$46,396[a]	$ 1,000	$ 322,612	ΔU = (6.97 SDy)-$1,000	$143.47
Wroten (1984)	Calculated selection utility for using various selection tests compared to random selection for Outside Mechanics.	1	8 yrs.	.15	1.55	.30	$26,438[a]	$ 1,000	$ 97,349	ΔU = (3.72 SDy)-$1,000	$268.82
Wroten (1984)	Calculated selection utility for using various selection tests compared to random selection for Welders.	1	17 yrs.	.15	1.55	.30	$18,291[a]	$ 1,000	$ 143,590	ΔU = (7.90 SDy)-$1,000	$126.58
Weekley, et al. (1985)	Examined selection utility of a test battery for convenience store managers.	1,000	1 yr.	.33	1.102	.45	$ 7,701[h]	$30,000	$3,788,926	ΔU = (495.9 SDy)-$30,000	$60.49

Table 3. Results of Studies Deriving Actual Program Utility Values (continued)

Reference	Setting	N_s	*T or F*	*SR*	$\bar{Z}_x$	$r_{x,y}$	SD_y	Cost	ΔU	Utility Formula	B-E SD_y
Burke & Frederick (1986)[i]	Examined selection utility of an assessment center to select mid-level sales managers. TAX=.49, i=.18, V=-.048.	29	4.2 yrs.	.22	.872	.59	$12,789[j]	$134,454[k]	$115,727	ΔU = (19.56 SDy)-$134,454	$6,874
Burke & Frederick (1986)[i]	Examined selection utility of an interview to select mid-level sales managers. TAX=.49, i=.18, V=-.048.	29	4.2 yrs.	.22	.872	.16	$12,789[j]	$ 25,747[k]	$ 42,098	ΔU = (5.30 SDy)-$25,747	$4,858
Burke & Frederick (1986)[i]	Examined selection utility of an assessment center minus interview selection utility to select mid-level sales managers. TAX=.49, i=.18, V=-.048.	29	4.2 yrs.	.22	.872	.43	$12,789[j]	$108,707[k]	$ 73,629	ΔU = (14.26 SDy)-$108,707	$7,625
Cascio & Ramos (1986)	Examined selection utility of an assessment center for promoting telephone company office managers.	1,116	4.4 yrs.	.32	1.17	.388	$10,421	$2,399,400	$20,830,312	ΔU=(2,229 SDy)-$2,399,400	$1,076
Cascio & Ramos (1986)	Examined selection utility of an interview for promoting telephone company office managers.	1,116	4.4 yrs.	.32	1.17	.14	$10,421	$1,046,250	$ 7,335,605	ΔU =(804 SDy)-$1,046,250	$1,301
Cascio & Ramos (1986)	Examined selection utility of an assessment center versus interview for promoting telephone company office managers.	1,116	4.4 yrs.	.32	1.17	.248	$10,421	$1,353,150	$13,494,707	ΔU=(1,425 SDy)-$1,353,150	$ 950
Cronshaw, et al. (1986)	Examined selection utility of a cognitive ability test to select one group of clerical/administrative employees for the Canadian military. i=.1125	470	18 yrs.	.333	1.09	.52	$10,680[j]	$159,900	$21,738,270	ΔU = (2,020 SDy)-$159,900	$79.16

Table 3. Results of Studies Deriving Actual Program Utility Values (continued)

Reference	Setting	*N*	*T or F*	*SR*	$\bar{Z}$	$r_{x,y}$	SD_y	Cost	Δ U	Utility Formula	B-E SD_y
Schmidt, et al. (1986)	Examined the selection utility of selecting one cohort of U.S. Government employees with a test of cognitive abilities versus non-test methods. (d=.487)	225,731	13 yrs.	.15[l]	1.55[l]	.31[l]	$5,429	Assume Zero	$7.87 Billion[m]	Δ U = (1.43 Million X SDy)	$0
Florin-Thuma & Boudreau (1987)	Examined the one-year utility of performance feedback for yogurt shop counter workers.	15	1 yr.	utility of production units (no SDy)				$409.75	$11,719	----------	----
Mathieu & Leonard (1987)[n]	Examined training program utility for Head Tellers in the Bank of Virginia. Utility calculated based on training 10 Head Tellers, who remain a maximum of 18 yrs. TAX = .46, i=.15, V=-.0668.	10	18 yrs.	Calculated d of .3146			$ 2,369	$ 3,089[o]	$ 9,790	Δ U = (5.44 SDy)-$3,089	$ 568
Mathieu & Leonard (1987)[n]	Examined training program utility for Operations Mgrs. in the Bank of Virginia. Utility calculated based on training 10 Operations Mgrs. who remain a maximum of 20 yrs. TAX = .46, i=.15, V=-.0729.	19	20 yrs.	Calculated d of .3146			$ 3,123	$5,830[o]	$ 28,694	Δ U = (11.05 SDy)-$5,830	$527
Mathieu & Leonard (1987)[n]	Examined training program utility for Branch Managers in the Bank of Virginia. Utility calculated based on training 36 Branch Managers who remain a maximum of 19 yrs. TAX = .46, i=.15, V=-.0287.	36	19 yrs.	Calculated d of .3146			$10,064	$14,814[o]	$ 156,400	Δ U = (17.01 SDy)-$14,814	$871

Table 3. Results of Studies Deriving Actual Program Utility Values (continued)

Reference	Setting	N_s	*T or F*	*SR*	$\bar{Z}_x$	$r_{x,y}$	SD_y	Cost	Δ U	Utility Formula	B-E SD_y
Rich & Boudreau (1987)[p]	Examined the selection utility of the PAT for computer programmers. TAX=.39, i=.15, V=0.0.	flows	11 yrs.	.398	.73	.73	$15,888	$229,101[k]	$3,198,258	ΔU = (216 SDy)-$229,101	$1,062
Rich & Boudreau (1987)[p]	Examined the selection utility of the interview for computer programmers. TAX=.39, i=.15, V=0.0.	flows	11 yrs.	.398	.73	.14	$15,888	$225,543[k]	$ 431,744	ΔU = (41.4 SDy)-$225,543	$5,452
Rich & Boudreau (1987)[p]	Examined the selection utility of the PAT minus the utility of the interview for computer programmers. TAX=.39, i=.15, V=0.0.	flows	11 yrs.	.398	.73	.59	$15,888	$3,557[k]	$2,775,889	ΔU = (174 SDy)-$3,557	$20.44

[a] As reported in Hunter & Schmidt (1982).

[b] Interview validity based on subsequent estimates by these authors (e.g., Schmidt, et al., 1984).

[c] Cost per interview estimated as $290.00, following Cascio & Silbey (1979); Authors provided no cost information.

[d] Cost per test estimated as $300.00, following the author's statement that the cost of testing would be an additional $10 per applicant.

[e] Conservatively estimated at the bottom of the 95% confidence interval.

