

Selection Utility Analysis: A Review
and Agenda for Future Research

by

John W. Boudreau

Working Paper 88-02

Prepared for the International Conference on Advances in Selection and Assessment, University of Manchester Institute of Science and Technology May 21-21, 1987.

This research was carried out with the support from the U.S. Army Research Institute, contract SRFC #MDA903-87-K-0001. The view, 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.

This paper will appear in a book of Conference papers and comments, Mike Smith and Ivan Robertson, editors. Published by John Wiley and Sons, 1988.

This paper has not undergone formal review or approval of the faculty of the ILR School. It is intended to make the results of Center research, conferences, and projects available to others interested in human resource management in preliminary form to encourage discussion and suggestions.

Introduction

Whether they are line managers, human resource management staff, or organizational psychologists, managers of human resources must make decisions (e.g., hiring, placement, training, compensation, performance appraisal, feedback, etc.) in which theories of human work behavior play an important role. I/O psychologists (and other social scientists) find the organizational environment a rich source of information for advancing knowledge and testing theories about employment relationships, their antecedents and consequences. Applied research articles inevitably discuss "practical implications," but what is the real value of human resource productivity improvement programs?

The Human Resource Management (HRM) functions of industrial organizations typically lack the influence and visibility of the more "traditional" management functions (such as marketing, finance, operations and accounting). Journals for HRM professionals routinely lament the slow progress of organizations in implementing programs that have gained wide acceptance by scientists (cf. Jain & Murray, 1984). Journals and books routinely admonish and instruct the HRM professional to "sell" their programs by emphasizing their effects on organizational goal attainment (Bolda, 1985; Gow, 1985; Jain & Murray, 1984; Sheppeck & Cohen, 1985). Indeed, the question of whether the HRM function contributes to corporate profit is still important enough to merit recent discussion in a widely-cited professional journal (Gow, 1985). It is difficult to imagine such a debate regarding the Finance, Marketing, Accounting, or Engineering departments. Is the contribution of I-O psychology

and other social sciences to human resource management really so intangible compared to these other management functions?

Utility analysis involves describing, predicting and explaining the usefulness or desirability of decision options, and analyzing how that information can be used in decision making. In I/O psychology, the term utility analysis has become associated with a specific set of models that reflect the consequences (usually performance-related consequences, such as output as sold, sales, net benefits, or reduced costs) of programs (e.g., selection, recruitment, training, performance feedback, goal setting, compensation, internal staffing and turnover control) to enhance work force productivity. Utility analysis offers a method for better understanding the role of I/O psychology and other social sciences in improving HRM decisions and organizational performance. Moreover, it offers an invitation to I/O psychologists and other social scientists to adopt a truly interdisciplinary approach to important scientific and practical questions.

A Conceptual Framework

Selection utility models can present very complex algebraic and statistical formulas, but their basic concepts are quite logical, simple, and direct.

Selection Utility Analysis Supports Decisions

Selection Utility models are "decision aids" (Edwards, 1977; Einhorn & McCoach, 1977; Einhorn, Kleinmuntz & Kleinmuntz, 1977; Fischer, 1976; Huber, 1980; Keeney & Raiffa, 1976), tools for predicting, explaining, describing and analyzing decisions. Decision aids assist decision makers in overcoming limited cognitive capacity ("limits on rationality" according to March & Simon, 1958)

by providing a consistent and structured framework within which to compare decision options. Selection utility models offer a consistent structured framework for considering selection consequences, and for communicating those consequences to constituencies in generally-understood units (e.g., dollars).

Applying selection utility analysis models requires: (1) a set of decision options to be considered (e.g., two or more different selection systems); (2) a set of attributes reflecting the characteristics of the options that affect valued outcomes (e.g., validity coefficients, costs, effect sizes, quantity of employees affected), combined with a "utility scale" that reflects the value of each attribute level to the decision maker (e.g., a scale translating selection validity into dollar-valued performance); and (3) a payoff function, the weighting scheme or other combination rule reflecting the relationships among the attributes in estimating total utility for each option (e.g., an algebraic formula describing the relationship between validity, cost and utility).

Viewing selection utility models as decision tools is quite consistent with their historical development, as will be discussed subsequently. Yet, little research has addressed whether they have any effect on actual decisions. Throughout this chapter, I will propose that selection utility research must proceed with a keen awareness of the decisions it is designed to support, and that such an awareness suggests some very different research questions and directions from those currently being pursued.

The Unit of Analysis: Human Resource Productivity Improvement Programs

The options considered by utility analysis models are human resource productivity improvement programs. Such programs are combinations of activities (or procedures governing activities) that affect the organizational value of the work force.

Decisions about individuals versus decisions about programs.

As Cronbach and Gleser (1965, p. 9) noted, utility analysis is most appropriate for programs that will be applied to many individuals over time. Thus, programs embody procedures, rules or "strategies" (Cronbach & Gleser, 1965, p. 19) intended to be applied to many individuals. Selection utility analysis typically focuses on programs using tests to select new employees. Such programs involve rules indicating how the test is to be applied to applicants, and how the test results are to be used in choosing among applicants.

A distinctive feature of "programs" is that they affect many decisions about individuals. The selection program, for example, will affect hiring decisions for many applicants. Selection programs work by providing information that causes certain applicants to be hired who otherwise might have been rejected (and vice versa), resulting in a more productive group of hires than would be possible without the test. The decision about each applicant, however, is not the focus of selection utility analysis. Rather, it is the decision to implement a program that will alter the way many applicants are evaluated for hiring. Because such 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, as we shall see.

The organization context. Selection programs "work" if they increase the correctness of those choices in ways that are important enough to offset the costs of the information. However, each organization or decision maker will define importance differently, depending on the constituents involved, the way it employs employee productivity, and how selection interacts with other programs such as compensation and training. Research has often proceeded as if these factors are held constant, but more integrative future research should examine these relationships more explicitly.

Three Basic Utility Analysis Variables: Quantity, Quality and Cost

Selection utility models can be expressed in terms of three basic attributes (Boudreau, 1984, 1987, in press a, in press b; Boudreau & Berger, 1985a, 1985b): (1) Quantity, reflecting the quantity of employees and time periods affected by program consequences; (2) Quality, reflecting the consequences (per person, per time period) associated with the program; and (3) Cost, reflecting the resources required to implement and maintain the program. The payoff from a selection program can be derived by taking the product of Quantity and Quality, and then subtracting Cost. Generally, the program exhibiting the largest positive difference is preferable.

Selection utility models differ in the manner in which they define each of these three variables, but they all can be understood within this framework.

