

2000

# Utility Analysis for Multiple Selection Devices and Multiple Outcomes

Michael C. Sturman

*Cornell University School of Hotel Administration, mcs5@cornell.edu*

Follow this and additional works at: <http://scholarship.sha.cornell.edu/articles>

 Part of the [Human Resources Management Commons](#)

---

## Recommended Citation

Sturman, M. C. (2001). *Utility Analysis for multiple selection devices and multiple outcomes*[Electronic version]. Retrieved [insert date], from Cornell University, School of Hotel Administration site: <http://scholarship.sha.cornell.edu/articles/923>

This Article or Chapter is brought to you for free and open access by the School of Hotel Administration Collection at The Scholarly Commons. It has been accepted for inclusion in Articles and Chapters by an authorized administrator of The Scholarly Commons. For more information, please contact [hlmdigital@cornell.edu](mailto:hlmdigital@cornell.edu).

---

# Utility Analysis for Multiple Selection Devices and Multiple Outcomes

## **Abstract**

Traditional utility analysis only calculates the value of a given selection procedure over random selection. This assumption is not only an inaccurate representation of staffing policy but also leads to overestimates of a device's value. This paper presents a more accurate method for computing the validity of a selection battery for when there are multiple selection devices and multiple criteria. Application of the method is illustrated using previous utility analysis work and an actual case of administrative assistants with eight predictors and nine criteria. A final example also is provided that includes these advancements as well as other researchers' advances in a combined utility model. Results reveal that accounting for multiple criteria and outcomes dramatically reduces the utility estimates of implementing new selection devices.

## **Keywords**

utility analysis, selection devices, outcomes, administrative assistants

## **Disciplines**

Human Resources Management

## **Comments**

### **Required Publisher Statement**

© Emerald. Final version published as: Sturman, M. C. (2001). Utility analysis for multiple selection devices and multiple outcomes. *Journal of Human Resource Costing & Accounting*, 6(2), 9-28. Reprinted with permission. All rights reserved.

## **Utility Analysis for Multiple Selection Devices and Multiple Outcomes**

Michael C. Sturman

Cornell University

Final version published in the *Journal of Human Resource Costing and Accounting* (2001), 6(2), 9-28

### **Author Note**

Michael C. Sturman, Department of Management: Operations, Human Resources, and Law  
School of Hotel Administration, Cornell University, Ithaca, NY 14853

Special thanks to John Boudreau, Tim Judge and Kevin Murphy for their helpful comments on earlier versions of this paper.

### **Abstract**

Traditional utility analysis only calculates the value of a given selection procedure over random selection. This assumption is not only an inaccurate representation of staffing policy but also leads to overestimates of a device's value. This paper presents a more accurate method for computing the validity of a selection battery for when there are multiple selection devices and multiple criteria. Application of the method is illustrated using previous utility analysis work and an actual case of administrative assistants with eight predictors and nine criteria. A final example also is provided that includes these advancements as well as other researchers' advances in a combined utility model. Results reveal that accounting for multiple criteria and outcomes dramatically reduces the utility estimates of implementing new selection devices.

## Introduction

Utility Analysis (UA) has evolved into a complex tool for estimating the value of human resource interventions. UA purportedly allows human resource decision makers to produce bottom line figures, which supposedly would add credibility to the perceived “soft” decisions commonly associated with human resources (Cascio, 2000). UA has overcome many of the limitations of the basic UA formula through the addition of algebraic modifications (c.f., Boudreau, 1983; Boudreau & Berger, 1985; De Corte, 1994; Hogarth & Einhorn, 1976; Murphy, 1986; Rynes & Boudreau, 1986). These algebraic modifications can have a substantial impact on the final results obtained from the UA calculation (Sturman, 2000). Yet UA has not had its intended effect on managerial decision making (Boudreau, 1996; Cascio, 1996). In part, this may be due to assumptions and inaccuracies still present in the UA formula. This paper addresses some of these assumptions by presenting a more accurate method of computing the validity coefficient.

This paper focuses on two assumptions of the basic UA formula: (1) its comparison of a new selection device to a policy of random selection (Boudreau, 1991; Cascio, 2000), and (2) the assumption of there existing a single predictor/criterion relationship described by the correlation coefficient  $r$  (Cascio, 2000). As I will describe below, both assumptions do not reflect the realities of the selection process, nor the current state of theory or practice with regard to the conceptualization of job performance. Using statistical methods originally designed to assess the correlation of multiple variables and a linear combination of outcomes, this paper presents a more accurate means of evaluation device validity to account for selection situations with multiple predictors and criteria. I then illustrate the use and implications of this adjustment with prior UAs and an actual case.