[f] Cost per test estimated as $290.00, based on the authors' statement that "it is unlikely to be greater than the cost per interview" (p. 493).

[g] SDy estimates based on the average of the Schmidt, et al. (1979) estimation method and the realistically anchored estimation method.

[h] SDy estimates based on CREPID estimation method (40% of salary produced $8,489; Schmidt, et al. produced $13,968).

[i] Adjusted for financial/economic factors.

[j] SDy estimate based on 40% of salary.

[k] After-tax cost.

[l] Derived from previous studies. Reported utility value is based on the empirically observed d_t value of .487.

[m] This value differs slightly from the reported utility value ($7.84 billion) which was based on an average across positions.

[n] Adjusted for financial/economic considerations and employee turnover over time.

[o] After-tax, discounted cost, includes $12,800 fixed program development cost allocated according to the number trained in each job.

[p] Adjusted for financial/economic factors as well as employee flows over time.

Table 4. Results of Studies Estimating SD_y Values

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Doppelt & Bennett (1953)[a]	Estimated savings in training costs.			Grocery Clerks: Mean = $ 308/yr SE = SD = % salary = 15% % mean y = NR Adding Machine Opers.: Mean = $214/year SE = SD = % salary = 10% % mean y = NR Produce workers: Mean = $179 SE = SD = % salary = 10% % mean y = NR
Roche (1961) Reported in Cronbach & Gleser (1965)	Estimated the value of production units of radial drill operators (N=291) in a manufacturing organization, by attaching value to each unit.	"The dollar profit which accrues to the company as a result of an individual's work."	Used cost accounting to attach "standard" costs (materials, labor and facility usage) and prices to units. Price less cost was value per unit, and SD_y was based on individual output quantities.	Mean = $1,217 ($.585/hour) SE = NA SD = NA % salary = 25%[a] % mean y = NR
Van Naersson (1963) Reported in Cronbach & Gleser (1965)	Used training time data to estimate the SD of training time costs across military driver trainees in the Dutch Army.	Reduction in training time costs.	Used training time data to estimate SD training hours (12.9), and then multiplied by average training cost/hour ($6.00) to produce SD of cost equal to $77.00.	Mean = $77.00 SE = NA SD = NA % salary = 4% % mean y = NR

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values	
Schmidt & Hoffman (1973)	Estimated savings from reducing turnover costs (hiring, training, administration, overhead)	Reduction in recruitment, hiring, and training costs due to longer tenure and fewer replacements. among nurse's	No SD_y calculated, but an SD_y estimate can be derived by working backward from the known $\overline{Z}_x$ $r_{x,y}$, and ΔU/selectee. aides in a hospital.	Mean SE SD % salary % mean y	= $624/year = NA = NA = 15% = NR
Lee & Booth (1974)*	Estimated the clerical employees' tenure, and calculated the cost savings due to reduced recruiting, hiring and training costs.	Reduction in recruitment, hiring, and training costs due to longer tenure and fewer replacements.	No SD_y measurement per se, but an SD_y measure can be derived working backward from the known $\overline{Z}_x$, $r_{x,y}$, and ΔU/selectee.	Mean SE SD % salary % mean y	= $1,238 = NA = NA = 20% = NR
Cascio & Silbey (1979)	Third-level food and beverage sales managers (N = 4) estimated yearly sales levels for incumbent second-level sales managers.	Yearly value of sales.	Global estimation of 15th, 50th, and 95th percentiles. Average difference between two endpoints equals SD_y estimate.	Mean SE SD % salary	= $9,500 = NA = NA = 42.2%
Schmidt, et al. (1979)	Supervisors (N =105) estimated "yearly value of products and services" and "cost of having an outside firm provide these products" for incumbent Federal Government (GS 9-11) computer programmers.	Value of products and services and/or cost of having an outside contractor provide them.	Global estimation of 15th, 50th, and 95th percentiles. Average difference between two endpoints equals SD_y estimate.	Mean SE SD % salary % mean y	= $10,413 = $ 1,354 = $13,874 = 55% = 31.6%
Arnold, et al. (1982)	Examined the amount shovelled by steel-workers.	The amount of work accomplished.	Attempted to provide a highly conservative SD_y estimate. (1) Evidence indicated that top 10% did 8-9 times as much as the bottom 10%, so the ratio of the top 1% to the bottom 1% was estimated to be 2 to 1. (2) a conservative estimate of yearly salary plus benefits was $18,000 per year. (3) Deduced a saving of $18,000 from hiring one top 1% worker instead of a bottom 1% worker. (4) Assumed top 1% and bottom 1% were 6 SD's apart.	Mean % salary	= $3,000 = 17%

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Dunnette, et al. (1982)	Power industry experts participated in a workshop in which they discussed critical incidents of effective and ineffective power plant operator performance, were presented with 8 previously-derived performance dimensions, retranslated 667 performance examples into the dimensions, and then made dollar contribution judgments for five jobs: (1) Hydroelectric plant operator (N=31); (2) Fossil fuel plant operator (N=20); (3) Fossil fuel plant control room operator (CRO, N=48); (4) Nuclear plant operator (N=19); and (5) Nuclear Plant CRO (N=34).	Dollar value contribution of operator performance.	Schmidt, et. al. method	Hydro Operator Mean = $20,790 SD = $14,110 SE = $ 2,530 Fossil Operator Mean = $30,350 SD = $19,120 SE = $ 4,280 Fossil CRO Mean = $88,720 SD = $56,210 SE = $ 8,110 Nuclear Operator Mean = $74,900 SD = $107,150 SE = $24,580 Nuclear CRO Mean = $213,730 SD = $226,600 SE = $ 38,860
Hunter & Schmidt (1982)	Supervisors (N = 62) of budget analysts estimated similar values to Schmidt, et al. (1979).	Value of products and services and/or cost of having an outside contractor provide them.	Global estimation of 15th, 50th, and 95th percentiles. Average difference between two endpoints equals SD_y estimate.	Mean = $11,327 SE = $ 1,120 SD = $ 8,818 % salary = 60% % mean y = NR

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Bobko, Karren & Parkington (1983)	Sales counselor supervisors (N=17) estimated SD_y for counselor sales and performance levels. Also, performance data was obtained for 92 actual insurance counselors.	"Total Yearly Dollar Sales" versus "Yearly value to the company of the overall products and services produced (considering the cost of having an outside contractor provide them).	Global estimation of the 15th 50th, 85th and 97th percentiles of both sales and value. Also gathered empirical sales data, computed by taking the number of policies sold and multiplying by the average policy value in his/her area.	Sales, 4 dist. points Mean = $47,967 SE = $ 9,969 SD = $34,533 % salary = 352% % mean y = 50% Value, 4 dist. points Mean = $ 4,967 SE = $ 2,089 SD = $ 7,533 % salary = 37% % mean y = 31% Sales, 3 dist points Mean = $56,950 SE = $15,365 SD = $55,400 % salary = 419% % mean y = 59% Value, 3 dist. points Mean = $ 5,550 SE = $ 2,413 SD = $ 8,700 % salary = 41% % mean y = 35% Actual Sales Data Mean = $124,882 SD = $ 52,308 % salary = 384% % mean y = 42%