Historical Development

Selection utility analysis models responded to the inadequacies of traditional measurement and test theory in expressing the

usefulness of tests. As Cronbach and Gleser (1965, pp. 135-136) stated:

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. 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.

Therefore, the history of selection utility analysis will be discussed from a decision-making perspective, focusing on each model's definition of "goodness", and using the concepts of Quantity, Quality and Cost to assess the decision value of each historical development.

Defining Payoff with the Validity Coefficient

Model description. The validity coefficient (the correlation between a predictor measure and some criterion measure of subsequent behavior, usually expressed as $r_{x,y}$) is the attribute of selection that has the longest history. The correlation coefficient and

indexes derived from it (e.g., the index of forecasting efficiency and coefficient of determination, Hull, 1928) lead to the conclusion that only relatively large differences in the validity coefficient produce important differences in the value of a test.

Evaluation from a decision-theory perspective. As indicators of a test's usefulness for decisions, such formulas are quite deficient. First, the implied payoff function is unrealistic. The correlation coefficient measures squared deviations from a predicted linear function, treating any deviation of the predicted value from the linear function as equally undesirable. Overpredicting the best candidate's future performance is treated as equivalent to underpredicting it, yet the former is unlikely to cause serious decision errors, while the latter easily could.

In terms of the three basic program attributes (i.e., quantity, quality and cost), correlation-based models reflect neither the quantity of time periods affected by the selection decisions nor the quantity of employees affected in each time period. They provide only indirect evidence of the predictor's effect on work force quality. Finally, they fail to acknowledge the fact that developing and applying selection programs entails costs.

Defining Payoff Based on the Success Ratio

Model description. The "success ratio" represents the percentage of selected individuals who are successful on the job. According to the Taylor-Russell (1939) tables, when other parameters are held constant: (1) higher validities result in 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 result in

more improved success ratios (because the lower the selection ratio, the more "choosy" is the selection decision, and the predictor scores of selectees lie closer to the upper tail of the predictor distribution); (3) base rates closer to .50 result in more improved success ratios because valid selection has less value as you approach a base rate of zero (where none of the applicants can be successful) or as you approach a base rate of 1.0 (where all applicants can succeed even without selection).

Evaluation from a decision-making perspective. Regarding the payoff function, a dichotomous criterion (i.e., selectees are either satisfactory or unsatisfactory) 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). The more typical (though not uniformly applicable) situation is where performance differences within the two groups exist. Under such situations, a continuous criterion would be more appropriate. Still, the Taylor-Russell model may provide adequate decision support for some situations. Cascio (1982, p. 146) 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). In terms of the three program attributes (i.e., quantity, quality and cost), the Taylor-Russell model (like its predecessors) ignores both the quantity of employees affected and the number of time periods during which that effect will last.

The model does a better job of describing the change in quality produced by the program (in that it provides some idea of increased success probabilities), but this quality measure must be interpreted differently from situation to situation. Finally, the Taylor-Russell model completely ignores costs.

Defining Payoff Based on the Standardized Criterion Level

Model description. The major criticism of the success ratio was that it's dichotomous criterion failed to reflect the true variation in performance. The next version of the selection utility model attempted to remedy this by defining a continuous criterion as the payoff function. Brogden (1946a, 1946b) used the principles of linear regression to demonstrate the relationship between the correlation coefficient and increases in a criterion (measured on a continuous scale).

Assuming a linear relationship between criterion scores and predictor scores, if we derived the best, linear, unbiased estimate of the change in standardized criterion scores (\bar{Z}_y) corresponding to a change in standardized predictor scores (\bar{Z}_x) in the applicant population, the linear prediction equation would be:

$$\bar{Z}_y = (r_{x,y})(\bar{Z}_x) \quad (1)$$

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 2.

$$\bar{Z}_y = (r_{\underline{x},y}) (\bar{Z}_x) \quad (2)$$

This utility model reflects a continuous criterion (expressed in standardized, Z-score units) as its payoff function, and includes as attributes both the validity coefficient (in the population of applicants to which the predictor will be applied) and the average standardized predictor score of those applicants chosen. The validity coefficient and its derivation were well established, and Naylor and Shine (1965) computed extensive tables showing, for each selection ratio, the corresponding average standardized predictor score (assuming normally distributed predictor scores and top-down applicant selection).

Evaluation from a decision-making perspective. This utility model addresses one shortcoming of the Taylor-Russell model by defining payoff based on a continuous criterion. However, because the criterion is expressed in standard-score units, it is difficult to interpret in units more natural to the decision process (e.g., dollars, units produced, reduced costs, etc.). Also, this payoff function 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 mean of the applicant population which, by definition, has a standard score of zero). The total utility from the program is not computed, only the increment over not using the predictor.

Considering the three basic utility model concepts (i.e., quantity, quality and cost), the quantity of employees and time

periods are not reflected, the quality criterion is in statistical rather than tangible units, and the cost of the selection program is still omitted.

Defining Payoff in Terms of Dollar-Valued Criterion Levels

Model description. The most obvious drawback of using standardized criterion levels as the payoff function is that they are difficult to interpret in "real" units. Selection device development and implementation activities are often expressed as costs (i.e., required uses of resources), usually scaled in dollars. 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 .5 standard deviations greater than what we would obtain without the predictor?" Obviously, many HRM managers may not even be familiar with the concept of a standard deviation. Almost certainly, they would find it difficult to attach a dollar value to a .5 standard deviation increase in the criterion (especially considering that the decision makers may never actually observe the appropriate population for that standard deviation--the population of applicants to which the predictor would be applied).

Both Brogden (1946a, 1946b, 1949) and Cronbach and Gleser (1965, pp. 308-309) eventually derived their utility formulas in terms of "payoff", rather than standardized criterion scores. They also both included the concept of costs. Thus, they originated the notion of expressing utility on a dollar-valued scale. To accomplish this, they introduced a scaling factor that translated standardized criterion levels into dollar terms, and they added a term for the costs of the selection program. The scaling factor

is simply the dollar value of a one-standard-deviation difference in criterion level (symbolized in various ways, including σ_y , σ_e , and \underline{SD}_y , the latter being used here). The cost factor is usually expressed as the cost to administer the predictor to a single applicant (usually symbolized as \underline{C}). Finally, the utility value is symbolized as $\underline{\Delta U}$, 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 3.

$$\underline{\Delta U} = (\underline{SD}_y)(r_{\underline{x},y})(\underline{Z}_{\underline{x}}) - \underline{C}/\underline{SR} \quad (3)$$

The per-applicant cost (\underline{C}) is divided by the selection ratio (\underline{SR}) to reflect total cost of obtaining each selectee (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).