The vast majority of organizations use multiple selection devices when hiring, such as reference checks (97%) and interviews (81%) (Gatewood & Feild, 2001). Other commonly used selection devices include structured interviews, ability tests, work sample tests, personality tests and integrity tests (Gatewood & Feild, 2001; Heneman, Heneman, & Judge, 1997) The impact of including these existing

selection devices on utility estimates can be substantial: the existence of a previous selection device lowers the utility estimate by an average of 59% (Sturman, 2000). Thus, if an organization is considering the addition of a new selection device, the utility estimation procedure should account for the current selection battery.

In addition to accounting for multiple selection devices, performance theorists have also recognized the multidimensional nature of job performance. Many have identified performance as including the fulfilment of duties, tasks and roles specifically associated with a job (Borman & Motowidlo, 1993; Campbell, 1990; Murphy, 1989; Rotundo & Sackett, in press), often labelled task performance. In addition to task performance, considerable attention has been paid to non-task related behaviours that otherwise contribute positively to the organization. These behaviours may not be specifically recognized or rewarded as part of the job, but do contribute to the goals of the organization by enhancing its social and psychological environment (Rotundo & Sackett, in press). Such behaviours have been described elsewhere, and include the behaviours labelled organizational citizenship (Organ, 1988), prosocial organizational behaviour (Brief & Motowidlo, 1986), contextual performance (Borman & Motowidlo, 1993; Motowidlo & Van Scotter, 1994) and citizenship performance (Rotundo & Sackett, in press). Other theoretical dimensions of performance have also been identified, such as counterproductive behaviours that detract from the goals of an organization (Murphy, 1989; Robinson & Bennett, 1995; Rotundo & Sackett, in press).

Besides performance theory, applications of performance appraisal often use multiple criteria (Milkovich & Newman, 2002). Performance appraisals often capture multiple aspects of task performance, recorded through a number of different rating scales. Although these scales may not capture the theoretical dimensions of performance identified by researchers, practitioners may be interested in identifying and predicting performance based on an array of duties associated with a specific job.

Because current UA models only concern themselves with a single criterion, traditional UA estimates may be inaccurate or inappropriate for a given setting. At a minimum, the basic UA formula does not accurately represent the nature of performance in the workplace. Additionally, useful selection

devices may measure multiple dimensions of performance (e.g., quantity of work, quality of work, etc.), in addition to impacting other outcome variables (e.g., turnover, trainability, etc.). Thus, UA models should be constructed to account for multiple criteria.

Given that there are both multiple predictors and criteria that are relevant to most selection processes, it is unreasonable to assume that an organization is debating using a single test alone as a selection device or reverting to a policy of random hiring. It may also be an oversimplification to only consider a single criterion, such as a global rating of performance, when evaluating the financial impacts of a selection procedure. If an organization considers using an additional selection device, calculations of that device's utility must be based upon the gain achieved by incorporating the device into the current selection process and the total predictive power of that device for multiple criteria.

### **Computing Validity for Multiple Predictors**

To expand the basic utility model to account for the above limitations, the incremental validity of the new selection device(s), the multidimensionality of performance, and the relationship between these two aspects must be determined. The basic UA formula, which emerged from work by Brogden (1949) and Cronbach and Glesser (1965), employs a variable,  $r$ , representing the correlation between a single predictor,  $p$ , and a single criterion,  $c$ . By expanding the analysis to a set of predictors and criteria, represented by the vectors  $p$  and  $c$  the computations needed to calculate a single correlation value become more complex. The analysis becomes even more complex given that the practitioner would likely want the predictors weighted to yield maximum potential predictive power, but the performance criteria would be determined independent of the utility analysis, such as from a job analysis.

The new correlation coefficient would therefore be based on the correlation between two linear combinations:  $u$ , which equals the scores on a set of predictors,  $p$ , times a set of (to be computed) weights,  $a$ ; and,  $v$ , which equals the criteria scores,  $c$ , times a set of weights,  $b$  (determined by the practitioner).

$$u = a \times p$$

$$v = b \times u$$

The set of weights,  $a$ , will be mathematically derived, but the set of criteria weights,  $b$ , will be predetermined (i.e., based on the relative importance of each of the performance dimensions, as determined by a manager and/or through a job analysis). The problem thus becomes calculating the correlation between the linear combinations  $u$  and  $v$ . The solution to the problem is based on the principles behind calculating the correlation of sums (Nunnally & Bernstein, 1994). One might also note that this problem is similar to the calculation of canonical correlations (Johnson & Wichem, 1998); however, because the set of weights,  $b$ , are not being modified to yield a maximal correlation, the resulting solution will differ. Figure 1 shows the formulae and necessary covariance matrix for calculating the correlation value. The derivation is shown in the Appendix.