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Ledvinka, Simonet, Neiner & Kruse (1983)	Claims processed per day for 15 insurance claims approvers were recorded for 2 months.	Dollar value to the company of claims processed per year.	Average dollar value of a processed claim was estimated by dividing total payroll plus benefits per year by the average number of processed claims per year. Assumption was that the wages and benefits paid to the average employee equals his/her value to the organization. Then, the standard deviation of claims processed per year (1679.29, as corrected for range restriction) times the average value/claim ($3.30) became the SD_y estimate.	Mean = $5,542 SD = NA SE = NA % salary = 42.6% % mean y = 31.4%
Burke & Frederick (1984)	Regional manufacturing sales managers (N=26) provided global ratings for SD_y regarding their subordinate district sales managers (N=69). They estimated four percentiles (15th, 50th, 85th and 97th).	Used actual yearly sales revenue, plus estimates of "total yearly value of services," considering the cost of "having an outside firm provide these services." (p. 484).	Yearly sales volume for the 69 sales managers in 1982 was one estimate of SD_y. 1982 annual salaries for these managers was another distribution. Three distributions were derived from estimates of the four percentiles. Standard procedure simply gathered individual estimates for each percentile and calculated the mean differences between them. Procedure A fed back the 50th percentile to 4 managers and had them reach consensus on the other three. Procedure B fed back the 50th percentile estimate to 18 managers and had them re-estimate the other three percentiles.	Standard Mean = $32,284 SE = $ 9,199 SD = $45,996 % salary = 105% % mean y = 32% Procedure A Mean = $38,333 SE = NA SD = NA % salary = 124% % mean y = 51% Procedure B Mean = $32,323 SE = 1,797 SD = $ 8,983 % salary = 104% % mean y = 43% Actual Sales Revenue Mean = $6.02 M SD = $2.634 M Actual Salary Mean = $30,900 SD = $ 4,600

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Schmidt, Mack, & Hunter (1984)	Park ranger supervisors (N=114) provided data for S, et al. SD_y estimates, by considering the park rangers they supervised. Only two supervisors could not estimate the 15th percentile.	Not reported, but assumed to follow the S, et al. method of "value of goods and services." Authors note subjects were asked to "consider what the cost would be of having an outside contracting firm provide the products or services to them." (p. 492).	Standard S, et al. method of surveying respondents.	85th -50th Mean = $3,801 SE = $ 239 SD = $2,546 % salary = 36% % mean y = 28% 50th-15th Mean = $5,101 SE = $ 357 SD = $3,813 % salary = 49% % mean y = 38%
Wroten (1984)	Groups of supervisors (N=3 to 4) of petroleum workers in 7 different organizations and 16 different locations provided SD_y estimates for six refinery jobs using six different methods.	Not reported, but probably similar to Schmidt, et al. (1979) because all measures were described as variants of this method.	Twelve estimation methods were used, for each of six jobs. The six jobs were: (1) Head Operator, (2) Outside Operator, (3) Pump Operator, (4) Instrument Technician, (5) Outside Mechanic, (6) Welder. The twelve estimation methods were of four types: (1) Direct Estimates (including the Schmidt, et al. method, obtaining y estimates for individuals and calculating SD_y from them, and obtaining percentile estimates by group consensus); (2) Actual Anchored Estimates (which replicated the first three methods, but provided accurate 50th percentile estimates first); (3) High Anchored Estimates (which replicated the first three methods, but provided a high 50th percentile estimate first); and (4) Low Anchored Estimates (which replicated the first three methods, but provided a low 50th percentile estimate first). Accurate anchor was derived from "cost accounting" and unanchored group's 50th percentile estimate. High anchor was twice actual, and low anchor was half of actual.	Direct, Head Operator Mean = $31,423 SD = $26,663 SE = $ 6,383 Direct, Outside Oper. Mean = $20,468 SD = $16,041 SE = $ 3,638 Direct, Pump Operator Mean = $13,950 SD = $ 9,532 SE = $ 2,124 Direct, Inst. Tech. Mean = $35,037 SD = $25,004 SE = $ 6,266 Direct, Outside Mech. Mean = $25,297 SD = $19,310 SE = $ 4,776 Direct, Welder Mean = $19,708 SD = $13,430 SE = $ 3,235

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Wroten (1984)				**Actual, Head Operator** Mean = $27,521 SD = $17,795 SE = $ 8,227 **Actual, Outside Oper.** Mean = $22,714 SD = $13,877 SE = $ 3,942 **Actual, Pump Operator** Mean = $14,461 SD = $ 8,969 SE = $ 2,516 **Actual, Inst. Tech.** Mean = $57,754 SD = $50,419 SE = $20,046 **Actual, Outside Mech.** Mean = $27,579 SD = $12,202 SE = $ 3,884 **Actual, Welder** Mean = $16,874 SD = $ 8,306 SE = $ 3,592 **High, Head Operator** Mean = $50,714 SD = $20,835 SE = $ 5,241 **High, Outside Oper.** Mean = $38,626 SD = $17,995 SE = $ 4,672 **High, Pump Operator** Mean = $27,368 SD = $14,174 SE = $ 3,848

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Wroten (1984) Continued				**High, Inst. Tech.** Mean = $96,554 SD = $48,360 SE = $15,248 **High, Outside Mech.** Mean = $79,789 SD = $36,766 SE = $24,344 **High, Welder** Mean = $60,356 SD = $29,735 SE = $11,432 **Low, Head Operator** Mean = $27,294 SD = $27,163 SE = $ 9,617 **Low, Outside Oper.** Mean = $20,571 SD = $18,353 SE = $ 6,263 **Low, Pump Operator** Mean = $10,358 SD = $ 8,868 SE = $ 2,863 **Low, Inst. Tech.** Mean = $17,501 SD = $10,307 SE = $ 3,574 **Low, Outside Mech.** Mean = $12,752 SD = $ 7,287 SE = $ 2,514 **Low, Welder** Mean = $10,718 SD = $ 5,761 SE = $ 792