Thus, Equation 3 depicts the incremental dollar-valued criterion level of those selected with a predictor (\underline{x}), in a population of applicants where the validity coefficient is $r_{\underline{x},y}$; where a one-standard-deviation difference in criterion levels equals \underline{SD}_y ; where the average standardized predictor score of those selected is equal to $\underline{Z}_{\underline{x}}$; and the per-selectee cost of using the predictor is ($\underline{C}/\underline{SR}$). To express the total gain from using the predictor to select \underline{N}_s selectees, we simply multiply the benefits by the number selected, change the symbol for incremental utility

from $\bar{\Delta U}$ to ΔU , and multiply the per-applicant cost by the number of applicants (N_{app}) as shown in Equation 4.

$$\Delta U = \left(\frac{N}{s}\right) \left(\frac{SD}{y}\right) (r_{x,y}) \left(\frac{Z}{x}\right) - (C) (N_{\text{app}}) \quad (4)$$

Cronbach and Gleser (1965, p. 39) also recommended computing the difference in utility between two tests, which simply involves substituting the difference in validities for $r_{x,y}$ and the difference in costs for C in Equations 3 and 4.

Finally, to incorporate the duration of the effects of better-selecting one employee cohort, Schmidt, Hunter, McKenzie & Muldrow (1979) multiplied the benefit component of these models by the expected tenure of the hired cohort (i.e., T).

Evaluation from a decision making perspective. Scaling the per-person, per-time-period incremental criterion level in dollars seems more in keeping with organizational objectives to increase dollar profits. The B-C-G utility model incorporates a scaling factor ($\frac{SD}{y}$) to translate standardized criterion levels into dollars. Measuring $\frac{SD}{y}$ has proven controversial as will be discussed subsequently.

The B-C-G model incorporates the three basic selection utility analysis components (i.e., quantity, quality and cost). Quantity is incorporated in the number selected and their average tenure. Quality is incorporated in the product of $r_{x,y}$, $\frac{Z}{x}$, and $\frac{SD}{y}$ (producing the per-person, per-time-period incremental dollar-

valued criterion level). Costs to develop and implement the selection program are contained in the cost factor (C).

Encouraging more widespread selection utility analysis applications. The B-C-G utility model remained largely unnoticed by I/O psychologists (at least in terms of published research studies), though this model represented a fundamental and important alternative to traditional measurement theory as a framework for I/O psychology research. The reasons for this lack of attention are unclear. It is likely that the algebraic complexity of these models proved daunting to managers, so researchers may have encountered difficulty communicating the purpose and importance of the models. Moreover, researchers may have incorrectly assumed that all model parameters must be fairly accurately measured to apply the models because the aim was to produce a point estimate of utility. This misconception still exists today, as discussed subsequently.

Hunter & Schmidt (1982) and Schmidt, et al. (1979) noted the limited application of the B-C-G models and proposed that three widely-held misconceptions might explain it: (1) The belief that utility equations are of no value unless the data exactly fit the linear homoscedastic model, and all marginal distributions are normal (in fact, the B-C-G model only introduces the normality assumption for "derivational convenience," Hunter & Schmidt, 1982, p. 243, as it provides an exact relationship between the selection ratio and the average standard test score of selectees); (2) The belief 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 (in fact, "validity generalization," Hunter, Schmidt, & Jackson, 1982a research suggests that much of the variability in validity coefficients observed across studies is due to artifacts of the studies rather than real differences in the predictor-criterion relationship); (3) the belief that the parameter scaling standardized values into dollars (i.e., $\frac{SD}{Y}$) is difficult or impossible to measure. A later section will discuss this measurement issue in detail.

Financial/Economic Factors: Utility Analysis as an Investment Model

The dollar-valued payoff model led to speculation that selection utility could provide a link between Personnel/HRM research and the more traditional management functions (Landy, Farr & Jacobs, 1982, p. 38; Cascio and Silbey, 1979). Recent research suggested enhancements to the traditional selection utility model designed to incorporate financial and economic considerations into the analysis. Boudreau (1983a) suggested that by measuring utility with a payoff function reflecting sales or "the value of output as sold," researchers were probably overstating HRM program effects on 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, "sales (or service) value" (i.e., the change in sales revenue or output as sold) differs from "service cost" (i.e., the change in organizational costs associated with changed service value), which differs from "net benefits" (i.e., the difference between service value and service costs) produced by an HRM

intervention. HRM programs that improve sales value can 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, in press). Boudreau (1983a) included the effects of HRM programs on service costs by multiplying the incremental service value increase by a proportion (V) reflecting the change in net benefits per change in sales value. Second, most organizations pay taxes on income to Federal, State and Local governments. Thus, pre-tax net benefits and costs may be overstated when they fail 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., $1 - \text{TAX}$) to reflect after-tax effects. Third, returns can often be invested to earn interest. A dollar received in the future is worth less than a dollar received today, because the latter can earn interest in future periods. Boudreau demonstrated how the interest rate earned on program returns (i.e., i) could be incorporated into the selection utility analysis model.

Boudreau proposed that by incorporating these financial adjustments, reported utility values would better reflect the economic realities of organizations, would be more comparable to investment values reported for programs in other management functions, and might be more credible to managers accustomed to working with financial analysis. Cronshaw and Alexander (1985) also argued for HRM programs as financial investments, suggesting

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).

The Employee Flows Utility Model

To move beyond selection utility models reflecting only the consequences of hiring one group of employees, Boudreau (1983b) proposed the "employee flows" model. Organizations seldom invest in a selection program to use it once and then stop. Rather they can continuously reapply the program as new members enter the workforce. To analyze only the first-cohort effects is tantamount to analyzing an investment in new manufacturing facilities based on only one production run. Clearly, such a focus omits a large part of the decision's effects.

Boudreau proposed that a more accurate approach would reflect the number of "treated" (i.e., better-selected) employees in the work force in each future period (i.e., N_k), and the costs incurred to select the employees joining the work force in each future time period (i.e., C_k). Boudreau acknowledged that any of the utility variables could change over time, and noted that the "flows" model could reflect such temporal changes. The flows model highlighted the importance of the Quantity concept (the number of employees and time periods affected by the intervention) in explaining the potentially huge effects of human resource management decisions.

To illustrate the effects of employee flows, Boudreau used data from the Schmidt, et al. (1979) study. However, whereas the earlier authors had computed quantity by multiplying the size of

the first selected cohort (i.e., 618) by its expected tenure (i.e., approximately 10 years), Boudreau assumed that each cohort of 618 selected employees remained on the job for 10 years, and then left to be replaced by a newly-hired cohort. The effect of these assumptions was that the number of better-selected employees in the workforce steadily rose (by 618 per year) during the first 10 years (as new employees were added to the workforce and joined previously-selected employees), until it reached 6,180. For the next five years, vacancies created by the separation of systematically-selected employees were filled with systematically-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. Boudreau used an analysis period of 25 years (i.e., the program was applied for 15 years and then stopped).