Calculation of the correlation between the set of predictors and the set of criteria must be accompanied by other changes to be logically implemented. When considering adding a new device to a given selection process, the incremental validity of the new selection device must be calculated. This means that the device cannot be evaluated relative to its improvement over random hiring. A new device may add value to the existing selection process by being able to improve prediction of the criteria. To properly utilize the UA formula, the validity of the new selection process as a whole must be computed and then compared to the previous method(s). The value of  $r$  that must then be used in the utility formula equals the difference between the new correlation and the old correlation.

Yet knowing only the validity of potential selection devices is insufficient to make an informed decision regarding their implementation. The base rate, selection ratio, payoff function, device cost and financial considerations all have an impact when considering the utility of a new selection device (Boudreau, 1991a; Cascio, 2000).

To illustrate the use of a more accurate method for calculating selection device validity this paper will present two sets of examples computing the expected utility of new selection devices. The first example applies the advances described here to previous work on UA. The second example is an actual implementation of the method in a UA for the selection of administrative assistants. The examples also

show how the more accurate validity coefficient estimation can yield results quite different from the traditional model.

---

Figure 1 Here

---

### **Application of the Revised Validity Estimate To Previous UA Research**

This section of the paper reapplies the more accurate validity estimation procedure to previous research where the simplest UA models were used. To do this, this paper relies on the review of UA literature by Boudreau (1991a). This review included an appendix containing summaries of 42 UAs, accumulated from the works of 19 authors. Of these, there are seven authors who performed analyses that involved multiple devices. In each study, the utility of a new selection device was compared to the utility of an interview. However, these studies only investigated the value of the new selection device over a policy of replacing the interview and did not provide information of the incremental utility achieved by implementing the new device in conjunction with the interview. Thus, the utility model derived above is used to obtain a more realistic estimation of the utility of implementing the new selection devices.

Before presenting the analysis, though, a special note needs to be made regarding the lack of some critical information. First, these studies only look at a single general criterion of performance. Therefore, although this paper proposes to examine performance as a vector of criteria, these examples will be limited to the general performance effect used by the seven authors (an example using multiple criteria will be presented later in this paper). Second, information on the correlation between the interview and the new selection device was not provided by the studies. Thus, the analyses are performed with assumed intercorrelations. In order to fully illustrate the implications of the utility modification, the analyses are performed for intercorrelation values of 0, 0.25 and 0.50. This method of performing UA

using different reasonable values of a specific variable has been used by other researchers to estimate the value of selection devices or to illustrate UA techniques (e.g., Cascio & Silbey, 1979; Murphy, 1986; Rich & Boudreau, 1987; Sturman, 2000). The values chosen here represent a broad range of realistic possible effects. Information on selection devices, validities and utilities for each study is shown in Table 1.

The results of the new UAs lead to a number of conclusions. Most notably, increases in validity from combining two devices were always less than the sum of the validities of the two selection devices analysed separately. In fact, the utility of adding the new device over random selection overestimates the value of adding the new advice to the interview by an average of 63%. This is similar to the mean reduction of 53% shown by Sturman (2000). This shows that simply adding or subtracting validity or total utility scores of selection devices computed separately overestimates the value of combining the selection devices.

The results of these validity calculations strongly suggest that the traditional UA model overestimates the utility of multiple devices. Despite increases in the test battery validity, the costs associated with administering two tests made the administration of both devices, on average, little better than implementing only the better of the two selection devices. Dollar values represented a gain of less than 5% or a loss for 18 of the 21 calculations.

It should be noted, however, that as mentioned earlier, the most likely scenario is that a new device is being considered in addition to an interview (given that the vast majority of companies use interviews). Consistent with the traditional utility model, the results here show positive returns for adding the new selection device for all of the examples. Thus, decision makers may still come to the same conclusions: implement the new selection device. Yet the methods employed here should provide more accurate information that may help human resource decision makers form a better idea of what dollar gain realistically can be expected from adding a new selection method.

---

Table 1 Here

---

## **Application of the Revised Validity Estimate To A Selection Program**

While the demonstration of this revised approach to validity estimating with previous UA research illustrates the general differences between the improved parameter calculation and the traditional method, it would be useful to illustrate the technique with an actual example where there are multiple predictors, multiple criteria and all the necessary intercorrelation data.

To make this illustration, I use a case of hiring administrative assistants at a specific, large, midwestern plant of a Fortune 500 company. Data were collected on the validity of eight selection tests. The utility of adding one or both of two tests-the Test of Learning Ability (Richardson, Bellows, Henry & Co., 1989) and the Wonderlic Personnel Test (Wonderlic & Associates, 1983)-are evaluated. These tests are considered as potential additions to a set of 6 office skills tests: the R. D. Craig Typing Test (R. D. Craig Assessments, 1990), SRA Checking test, SRA Coding test, SRA Filing test, SRA Grammar test and SRA Punctuation test (SRA/London House Office Skills Tests, 1977).