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Bolda (1985)	Estimated the job performance value for employees in maintenance and toolroom jobs in a manufacturing operation.	The dollars-per-hour value of the employee's performance	Gathered estimates of the 15th, 50th and 84th percentiles, and used the average difference as SD_y.	Mean = $6.00/hr. SE = NR SD = NR % salary = 46% % mean *y* = 46%
Burke (1985)	Supervisors of clerical workers made global y ratings for one of the three job classes they supervised. 132 gave estimates of the 50th percentile (mean = $22,045). This mean was fed back to two groups.	Used the Schmidt, et al. (1979) function of the "total yearly value of services" considering how performance contributes to the "sales value of products sold."	After making the global estimate, the 50th percentile was fed back to two groups of supervisors. The first group (N=50) estimated the 15th before the 85th percentile, while the second group (N=41) did the opposite. They also provided self-reported dimensions used in the estimates. Sixteen of the original 118 surveys produced inconsistent estimates. Dimensions used tended to follow job eval.	Mean = $ 5,529 SE = $ 400 SD = $ 3,800 % salary = NR % mean *y* = 25%
Eaton, Wing Lau (1985)	Supervisors of soldiers in 5 military occupations (MOS) provided data estimating the value of first-term soldiers operating at different performance levels. The 5 MOS's were identified as: Infantryman (11B), Armor Crewman (19E), Vehicle Mech. (63B), Medical spec. (91B), Radio oper. (05C). Total number of supervisors was 270. Computed equivalent civilian salary levels to be approximately $16,000/yr.	Used Questionnaires similar to Eaton, et al. (1985a,b). The payoff scale was the "worth to the Army" of the soldiers, considering "such factors as salary, output, responsibility and equipment" (p. 4).	Used the GLOBAL technique of S, et al. (1979), the Superior Equivalents Technique (EQV), and examined the "40-70% Rule." Only the 85th and the 50th percentiles were estimated.	11B, EQV Mean = $12,881 SD = NR SE = NR % salary = 81% % mean *y* = 81% 11B, GLOBAL Mean = $ 9,774 SD = NR SE = NR % salary = 61% % mean *y* = 51% 19E EQV Mean = $13,630 SD = NR SE = NR % salary = 84% % mean *y* = 84%

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Eaton, Wing Lau (1985) Continued				19E GLOBAL Mean = $6,254 SD = NR SE = NR % salary = 39% % mean y = 45% 91B EQV Mean = $16,720 % salary = 105% % mean y = 105% 91B GLOBAL Mean = $ 9,132 % salary = 57% % mean y = 51% 63B EQV Mean = $15,068 % salary = 94% % mean y = 94% 63B GLOBAL Mean = $10,625 % salary = 66% % mean y = 68% 05C EQV Mean = $16,653 % salary = 104% % mean y = 104% 05C GLOBAL Mean = $11,150 % salary = 70% % mean y = 61%

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Eaton, Wing & Mitchell (1985)	Trainers/Supervisors of U.S. Army Tank commanders (N=40 and 48) estimated the "value" of different levels of TC performance. Dollar-valued anchors were the average yearly compensation of TCs ($30,000), a subjective estimate of average value, and the average tank cost.	The Superior Equivalents Technique (EQV) derives the number of 85th and 15th percentile performers it takes to equal 17 average performers. The System Effectiveness Technique (EFF) derives the ratio of SD_y to standard y (where y is expressed in natural production units per tank), indicating how much of a standard unit (tank) can be "saved" for each SD productivity increase.	The difference between the median supervisor 85th and 50th percentile dollar-valued performance global subjective estimate was called SD$. Average compensation and the 50th percentile global estimate formed the anchor for the EQV technique. The value for the EFF estimate of the ratio of SD_y to standard y (i.e., .2) was from previous validation studies. The cost of a tank was estimated at $300,000 per year.	Global (SD$) method Mean = $40,000 SD = $235,000 (3rd - 1st quartile) SE = NR % salary = 133% % mean y = 123% EQV, salary anchor Mean = $26,700 SD = $21,857 (3rd - 1st quartile) SE = NR % salary = 89% % mean y = 89% EQV, global anchor Mean = $28,900 SD = $23,678 (3rd - 1st quartile) SE = NR % salary = 96% % mean y = 89% EFF Mean = $60,000 SD = NR SE = NR % salary = 200% % mean y = 185%
Eulberg, O'Connor, & Peters (1985)	Air Force supervisors (N=69) and job incumbents (N=113) from the medical technician speciality provided ratings necessary for CREPID estimates. These ratings were then combined with actual performance ratings and average salary for 95 technicians to derive a CREPID SD_y estimate, which was compared to 40% of salary.	Not reported, but the authors note that in this job, pay is "strictly a function of rank and years of service" (p. 7) which makes it unlikely that average pay is equal to average service value or net benefits.	CREPID estimates were derived using task importance ratings from either supervisors or incumbents, combined with the same performance ratings and average salary levels. 40% estimates were derived by multiplying average salary by .40.	Supervisors Mean = $ 3,336 % salary = 37% % mean y = NR Incumbents Mean = $ 3,307 % salary = 37% % mean y = NR 40% rule Mean = $ 3,581

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Mitchell, Eaton, & Wing (1985)	Collected ratings from cannon crewmen (N=35)	(1) "Yearly value to the Army of an average entry- and motor transport operators (N=26) job incumbents in the U.S. Army.	See Payoff Scale section	GLOBAL, Crewman Mean = $ 6,000 SD = $10,000 (3rd - 1st quartile) SE = NR % salary = NR % mean y = 40% FEEDBACK, Crewman Mean = $ 4,000 SD = $0 (3rd-1st quartile) SE = NR % salary = NR % mean y = 25% EQV, Crewman Mean = $ 9,600 SD = NR SE = NR % salary = NR % mean y = NR GLOBAL, Transport Mean = $ 7,000 SD = $7,000 (3rd-1st quartile) SE = NR % salary = NR % mean y = 47% FEEDBACK, Transport Mean = $ 6,000 SD = $12,000 (3rd-1st quartile) SE = NR % salary = NR % mean y = 40% EQV, Transport Mean = $ 10,000
		level soldier in their field." (also an 85th percentile performer) GLOBAL. (2) "Number of superior performers needed to obtain the output of 10 average producers working for an equivalent amount of time" (EQV). (3) "re-estimate dollar values for average and superior performers in light of dollar values cited for soldiers in other specialties. (FEEDBACK)		

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Reilly & Smither (1985)	Graduate students (N=16) with prior management experience played a computerized management simulation. They were provided sales data on 10 representatives, based on 3 job components.	Job performance was measured through 3 job components: (1) selling estab. products, (2) selling new products, (3) expense control. The first and third were reported in dollars, and the second could be computed in dollars using a formula. In addition, information on variable cost levels was provided.	Used CREPID to obtain ratings of performance dimensions that could be compared to actual performance. Used the Schmidt, et al. method to obtain SD_y estimates of estab. prod. sales, new prod. sales, net sales less expenses, and value of "overall products and services produced."	Sim. Repeat Sales Mean = $1,093,641 SD = $ 170,119 Simulated New Sales Mean = $156,225 SD = $ 24,302 Simulated Net Revenue Mean = $ 175,600 SD = $ 43,639 Schmidt,et al. repeat sales Mean = $178,725 SE = $ 13,651 SD = $ 54,604 % salary = 357% Schmidt, et al. new sales Mean = $ 29,477 SE = $ 3,374 SD = $ 13,496 % salary = 60% Schmidt, et al. net revenue Mean = $ 119,605 SE = $ 57,773 SD = $ 231,092 % salary = 242% Schmidt, et al.overall worth Mean = $ 83,994 SE = $ 25,247 SD = $ 100,988 % salary = 170% CREPID $ performance Mean = $ 26,485 SE = $ 1,381 SD = $ 5,524 % salary = 54% % mean y = 49%