Using the one-cohort model, Schmidt, et al. proposed that the effects of improved selection would affect 6,180 person-years of productivity (i.e., 10 years average tenure times 618 hired employees). Using the employee flows model, Boudreau calculated that the program would affect 92,700 person-years of productivity (i.e., the sum of N_k over all 25 years). This is a critically important point. Repeated application of improved selection programs can affect huge numbers of employee-years of productivity producing massive potential productivity effects. The key to understanding this is to recognize that HRM programs are likely to be re-applied over time, rather than applied only once. 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 attempt to justify investments in HRM programs based only on the first cohort affected.

Integrating Employee 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), 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.

Two factors may make such assumptions simplistic compared to organizational reality. First, as Boudreau and Rynes (1985) noted, common wisdom in the recruitment literature suggests that more rigorous or intrusive selection methods may affect the size and/or characteristics of applicant pools (though there is little research to support or refute this suggestion). Second, 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 proposed that every parameter of the utility model could be affected by recruitment strategies, or could be affected by applicant reactions to selection devices. For example, applicant populations might be more homogeneous (reducing both \overline{SD}_y and the correlation coefficient) if more stringent standards were applied to recruitment sources. 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). In the Boudreau-Rynes model, utility values are represented on an absolute scale, reflecting both the average and the incremental value of the selectees. They demonstrated how the incremental selection utility model alone may severely understate the combined value of integrated recruitment and selection, and how improved base rates may actually offset reductions in selection effectiveness caused by restrictions in range or smaller applicant pools.

The State of the Art in Empirical Research:

Utility Values for Selection Programs

I recently reviewed empirical studies through 1986, with utility values for 39 interventions. The unavoidable conclusion is that selection programs pay off handsomely. Virtually every study has produced dollar-valued payoffs that clearly exceeded costs. Studies dealing with many employees and multiple-year tenure can produce utility values as high as \$20 to \$30 million (e.g., Schmidt, et al., 1979; Cascio & Ramos, 1986). The clear positive payoff from selection programs remains evident in studies with relatively small $\frac{SD}{\bar{y}}$ values as well as in studies with large $\frac{SD}{\bar{y}}$ values, and with selection ratios as high as 50%. As would be expected, the largest utility values result in studies where large numbers of individuals are affected by the program (i.e., where $N_{\bar{s}}$ is large). Many of the studies were designed to examine whether substituting a more-valid selection method for a less-valid one (usually in interview) produced greater dollar-valued payoff (Burke & Frederick, 1985; Cascio & Ramos, 1986; Cascio & Silbey, 1979; Ledvinka, Simonet, Neiner & Kruse, 1983; Schmidt, et al., 1979;

Schmidt, Mack & Hunter, 1984; Rich & Boudreau, 1987). In every case the more valid (and usually more costly) selection procedure produced the greater utility.

Utility values as measured by the B-C-G model appear to be quite high. Moreover, the measured costs of improved selection are usually minuscule compared to the benefits. Such a conclusion seem inconsistent with the ongoing debate over whether human resource management's contribution is ignored, whether HRM issues should be considered in organizational planning, and whether HRM programs represent appropriate uses for organizational resources.

The State of the Art In Empirical Research:

Measuring the Standard Deviation of Performance in Dollars

The standard deviation of dollar-valued job performance in the applicant population (\underline{SD}_y) was characterized as the "Achilles' Heel" of utility analysis by Cronbach and Gleser (1965, p. 121). The large amount of recent research aimed at better estimating this elusive concept suggests that many of today's selection utility researchers agree, and regard accurate measurement of \underline{SD}_y as a fundamental requirement for useful selection utility analysis research (Burke & Frederick, 1984, 1985; Weekley, O'Connor, Frank & Peters, 1985). I located 26 studies, with over 100 individual \underline{SD}_y estimates. The trend in research activity is evident, with only 5 studies between 1953 and 1979 and 21 studies between 1979 and 1986.

Effects of Setting

A wide variety of occupations has been examined, with the choice of occupation usually determined by whatever research setting

presented itself to the researchers. Jobs where workers exercise more discretion regarding the quantity and quality of production and/or where variation in production has large implications for organizational goals should exhibit higher \underline{SD}_y values than jobs without these characteristics, assuming the same variability in skill and motivation in each work force. Of course, different jobs probably face different ranges of skill and motivation among job incumbents, which could cause even jobs of high discretion and importance to produce lower observed \underline{SD}_y values than jobs without those characteristics. Most \underline{SD}_y studies focused on only one job, making across-job comparisons difficult (because jobs, measurement methods, settings and time periods are confounded). In studies of more than one job (Wroten, 1984; Eaton, Wing & Lau, 1985), but results suggested the \underline{SD}_y estimation method affected whether cross-job differences were observed.

Every \underline{SD}_y study used job titles to identify employees holding similar job duties and tasks. Such an approach may inadvertently include across-job differences in the \underline{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. If the selection test will be used primarily to select programmers assigned as coders, this will overstate \underline{SD}_y (and vice versa). Moreover, if jobs with different titles actually share duties, this could explain the lack of consistent job-to-job differences.

Effects of Payoff Scales

The most 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). Payoff measures should 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.) in different decision situations, consistent with the objectives of decision makers (see Cronbach & Gleser, 1965, p. 23). The payoff scales used in selection utility analysis research focus on the consequences of increased labor force quality, so the payoff from improved selection depends heavily upon how the quality enhancements caused by such programs will be used.

Three general uses for improved labor force quality are: (1) raising the quantity of production; (2) raising the quality of production; and (3) reducing production costs. Managers may use labor force quality increases in any or all of these ways, or may combine them. A payoff scale defined in terms of profit can reflect any or all of these uses. Payoff scales reflecting only quantity or quality or cost may fail to reflect the two omitted uses. Payoff scales reflecting revenue enhancements (through higher quality or quantity) and cost reductions dominate the selection utility literature, though profit-based scales are emerging.

Payoff as cost reduction. Most of the earliest selection utility analysis applications focused on cost reduction as the salient outcome of 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 replacing separations. More recently, Eaton, Wing & Mitchell (1985) measured payoff in terms of the avoided costs of additional tanks, and Schmidt and Hunter (1983) noted that increased work force productivity might reduce "payroll costs" by producing the same amount of output with a smaller number of employees (p. 413). Arnold, Rauschenberger, Soubel & Guion (1983) adopted the premise that improved selection would allow hiring fewer employees to do an equivalent amount of work. These payoff functions are also consistent with the "behavioral costing" approach to HRM program analysis described by Cascio (1982). While cost reductions often offer a highly visible and salient payoff function in situations where cost reduction is the dominant consideration, its deficiencies have led researchers to explore further options.

Payoff as the "value of output as sold". Schmidt, et al. (1979) measured $\frac{SD}{y}$ in terms of the "yearly value of products and services", and the "cost of having an outside firm provide these products." 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." Most research has focused on similar payoff scales (Bobko, Karren & Parkington, 1983; Bolda, 1985; Burke & Frederick, 1984, 1985; Cascio, 1982; Cascio & Ramos, 1986; Cascio & Silbey, 1979; Eaton, Wing & Lau, 1985; Eaton, Wing & Mitchell, 1985; Eulberg, O'Connor & Peters, 1985; Ledvinka, et al., 1983; Mathieu & Leonard, 1986; Reilly & Smither, 1985; Schmidt, et al., 1984; Weekley, et al., 1985; Wroten, 1984). The "sales

value" payoff scale reflects the increased revenue generated by employees as a result of the program, but it may be a deficient payoff definition. When organizational investments are evaluated based on profit contribution, evaluating selection investments based only on revenue contribution may artificially inflate selection utility values.

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. This suggests defining payoff as the contribution to organizational profits. Brogden and Taylor (1950) proposed the "dollar criterion" reflecting the sales revenue generated when a product is sold, less any production costs. 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. Only one study actually adopted a payoff function reflecting profit contribution (Reilly & Smither, 1985), with results suggesting that the graduate students in their simulation differed most in their $\frac{SD}{Y}$ estimates when they were asked to consider "net revenue" rather than "new sales," or "overall worth". Similarly, Bobko, et al. (1983) found that sales counselor supervisors exhibited much more variable $\frac{SD}{Y}$ estimates when attempting to estimate "yearly value to the company" rather than "total yearly dollar sales."

Summary. Although costs, sales and profits have enjoyed some attention as payoff functions, any payoff function should be judged

on its ability to improve decision quality or better describe, predict and explain decisions. All existing payoff scales reflect a concern with productivity-based outcomes, virtually ignoring other consequences of selection decisions (e.g., community relations, work force attitudes, adherence to a code of ethics). Every payoff function will be deficient in some way, so the fundamental consideration is how the organization will use such quality improvements (i.e., increasing revenue versus reducing costs versus increasing profit). The typical research approach of comparing $\frac{SD}{y}$ estimates based on different payoff functions fails to reflect this fundamental decision context. Future research should focus on how productivity improvements are actually used so that the payoff function better reflects the actual decision.

Effects of the Focus Population

Virtually all $\frac{SD}{y}$ estimates are based on the incumbent population (e.g., Bobko, et al., 1983; Burke & Frederick, 1985; Cascio & Silbey, 1979; Eaton, Wing & Mitchell, 1985; Janz & Dunnette, 1977; Schmidt, et al., 1979; Schmidt & Hunter, 1983; Wroten, 1984), probably because the incumbent population is most familiar to job supervisors. However, the incumbent population is not technically the appropriate population of interest. For selection utility, the appropriate population is the applicant population to which selection procedures will be applied. This population may differ from the incumbent population for a number of reasons.

First, if certain procedures (for example, promoting out the best performers and dismissing the worst performers) make the incumbent population a restricted sample of applicant job

performance (Schmidt, et al., 1979), then \underline{SD}_y estimated on job incumbents will be downward-biased. Second, applicant population changes over time due to different recruitment procedures or labor market influences (Becker, 1985; Boudreau & Rynes, 1985) may operate either to increase or decrease performance variability, and make applicant \underline{SD}_y levels either higher or lower than \underline{SD}_y among job incumbents. Third, estimating \underline{SD}_y on job incumbents encourages estimators to consider all of the incumbents they have had experience with, including incumbents with different tenure levels. If performance varies with tenure, then \underline{SD}_y estimated on the incumbents will reflect this tenure variability. However, this variability will not be present among cohorts of selectees because each cohort of hired applicants will have equal tenure throughout their employment. Thus, where job tenure and performance are related, an \underline{SD}_y estimate based on job incumbents may overestimate applicant \underline{SD}_y . Fourth, as noted earlier, virtually all utility analysis research groups employees with similar job titles to form the focus population. However, if task assignments or work environments differ within the same job the variability of performance may differ as well. \underline{SD}_y estimates based on incumbent populations may be inaccurate reflections of the actual \underline{SD}_y in the selection system (Bobko, et al., 1983 and Rich & Boudreau, 1987 discuss this issue).

Most authors who discuss this issue adopt the argument that incumbent-based \underline{SD}_y estimates are conservative due to restricted range. However, there is as yet no evidence regarding the possible biasing effects of variability due to tenure, different recruiting approaches, or different labor market conditions. Indeed, not

one study has compared $\frac{SD}{\bar{y}}$ estimates based on the applicant population to those based on the incumbent population.

Effects of Measurement Technique

Because $\frac{SD}{\bar{y}}$ was characterized as the Achilles' Heel of utility analysis by Cronbach and Gleser (1965) and because differences in $\frac{SD}{\bar{y}}$ can cause such large differences in total utility estimates, many have argued that it is important to develop better $\frac{SD}{\bar{y}}$ measures. Though variations on each theme are prevalent, it is possible to divide $\frac{SD}{\bar{y}}$ measurement methods into four categories: (1) Cost accounting, (2) Global estimation, (3) Individualized estimation, and (4) Proportional rules.

Cost accounting. These methods use accounting techniques to attach a value to units of performance or output for each individual, with the standard deviation of these individual performance values representing $\frac{SD}{\bar{y}}$ (e.g., Roche, 1961; Van Naersson, 1963; Schmidt & Hoffman, 1973; Lee & Booth, 1973). The difficulty and arbitrariness of cost accounting has frequently been cited as arguing in favor of simpler methods (e.g., Cascio, 1982; Cascio & Ramos, 1986; Hunter & Schmidt, 1982; Schmidt, et al., 1979).