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Weekley, et al. (1985)	Supervisors of store managers (N=110) provided global SD_y estimates as well as ratings for CREPID. Subjects worked for a convenience store chain. CREPID ratings were obtained for 805 store managers.	Global estimation was based on subjects estimates of the "yearly value of the output produced" to the company. No reference to subcontracting was made.	Schmidt, et al. (1979) estimation was used for the global method. Standard CREPID method was also used, and both were compared to 40% of salary.	CREPID Mean = $ 7,701 % salary = 36% Schmidt, et al. Mean = $13,968 % salary = 66% % mean y = 51% 40% of salary Mean = $ 8,850
Burke & Frederick (1986)	Same as Burke & Frederick (1984)	Same as Burke & Frederick (1984)	Same as Burke & Frederick (1984) except that estimates were made that omitted the 97th percentile.	Standard (3 pt est) Mean = $35,192 Proced. A (3 pt. est) Mean = $27,500 Proced. B (3 pt. est) Mean = $28,151
Cascio & Ramos (1986)	Second-level managers provided CREPID ratings for 602 first-level managers in a telephone operating company.	Not reported, but assumed to be the standard CREPID notion of payoff to the organization, based on salary. These were combined with salary data to derive CREPID estimates, and then were compared to 40% of salary.	Standard CREPID method, but the originally-derived estimate of $10,081 was adjusted for restricted range because this sample of managers were job incumbents, not applicants. the authors report (p. 28) that SD_y varied between 40% and 60% of annual wage across job classes.	CREPID Mean = $10,081 SE = NR SD = NR % salary = 35% % mean y = 34%
Cronshaw, et al. (1986)	Clerical/administrative employees in the Canadian military.	Similar to Schmidt, et al. (1979)	40% of average salary.	Mean = $10,680
DeSimone, et al. (1986)	Surveyed supervisors (N=27) of medical claim approvers in a large financial service company.	Similar to Schmidt, et al.	Similar to Schmidt, et al. (1979)	Mean = $3,871 SD = $1,765 SE = $ 334 % Salary = 25%
	Actual medical claims approver performance in a large financial services company.	Payroll cost reductions that could be achieved by having fewer approvers process a similar number of claims.	Used 12 monthly averages of claims processed per day for 176 approvers. Mean was 47, SD was 11.54, extrapolated to yearly mean and SD of 11,139 and 2,735. Salary and benefits per claim were $1.79.	Mean y = $19,939 SD_y = $ 4,896 % Salary = 31%

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values	
Schmidt, et al. (1986)	U.S. Government employees across levels GS-1 to GS-18.	Value of output as sold.	40% of lowest 1984 salary level in each grade, averaged by weighting according to the number hired at each GS level.	Mean SD	= $5,429 = $2,251
Day & Edwards (1987)	Estimated utility for 43 Account Executives using ratings from 34 supervisors, in a Midwestern U.S. transportation company.	Similar to Schmidt, et al. (1979) (N=17 supervisors)	Same as Schmidt, et al. (1979)	Mean SD SE % salary % mean y	= $161,471 = $252,639 = $61,274 = 471% = 80%
		Similar to Schmidt, et al. (1979)	Modified Schmidt et al. (1979) by omitting the instruction, "In placing an overall dollar value on this output, it may help to consider what the cost would be of having an outside firm provide these products and services." (N=17 supervisors)	Mean SD SE % salary % mean y	= $180,382 = $248,153 = $60,186 = 526% = 80%
		"Worth in dollars of an employee's overall job performance".	% ROI method: (1) calculated the "average annual investments" (sum of salary plus incentive pay plus benefits, plus 40% "overhead", which equalled $65,280 per position. (2) had Had N=34 supervisors estimate the percent return on this investment (ROI) corresponding to each of the seven performance appraisal scale points. (3) Applied these figures to each Account Executive based on their actual appraisal.	Mean y SD_y % salary % mean y	= $180,920 = $ 34,103 = 99% = 19%
		Similar to Cascio (1982)	CREPID method as described in Cascio (1982)	Mean y SD_y % salary % mean y	= $ 45,230 = $ 13,392 = 39% = 30%
		Similar to Schmidt, et al. (1979)	40% of average salary	SD_y	= $ 13,723

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values	
Day & Edwards (1987) Continued	Estimated utility for 107 Mechanical Foremen using ratings from 28 supervisors in a Midwestern U.S. transportation company.	Similar to Schmidt, et al. (1979)	Same as Schmidt, et al. (1979) (N=13 supervisors)	Mean SD SE % salary % mean y	= $41,423 = $38,698 = $10,733 = 129% = 54%
		Similar to Schmidt, et al. (1979)	Modified Schmidt et al. (1979) (N=15 supervisors)	Mean SD SE % salary % mean y	= $134,335 = $258,618 = $ 66,775 = 417% = 65%
		"Worth in dollars of an employee's overall job performance"	% ROI	Mean y SD_y % salary % mean y	= $ 95,744 = $ 14,440 = 45% = 15%
		Same as Cascio (1982)	CREPID	Mean y SD_y % salary % mean y	= $43,237 = $11,988 = 37% = 28%
		Same as Schmidt, et al. (1979)	40% of salary	SD_y	= $12,881
Greer & Cascio (1987)	Estimated the performance value of route salesmen (N=62) for a Midwestern U.S. soft drink company.	Value of output as sold, similar to Schmidt, et al. (1979)	Global Estimation Model using the questionnaire-based procedure of Schmidt, et al. (1979), completed by supervisors (N=29).	Mean y SD_y % salary % mean y	= $31,979 = $14,636 = 55% = 46%
		"Contribution of labor".	CREPID method (Cascio & Ramos, 1986).	Mean y SD_y % salary % mean y	= $38,435 = $8,988 = 34% = 23%
		"Contribution Margin of salesmen" defined as revenue less variable costs.	"Cost-accounting" method that calculated the revenue less cost per unit sold, and multiplied by the quantity of units sold by each salesman.	Mean y SD_y % salary % mean y	= $44,985 = $15,864 = 60% = 35%