Global estimation. These methods have experts estimate the total yearly dollar-valued performance at two, three, or four percentiles of an hypothetical performance distribution, with average differences between these percentile estimates representing $\frac{SD}{\bar{y}}$ (e.g., Bobko, et al., 1983; Bolda, 1985; Burke & Frederick, 1984, 1985; Cascio & Silbey, 1979; Eaton, Wing, & Lau, 1985; Eaton, Wing & Mitchell, 1985; Hunter & Schmidt, 1982; Mathieu & Leonard, 1986; Rich & Boudreau, 1987; Schmidt, et al., 1979; Schmidt, Mack

& Hunter, 1984; Weekley, et al., 1985; Wroten, 1984). Subjects often find the task difficult, refuse to do it, or produce high inter-rater variance (e.g., Bobko, et al., 1983; Mathieu & Leonard, 1987; Rich & Boudreau, 1987; Reilly & Smither, 1985). Providing a common 50th percentile anchor can reduce inter-rater variability as can consensus ratings (Burke & Frederick, 1984; Wroten, 1984) but this makes \underline{SD}_y dependent on the anchor (Bobko, et al., 1983; Schmidt, Mack & Hunter, 1984; Wroten, 1984), violating the statistical assumption of independence between the mean and the standard deviation.

Only limited evidence exists on the accuracy of global \underline{SD}_y estimation, and tests are usually based on arguably deficient objective performance measures. 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 \underline{SD}_y estimate based on the averaged difference between the 85th and 50th and the 50th and 15th percentiles was not significantly different from the actual sales distribution (although the percentile estimates were quite different). However, when respondents 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 than sales. Burke and Frederick (1984) also found \underline{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 \underline{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 \underline{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. Thus, the research is sparse and the results are mixed, providing little evidence that global \underline{SD}_y estimates reflect actual sales or productivity information.

Individualized estimation. This method translates some measurable characteristic of each individual in the sample (e.g., pay, sales activity, performance ratings) into dollars using a scaling factor such as average salary or average sales, with the standard deviation of these values representing \underline{SD}_y (e.g., Arnold, et al., 1982; Bobko, et al., 1983; Burke & Frederick, 1984; Cascio & Ramos, 1986; Dunnette, Rosse, Houston, Hough, Toquam, Lammlein, King, Bosshardt, & Keyes, 1982; Eulberg, O'Connor & Peters, 1985; Janz & Dunnette, 1974; Ledvinka, et al., 1983; Reilly & Smither, 1985). Three individualized estimation techniques have emerged.

First, Cascio (1982) and Cascio & Ramos (1986) used CREPID (Cascio-Ramos Estimate of Performance In Dollars). This method breaks a job into important "principal activities." Then, each activity is rated on four dimensions (time/frequency, level of difficulty, importance, and consequence of error), 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 employees' performance on each principal activity, using a 0 to 2 scale. To translate these ratings into dollars, they are multiplied by the dollar value assigned to that activity. After each employee has been assigned a dollar value for each activity, these activity values are summed to provide the total dollar value of yearly performance for that employee.

Janz & Dunnette (1977) proposed a second technique. Their approach also involves identifying critical job activities. However, rather than allocating salary to each activity based on its time/frequency, importance, etc., 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), by tracing the consequences of the various levels of effectiveness to determine their impact on activities to which costs and/or value can be attached.

A third approach to individualized estimation involves having experts simply assign dollar values to individual employees directly. Bobko, et al. (1983) used this method to derive an $\frac{SD}{y}$ estimate based on sales levels (sales volume times average insurance policy value), 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) 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.

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 judgments. Methods involving behavioral job analysis (e.g., CREPID or Janz-Dunnette) may be more understandable or credible to those familiar with the job, though absolutely no evidence exists on this issue. Still, 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 not supported by economic theory and clearly violated in organizations with tenure-based pay systems, pay systems based on rank, and many hourly-based pay systems. 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). Janz-Dunnette is based on the assumption that job behavior effects on costs and revenues can be accurately traced by managers. Individualized estimation methods are often more complex, costly and time consuming than other estimation methods, and as yet no evidence suggests whether they improve decisions.

Proportional rules. This measurement approach multiplies some available productivity-related variable (e.g., average wage, average sales, average productivity value) by a proportion to

estimate $\frac{SD}{Y}$ (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, 1986). Proportional rules emerged in part from observations concerning the relationship between $\frac{SD}{Y}$ estimates and average salary levels, and in part from the desire to provide a straightforward $\frac{SD}{Y}$ measurement method that could be used even when the global estimation procedure is not feasible. The method involves multiplying average salary in a job by some proportion (e.g., between 40% and 70%) to derive the $\frac{SD}{Y}$ estimate for the incumbent employee group.

Hunter and Schmidt (1982, pp. 257-258) reviewed empirical studies and compared their $\frac{SD}{Y}$ estimates to reported or derived average salary levels. They discovered that $\frac{SD}{Y}$ averaged about 16% of average salary. The authors also reviewed two of their own studies (using a global estimation procedure) and noted that $\frac{SD}{Y}$ 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 $\frac{SD}{Y}$ falls somewhere in the range of 40 to 70% [of average salary]" (p. 258). They also reviewed empirical data on productivity levels measured in units of output. Their review indicated that for non-piece-rate situations the average ratio of the standard deviation to the mean productivity was .185, while in piece-rate situations the average ratio was .150, and in uncertain compensation systems the average ratio was .215. They concluded 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 \underline{SD}_y as 40% of mean salary" (p. 412).

The proportional rules proposed by Schmidt and Hunter are intriguing because they suggest that simple \underline{SD}_y estimation may be quite feasible in virtually all situations. However, this simplification is obtained by assuming that average salary is indeed equal to about half the value of the average value of products "as sold", which may be violated by tenure-based pay systems, negotiated pay systems, labor market conditions such as unemployment, and internal labor markets (e.g., Becker, 1964). One must also assume that \underline{SD}_y is equal to about 20% of the average value of products "as sold", which was not the case in a number of the studies reviewed by Schmidt and Hunter (1983).

My review of utility studies through 1986 uncovered 17 \underline{SD}_y estimates below 40% of salary, 18 estimates within the 40% to 70% range, and 29 estimates above 70% of salary. No \underline{SD}_y estimates fell below 20% of mean output, 9 fell between 23% and 34%, and 27 fell above 34% (many of these were quite substantially above 34%, some as high as 100% or more). While these results support the conservatism of this decision rule, it should be noted that a detailed meta-analysis has not been conducted. Moreover, the results suggest that using the Hunter-Schmidt proportional rule may produce such conservative estimates that severely understated utility estimates and rejection of potentially useful HRM programs might result.

Summary. Existing research suggests that differences between \underline{SD}_y estimates using different methods are often less than 50% (and may be less than \$5,000 in many cases). However, it is tempting

to consider the fact that these differences may be multiplied by factors of hundreds or thousands in deriving the final total utility value. Even a small difference multiplied by such large values can imply vast total utility differences, tempting some to conclude that we need substantially more research on $\frac{SD}{y}$ measurement to whittle down such differences and provide more precise total utility estimates. However, carefully considering the role of uncertainty in selection utility tempers this conclusion.

The Role of Uncertainty and Risk in Utility Analysis

How is it that selection utility analysis research can simultaneously produce such clear evidence of program payoff (i.e., virtually every study showed positive and often quite large utility values) and such a raging debate on the proper measurement method for one utility parameter ($\frac{SD}{y}$)? One explanation is that although the expected utility values are quite high, there is also substantial uncertainty associated with these utility estimates, and that uncertainty stems largely from measurement error in $\frac{SD}{y}$. Properly investigating this issue requires changing the focus of utility analysis from attempting to develop the most accurate estimate of expected utility to attempting to estimate both the expected value and the distribution of values (Boudreau, 1984; Rich & Boudreau, 1987). It focuses attention away from measurement and toward uncertainty and risk in the decision situation.

Four Alternative Approaches for Estimating Uncertainty

Rich and Boudreau (1987) provided a conceptual framework for uncertainty in utility analysis and empirically applied of four alternative methods to account for uncertainty: (1) sensitivity analysis; (2) break-even analysis, (3) algebraic derivation of

utility value distributions; and (4) Monte Carlo simulation analysis.

Sensitivity analysis. Several utility analysis applications have addressed possible variability in utility parameters through sensitivity analysis (e.g., Boudreau, 1983a, 1983b; Boudreau & Berger, 1985a; Cascio & Silbey, 1979; Florin-Thuma & Boudreau, 1987; Schmidt, et al., 1979; Schmidt, et al., 1984). Sensitivity analysis varies each of the utility parameters from its lowest to its highest value, 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. 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., Schmidt, et al., 1979; Hunter & Schmidt, 1982; Schmidt, et al., 1984). If resulting utility values remain positive in spite of such conservatism, it is presumed that they will turn out to be positive in the actual application.

Though valuable, sensitivity analysis usually provides no information about the effects of simultaneous changes in several utility parameters (though Boudreau & Berger, 1985a and Boudreau, 1987 expressed the effects of simultaneous changes in utility parameters) and provides no information regarding the utility value distribution nor the probabilities associated with particular parameter value combinations (Hillier, 1963, p. 444). Setting

all parameters at their most conservative levels (a statistically unlikely event) risks incorrectly concluding that some programs will not pay off.

Break-even analysis. Boudreau (1984) proposed that a relatively simple 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 that such analysis was useful not only when analyzing a single program, but also when multiple alternatives are involved (with more expensive alternatives offering greater potential payoffs). Boudreau's approach is relatively simple and explicitly focuses on the decision context. Rather than advocating improved measurement in all situations, Boudreau proposed that one should first maximize the knowledge to be gained from existing information (usually the quantity of employees affected and the costs of the program) by estimating the critical values for the unknown parameters, and then determining whether further measurement effort is warranted. Because controversy surrounded the accuracy and validity of $\frac{SD}{y}$ estimates, Boudreau concentrated his analysis on that utility parameter, demonstrating that the break-even $\frac{SD}{y}$ values for the studies by Cascio and Silbey

(1979), and Schmidt, et al. (1979) were substantially lower than the expected \underline{SD}_y value they derived.

I applied break-even analysis to my more recent review of empirical utility studies. This not only verified the conclusion that selection program utility is uniformly high, but also shed some light on the \underline{SD}_y controversy. Without exception, the break-even \underline{SD}_y values fell below 60% of the estimated \underline{SD}_y value. In many cases, the value necessary to break even was less than 1% of the estimated value! In fact the break-even \underline{SD}_y value exceeded 20% of the estimated value in only 6 of the 41 analyses. The vast majority of utility analysis applications concluded that the more-valid selection device is worth the extra costs. Break-even analysis supports this conclusion, suggesting that it could probably have been reached without ever actually measuring \underline{SD}_y (or at least by measuring it in the simplest and most conservative manner possible). The break-even \underline{SD}_y values often fall several standard deviations below the expected value. Sometimes (e.g., Rich & Boudreau, 1987) the break-even \underline{SD}_y value falls below the lowest value estimated by any of the subjects. Recent research incorporating Boudreau's break-even analysis approach has reached similar conclusions (e.g., Burke & Frederick, 1985; Cascio & Ramos, 1986; Eaton, Wing and Lau, 1985; Florin-Thuma & Boudreau, 1987; Karren, NKomo, & Ramirez, 1985).

Algebraically deriving utility value variability. Recently, statistical formulas for the variance of products of random variables have been adapted to utility analysis. 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 (1986) to produce a formula for the standard error of utility values associated with a one-cohort selection utility model. 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.

Monte Carlo analysis of utility value variability. Monte Carlo analysis involves describing each utility model parameter in terms of its expected value and distribution shape. In each 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 the total utility value for that trial. Repeated application of this choosing and calculating procedure (using a computer) produces a sample of trials from which the distribution properties of the utility values can be derived. 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 (1987) applied Monte Carlo analysis (and compared it to each of the other three uncertainty estimation methods) using an application of the Programmer Aptitude Test (PAT) to select computer programmers in a mid-size computer manufacturer.

They discovered that all of the utility parameters were subject to some degree of uncertainty or variability over time. They also discovered that \underline{SD}_y variability heavily influenced the utility value distribution and that the distribution of \underline{SD}_y values was positively skewed as in other studies (Bobko, et al., 1983; Burke & Frederick, 1984; Schmidt, Mack & Hunter, 1984; Mathieu & Leonard, 1987). Rich and Boudreau's (1987) findings suggested that the distribution of utility values was greatly affected by the assumptions about the \underline{SD}_y distribution. They also found that the simulation suggested a greater amount of 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: positive payoff from the selection program was very likely.

Future Research Implications

Promising future research directions include integrating selection utility with employee separations and job changes within the organization, extending the payoff function to encompass consequences other than productivity, and studying how selection utility information affects decisions.

The External Employee Movement Utility Model

Boudreau and Berger (1985a, 1985b) suggested that there were important analogies between employee acquisitions and employee separations. Selection utility involves choosing a subset of employees to join the workforce from a pool of applicants. Retention utility involves "choosing" a subset of the previous-

period's incumbent workforce to remain with the organization (though retentions are more bilaterally chosen than acquisitions, the analogy for utility purposes is still correct). The utility of both acquisitions and retentions depends on the quantity of employees affected (i.e., the number hired and the number retained), on the quality of affected employees (i.e., the per-person, per-time period effects of selection strategies; and the per-person, per-time period effects of the retention pattern), and on the costs incurred to implement or accommodate the movements (i.e., selection device development/implementation costs and separation costs such as severance pay, relocation assistance, etc.). Boudreau & Berger (1985b, pp. 598-599) concluded that a utility analysis based only upon selection consequences risks not only producing deficient utility values, but also producing values that could lead to faulty decision making. Their results also suggested that selection utility models that ignore retention utility effects may substantially overstate utility values (when less-valuable employees are retained) or understate utility values (when more valuable employees are retained).