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values
Mathieu & Leonard (1987)	Supervisors of bank employees (Head Tellers, Operations Managers and Branch Managers) responded to a questionnaire similar to that used by Schmidt, et al.	Similar to Schmidt, et al.	The original distribution of SD_y estimates was examined for non-normality, which was assumed to result from "systematic error". The 50-15 SD_y estimates (SD_y1) differed from normal for Branch Managers and Operations Managers and marginally for Head Tellers. The 85-50 distribution (SD_y2) differed from normality for Operations Managers. So, the authors trimmed a number of "outliers" from each distribution, which normalized them. The average SD_y was used.	Head Tellers, Trimmed Mean = $2,369 SD = NR SE = NR % salary = 19% % mean y = NR Oper. Mgr., Trimmed Mean = $3,123 SD = NR SE = NR % salary = 17% % mean y = NR Branch Mgr., Trimmed Mean = $10,064 SD = NR SE = NR % salary = 44% % mean y = NR Head Teller, Untrim Median = $2,150 % salary = 17% Oper. Mgr., Untrim Median = $3,250 % salary = 18% Branch Mgr. Untrim Median = $10,000 % salary = 44%
Rich & Boudreau (1987)	Supervisors of computer programmers (N=29) in a computer manufacturing organization estimated the value of performance for computer programmers.	Similar to Schmidt, et al.	Gathered estimates of the 15th, 50th and 85th percentiles, and used the average difference as SD_y.	Mean = $15,888 SD = $14,617 SE = $ 2,761 % salary = 60% % mean y = 47%

Table 4. Results of Studies Estimating SD_y Values (Continued)

Reference	Setting	Utility Scale	Estimation Method	SD_y Values	
Edwards, et al. (1988)	Directors of Sales, Regional Managers, and Field Personnel Managers (N=33) in a National manufacturing company estimated performance value for the job of District Sales Managers.	See description for Burke & Frederick (1984) Procedure B.	Burke & Frederick (1984), Procedure B.	Mean SD SE % salary % mean y	= $63,326 = $16,177 = $ 2,816 = 174% = 72%
			CREPID-O, followed CREPID using job components from Burke & Frederick (1984), applied to 33 District Sales Managers.	Mean y SD_y % salary % mean y	= $42,002 = $12,170 = 33% = 29%
			CREPID-AP, followed CREPID using job components from Burke & Frederick (1984), applied to 33 District Sales Managers, but used archival data on performance.	Mean y SD_y % salary % mean y	= $33,475 = $11,342 = 31% = 34%
			CREPID-AJ, followed CREPID using job components from Burke & Frederick (1984), applied to 33 District Sales Managers, but used archival job analyisis data on activity frequency and importance.	Mean y SD_y % salary % mean y	= $42,318 = $11,160 = 31% = 26%
			CREPID-AA, followed CREPID using job components from Burke & Frederick (1984), applied to 33 District Sales Managers, but used archival data on performance and job analysis.	Mean y SD_y % salary % mean y	= $38,293 = $ 7,890 = 22% = 21%

* As reported by Hunter & Schmidt (1982).

Table 5. *Financial One-Cohort Entry-Level Selection Utility Model*

Cost-Benefit Information	*Entry-Level Computer Programmers*
Current Employment	4,404
Number Separating	618
Number Selected (N_s)	618
Average Tenure (T)	10 years
Test Information	
Number of Applicants (N_{app})	1,236
Testing Cost	$10/applicant
Total Test Cost (C)	$12,360
Average Test Score ($\bar{Z}_x$)	.80 SD
Test Validity ($r_{x,y}$)	.76
SD_y (per person-year)	$10,413
Financial Information	
Variable Costs (V)	5%
Tax Rate (TAX)	45%
Interest Rate (i)	10%

Utility Computation

Unadjusted Quantity = Average Tenure X Applicants Selected
= 10 Years X 618 applicants
= 6,180 person-years

Unadjusted Quality = Average Test Score X Test Validity X SD_y
= .80 X .76 X $10,413
= $6,331 per person-year

Adjusted Costs (After Taxes) = Test Costs - Tax Savings
= $12,360 - (.45 X $12,360), or .55 X $12,360
= $6,798

Utility	= Unadjusted Quantity	X	Unadjusted Quality	X	Variable Cost Adjustment	X	Tax Cost Adjustment	X	Discount Rate Adjustment	-	Adjusted Costs
	= [6,180	X	$6,331	X	.95	X	.55	X	.614]	-	$6,798

= $12.55 million

Adapted with permission from: Boudreau (1988, Table 5).

Table 6. *Entry-Level Selection Utility Decision With Financial/Economic Considerations and Employee Flows*

Cost-Benefit Information	*Entry-Level Computer Programmers*
Current Employment	4,404
Number Separating (N_s)	618
Number Acquired (N_a)	618
Average Tenure (T)	10 years
Test Information	
Number of Applicants (N_{app})	1,236
Testing Cost	$10/applicant
Total Test Cost (C)	$12,360/year
Average Test Score ($\bar{Z}_x$)	.80 SD
Test Validity ($r_{x,y}$)	.76
SD_y (per person-year)	$10,413/yr.
Financial Information	
Variable Costs (V)	5%
Tax Rate (TAX)	45%
Interest Rate (i)	10%
Flow Information	
Analysis Period	10 years
Test Application Period	7 years
Person-Years Affected	31,282
After-Cost, After Tax, Discounted Utility Increase over Random Selection (Millions)	Benefit - Cost $54.32 - $.04 = $54.28

Adapted with permission from Boudreau (1988), Table 6.

Table 7. *Entry-Level Recruitment/Selection Utility Decision With Financial/Economic Considerations and Employee Flows*

Cost-Benefit Information	*Entry-Level Computer Programmers*
Current Employment	4,404
Number Separating (N_{s_t})	618
Number Acquired (N_{a_t})	618
Average Tenure (T)	10 years
Test Information	
Number of Applicants (N_{app})	1,236
Average Test Score ($\bar{Z}_x$)	.80 SD
Financial Information	
Variable Costs (V)	5%
Tax Rate (TAX)	45%
Interest Rate (i)	10%
Flow Information	
Analysis Period	10 years
Test Application Period	7 years
Person-Years Affected	31,282

Workforce Utility Results

Staffing Variable	*Recruitment Advertising*	*Recruitment Agency*
Test Validity ($r_{x,y}$)	.76	.60
Testing Cost (C_s)	$10/applicant	$10/applicant
Recruitment Cost (C_r)	$ 1,250/Selectee	$ 2,225/selectee
Average Applicant Service Value	$52,065	$60,000
Average Applicant Service Cost	$36,445	$40,000
SD of Applicant Value (SD_y)	$10,413	$ 8,500
Value of Random Selection (Millions)	$141.04	$180.50
Cost of Random Selection (Millions)	-$ 4.55	-$ 8.10
Value Added by Testing (Millions)	$ 54.32	$ 35.00
Cost Added by Testing (Millions)	-$ 0.04	-$ 0.04
Total After-Tax, After-Cost Discounted Workforce Value (Millions)	$190.76	$207.45

Adapted with permission from: Boudreau (1988), Table 7

Table 8. *Entry-Level Recruitment/Selection/Retention Utility Model With Financial/Economic Considerations*

Cost-Benefit Information	*Entry-Level Computer Programmers*
Current Employment	4,404
Beginning Average Service Value	$52,065
Beginning Average Service Cost	$36,445
SD of Incumbent Service Value (SD_y)	$10,413/person-year
Number Separating (N_{s_t})	618
Number Selected (N_{a_t})	618
Acquisition Cost	$7,000/Hire
Separation Cost	$7,000/Separation
Number of Applicants (N_{app})	1,236
Average Applicant Service Value	$52,065/year
Average Applicant Service Cost	$36,445/year
Average Test Score ($\bar{Z}_x$)	.80 SD
SD of Applicant Service Value (SD_y)	$10,413/person-year
Testing Cost	$10/applicant
Variable Costs (V)	5%
Tax Rate (TAX)	45%
Interest Rate (i)	10%
Analysis Period	10 years

Workforce Utility Results

Staffing Variable	*Option 1*	*Option 2*	*Option 3*	*Option 4*
Test Validity ($r_{x,y}$)	0.00	0.76	0.76	0.76
Separation Effect	$0	$0	$2,707	-$2,707
After-Tax, After-Cost Discounted Work Force Value (Millions)	$200.31	$242.10	$351.69	$132.50

Adapted with permission from: Boudreau (1988), Table 8

Table 9. *Internal/External Recruitment/Selection/Retention Utility Decision With Financial/Economic Considerations*

Cost-Benefit Information	*Entry-Level Computer Programmers*	*Upper-Level Data System Manager*
Current Employment	4,404	1,000
Beginning Average Service Value	$52,065	$57,272
Beginning Average Service Cost	$36,445	$40,000
Number Separating	618	100
Number Selected	718	0
Number Promoted	100	100
Acquisition Cost	$7,000	NA
Separation Cost	$7,000	$8,000
Promotion Cost	NA	$8,000
Number of Applicants	1,436	3,786
Average Applicant Service Value	$52,065/yr.	1.10 times average Programmer value
Average Applicant Service Cost	$36,445/yr.	1.10 times average Programmer value
Average Test Score	.80 SD	2.32 SD
SD of Applicant Value (SD_y)	$10,413/yr.	$11,454/yr.
Testing Cost	$10/applicant	NA
Assessment Center Cost	NA	$1.44 million/yr.
Variable Costs	5%	5%
Corporate Tax Rate	45%	45%
Corporate Interest Rate	10%	10%
Analysis Period	10 years	10 years

Table 9. *Interna/External Recruitment/Selection/Retention Utility Decision With Financial/Economic Considerations* (Continued)

Total Workforce Utility Results

	Options				
HRM Activity	1	2	3	4	5
Programmer Selection Validity	0.00	0.76	0.76	0.76	0.76
Programmer Promotion Effect	$0	$0	$0	-$625	-$625
Manager Promotion Validity	0.00	0.00	0.35	0.35	0.35
After-Cost, After-Tax, Discounted Total Workforce Value (Millions)	$249.86	$296.90	$302.51	$278.68	$198.38

Adapted with permission from: Boudreau (1988, Table 9).

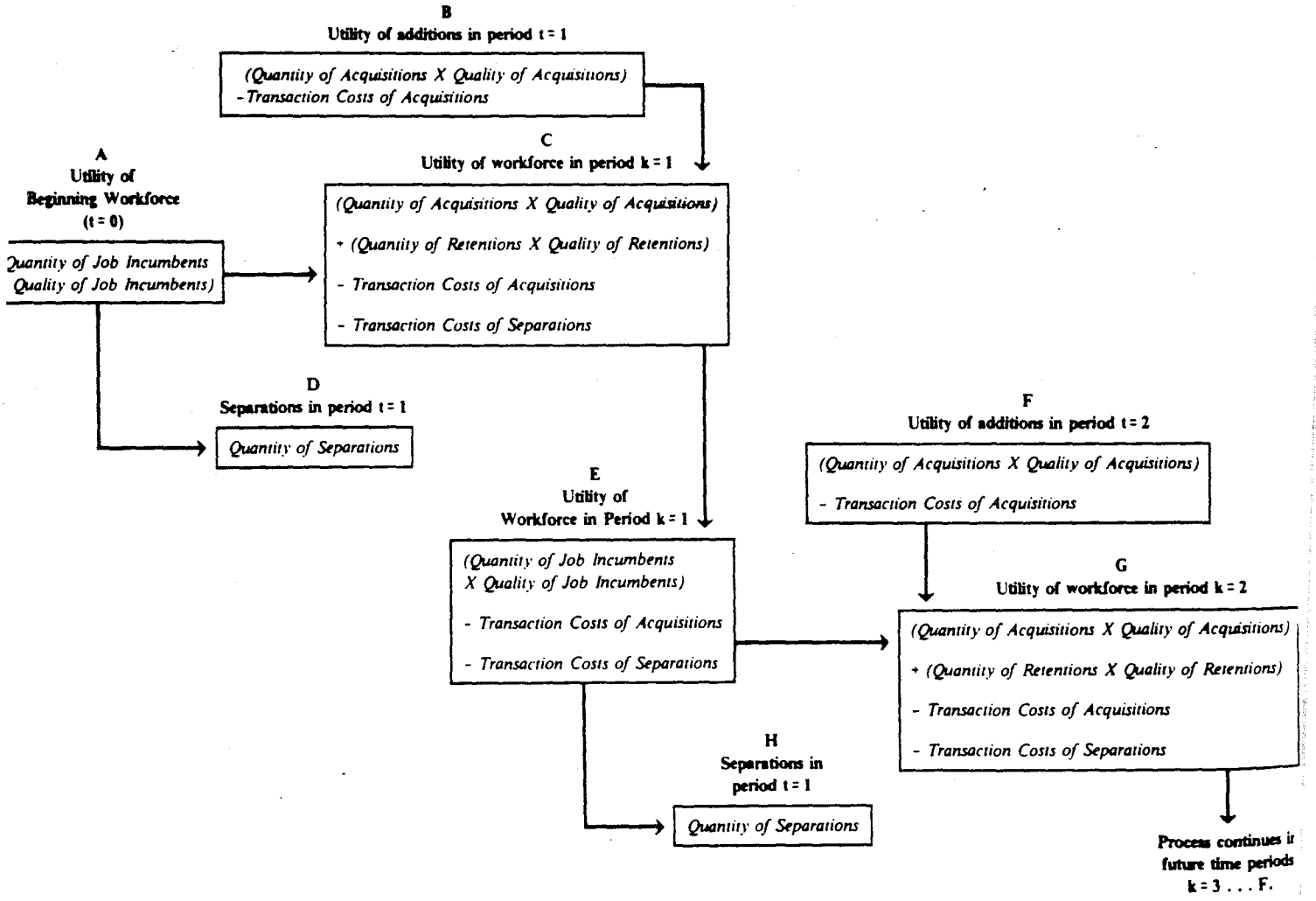
B
Utility of additions in period t = 1
(Quantity of Acquisitions X Quality of Acquisitions)
- Transaction Costs of Acquisitions
A
Utility of
Beginning Workforce
(t = 0)
Quantity of Job Incumbents
Quality of Job Incumbents)
C
Utility of workforce in period k = 1
(Quantity of Acquisitions X Quality of Acquisitions)
+ (Quantity of Retentions X Quality of Retentions)
- Transaction Costs of Acquisitions
- Transaction Costs of Separations
D
Separations in period t = 1
Quantity of Separations
F
Utility of additions in period t = 2
(Quantity of Acquisitions X Quality of Acquisitions)
- Transaction Costs of Acquisitions
E
Utility of
Workforce in Period k = 1
(Quantity of Job Incumbents
X Quality of Job Incumbents)
- Transaction Costs of Acquisitions
- Transaction Costs of Separations
G
Utility of workforce in period k = 2
(Quantity of Acquisitions X Quality of Acquisitions)
+ (Quantity of Retentions X Quality of Retentions)
- Transaction Costs of Acquisitions
- Transaction Costs of Separations
H
Separations in
period t = 1
Quantity of Separations
Process continues in
future time periods
k = 3 . . . F.