Omitting retention utility considerations may severely bias selection utility estimates. When improved selection causes the retention pattern to become less optimal or when the retention pattern is such that the value of improved selection is lost quickly, selection utility values based on simpler models may lead to incorrect decisions. This suggests that employee turnover research and employee selection research should be integrated, with both areas attending to the effects of the other. Moreover, it suggests that studies of employee separations (e.g., "turnover")

that focus only on the costs of separations or on 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 fact that they fail to consider the effects of those acquired to replace the separations. Thus, the external employee movement utility model provides the framework for integrating and expanding selection utility research.

Integrating Selection with Internal Staffing

Selection 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 can recognize. Conversely, if selection devices are targeted to skills exclusively applicable to a lower-level job, but employees routinely move to upper-level jobs using other skills, then maximizing lower-level job selection utility may simultaneously reduce utility for the upper-level job. These phenomena require an explicit framework integrating the consequences of internal and external employee movement and suggesting the variables likely to determine the utility of such movements--a utility model for internal and external employee movement.

Boudreau (1987) developed such a utility model. His utility model draws upon the analogies between internal and external employee movement. Specifically, Boudreau proposed 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. Boudreau's results suggested that decisions based solely on an external selection utility model (or even the external movement utility model including acquisitions and separations/retentions) may produce utility values leading to erroneous conclusions when internal movement patterns offset the apparent positive effects of external selection and retention, and vice versa. Future research should adopt a broader perspective by incorporating internal movement consequences into investigations of productivity-enhancement interventions.

Using Selection utility analysis Models to Examine Actual Decision Processes

While selection utility analysis results are often reported as if they will influence decisions, enhance credibility, and encourage a broader decision focus, existing research has ignored these phenomena. Studies are needed to examine whether the results of selection utility analysis actually affect managerial decisions, whether decision maker's reactions to selection utility results are affected by different estimation techniques, and whether selection utility models accurately reflect the concerns of decision makers. Florin-Thuma and Boudreau (in press) derived the utility of a performance feedback intervention in a small organization, asked decision makers to estimate the parameters of the utility model, and to develop their own decision model. Decision makers underestimated the magnitude of the performance problem and the intervention's effect. Factors considered by decision makers but

not included in the utility model (e.g., customer dissatisfaction) argued against the intervention. However, when dollar values were attached to these factors and when the decision maker's assumptions were incorporated into the model, the utility results still suggested substantial payoffs. Apparently, the decision makers had failed to implement the performance feedback intervention because they had simply never considered the performance problem serious enough to warrant systematic consideration. Utility analysis improved their awareness of the problem and their decision.

These results suggest research questions and methodologies to more fully explore the effects of utility analysis on decisions. Such research should draw on the substantial body of knowledge regarding irrationality in decision making (e.g., Kahneman & Tversky, 1972, 1973; March & Simon, 1958). Selection utility analysis models offer detailed frameworks for program analysis. They are normative descriptions of the factors that decision makers "should" consider in making selection decisions. However, Etzioni (1986) has suggested that rational decision making must be induced because it is contrary to natural inclinations. Selection utility analysis may provide such inducements, but we must first understand how actual decisions depart from selection utility analysis prescriptions, and focus our efforts to induce more rational decision making. Research linking selection utility analysis to actual decisions may discover how selection utility analysis models can be enhanced by better reflecting actual decision considerations.

Selection utility analysis offers vast research potential. Moreover, the results of such research are likely to have very important implications for the ways Human Resource managers (and

those who assist them) apply findings from industrial psychology and other social sciences. Future research should emphasize the decisions supported by utility analysis, should incorporate economic information into utility analysis, should adopt a broader and more integrative perspective regarding multiple interventions, and should attend more closely to the effects of utility analysis on managerial decisions. With attention to the research questions noted above, researchers and decision makers will produce decision tools that truly reflect a partnership between applied social science research and managerial decisions regarding human work behavior.

References

- Alexander, R. A., & Barrick, M. R. (1986). Estimating the standard error of projected dollar gains in utility analysis. Working paper, University of Akron.
- 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, G. (1964). Human capital. New York: National Bureau of Economic Research).
- Becker, B. (1985). Utility analysis of human resources programs: Some caveats. Unpublished manuscript.
- Bierman, H. Jr., Bonini, C. P., & Hausman, W. H. (1981). Quantitative analysis for business decisions. Homewood, Ill: Irwin.
- 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. (1985). Utility: A productivity planning tool. Human Resource Planning, 111-132.
- 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. (1984). Decision theory contributions to HRM research and practice. Industrial Relations, 23, 198-217.

- Boudreau, J. W. (1987). Utility analysis applied to internal and external employee movement: An integrated theoretical perspective. Ithaca, New York: Author.
- Boudreau, J. W. (in press, a). Utility analysis: A new perspective on human resource management decisions. In L. D. Dyer & G. Holder (Eds.) Human resource management: Evolving roles and responsibilities. Vol. 1, ASPA/BNA Handbook of Human Resource management. Washington, D.C.: Bureau of National Affairs.
- Boudreau, J. W. (in press, b). Utility analysis for human resource management decisions. In M. D. Dunnette (ed.), Handbook of industrial/organizational psychology (2nd ed.).
- 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. & 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. (1985). A comparison of economic
utility estimates for alternative rational SDy estimation
procedures. Paper presented at the 45th annual meeting of
the Academy of management. San Diego, CA.
- Cascio, W. F. (1982). Costing human resources: the financial
impact of behavior in organizations. 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.
- 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.

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

Dalton, D. R., Krackhardt, D. M., & Porter, L. W. (1981).

Functional turnover: An empirical assessment. Journal of Applied Psychology, 66, 716-721.

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.

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.
- 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.
- Florin-Thuma, B. C. & Boudreau, J. W. (in press). Performance feedback utility in a small organization: Effects on organizational outcomes and managerial decision processes. Personnel Psychology.
- 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.
- Hillier, F. S. (1963). The derivation of probabilistic information for the evaluation of risky investments. Management Science, 9, 443-457.
- 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: Pergamon Press.
- 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. & Jackson, G. B. (1982). Meta analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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. & Boudreau, J. W. (1987). The effects of variability and risk on selection utility analysis: An empirical 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.
- 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., 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., 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.
- Shepbeck, M. A. & Cohen, S. L. (1985). Put a dollar value on your

training programs. Training and Development Journal, November, 59-62.

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.

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

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.

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.