Figure 1. *Diagram of External Movement Utility Model Concepts*

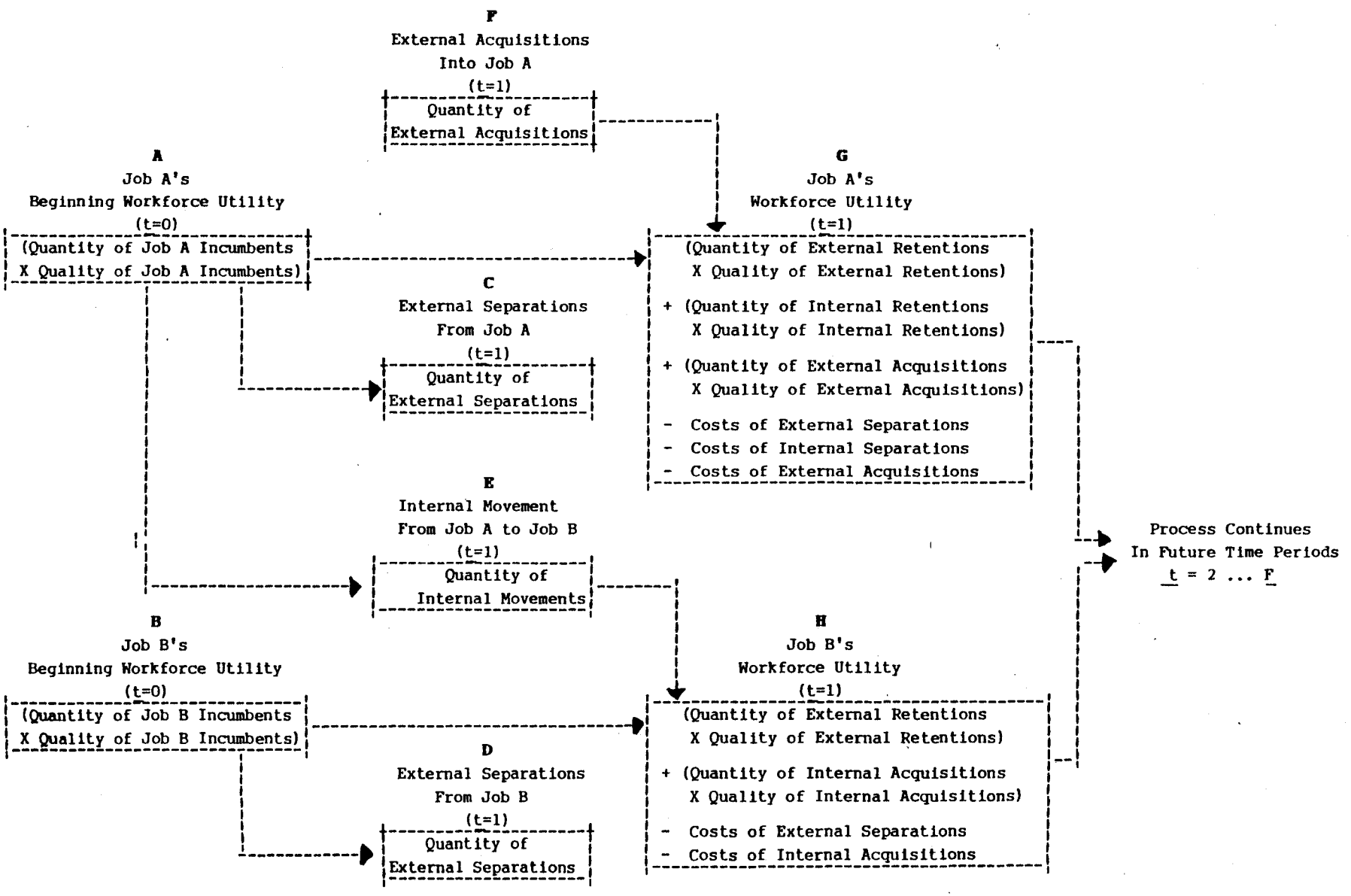
F
External Acquisitions
Into Job A
(t=1)
Quantity of
External Acquisitions
A
Job A's
Beginning Workforce Utility
(t=0)
(Quantity of Job A Incumbents
X Quality of Job A Incumbents)
G
Job A's
Workforce Utility
(t=1)
(Quantity of External Retentions
X Quality of External Retentions)
+ (Quantity of Internal Retentions
X Quality of Internal Retentions)
+ (Quantity of External Acquisitions
X Quality of External Acquisitions)
- Costs of External Separations
- Costs of Internal Separations
- Costs of External Acquisitions
C
External Separations
From Job A
(t=1)
Quantity of
External Separations
E
Internal Movement
From Job A to Job B
(t=1)
Quantity of
Internal Movements
Process Continues
In Future Time Periods
t = 2 ... F
B
Job B's
Beginning Workforce Utility
(t=0)
(Quantity of Job B Incumbents
X Quality of Job B Incumbents)
H
Job B's
Workforce Utility
(t=1)
(Quantity of External Retentions
X Quality of External Retentions)
+ (Quantity of Internal Acquisitions
X Quality of Internal Acquisitions)
- Costs of External Separations
- Costs of Internal Acquisitions
D
External Separations
From Job B
(t=1)
Quantity of
External Separations

Figure 2. *Diagram of Internal-External Movement Utility Model Concepts*

Figure 3. Matrix of Research Issues in Utility Analysis

Type of Research Study

Research Content	**V** Conceptual Extension	**W** Simulation	**X** Empirical Demonstration	**Y** Data-Based Inference	**Z** Theory Testing
Outcome Evaluation **A**	*	*	*		
Human Resource Planning **B**					
Recruiting **C**	*	*			
Selection **D**	*	*	*	*	
Training **E**	*	*	*	*	
Compensation **F**					
Internal Movement **G**	*	*			
Turnover and Layoffs **H**	*	*	*	*	
Performance Assessment **I**	*		*	*	
HRM Decision Processes **J**	*		*		