This decision situation reflects a realistic problem that a human resource manager may face. Many companies use skill performance tests or work samples to make selection decisions, whereas less than a third employ mental ability tests (Bureau of National Affairs, 1988; Gatewood & Field, 2001); however, the use of mental ability tests is increasing (Gatewood & Field, 2001). Thus, it is quite plausible that a company may be considering adding a mental ability test to its current battery of work sample tests.

Administrative assistants in this sample are evaluated on nine dimensions: administrative skills, ability to handle stress, adaptability to change, customer service, attention to detail, writing ability, computer skills, numerical ability and proofreading ability. Additionally, each administrative assistant received an overall performance rating. A validity study was conducted on 296 current administrative assistants at the company by administering the current and new tests to the workers. The resulting correlation matrix, including the specific tests and criteria, are shown in Table 2.

A utility analysis, using the techniques prescribed in this paper for computing  $r$  is compared to a traditional utility analysis. The other aspects of the utility equation—the number of hires ( $N$ ), the dollar value of someone who performs a standard deviation above average ( $SDy$ ), the average performance on the selection device of those who are hired ( $Zx$ ), the initial cost of implementing the selection device ( $Ci$ ) and the cost per application ( $Ca$ )—are the same for both calculations. The number of applicants and number selected are based on estimates from the company that are representative of past hiring situations.

The estimated number of new administrative assistants needed each year is 20. The average tenure for an administrative assistant is six years. For each position, roughly 30 applicants apply (600 applicants per year). Thus, over the course of a year, the selection ratio averages .033, yielding an average  $Z\text{-bar } X$  score of 2.23. At the time of this study, the Wonderlic Personnel Test (Wonderlic, 1983) had an initial cost of \$130, which included 50 tests, a manual and a score key. One hundred additional tests could be purchased for \$158. Also, at the time of this study, the Test of Learning Ability (Richardson, Bellows, Henry & Co., 1989) cost \$55 for 25 tests, and a single answer key for \$7.50 must also be purchased. Both the Wonderlic Personnel Test and the Test of Learning Ability are multiple-choice tests and have 12-minute time limits. The prospective tests would be administered with the other 6 tests, given to the group of 30 applicants in a single sitting. An additional quarter of a person hour (at \$30/hour) was estimated as being necessary to administer each new test. Scoring the test was estimated to take one minute per applicant (again, at a rate of \$30/ hour).

Finally,  $SDy$  was based on the 40% rule (Schmidt & Hunter, 1983). Although there are other methods of estimating  $SDy$  (Cascio, 2000), the 40% rule seems adequate for this study. Thus, based on a reported starting pay of administrative assistants at the company,  $SDy$  equals roughly \$6,000.

---

Table 2 Here

---

---

Table 3 Here

---

## Utility Evaluations

To demonstrate the implications of the methods described in this paper, a number of sets of UAs are performed. For all six sets (to be described below), three separate calculations were made. First, the total utility in year one is given. Second, the utility in year one, after accounting for a 45% tax rate (Boudreau, 1983), is calculated. Third, the utility given a 10- year forecast was computed as a present value (Boudreau, 1983) and given a turnover rate of 20/296, which employs the employee flows model of Boudreau and Berger (1985).

The first UA set is a conventional analysis, evaluating the utility of the two tests over a policy of random selection. Then, a slightly more sophisticated analysis is conducted in that the combined  $r$  of the two selection devices over random selection is used in a basic UA. In both of these cases, the correlation between the selection devices and the overall measure of performance is used.

The second set of analyses involves determining the incremental validity of the two selection devices over the current battery of six tests. This case, though, still employs the overall measure of performance as the criterion for the correlations.

The third set of UAs uses all the methods proposed in this paper. The utility of the two selection devices, both separately and in conjunction with each other, are evaluated in addition to the current battery of six tests, and the individual facets of performance are used to weight the tests. Additionally, examples are given for different weighting values: (1) all the measures of performance weighted equally; (2) a fast paced administrative aid job; and (3) a job with inverse weights of the previous example.

A fourth set UAs is performed which combines the advances of this paper with other researchers' work refining the utility model. The basic example remains the same as the second case of the third set of analysis (i.e., the high-paced work environment), but the UA also includes corrections for non-top-down hiring (Murphy, 1986) and a probationary period for new employees (De Corte, 1994). A summary of the utility estimates for each of the examples is given in Table 3.

### Utility of Devices over Random Hiring

The conventional UA compares each of the two selection devices to a policy of random hiring. Both the Test of Learning Ability and the Wonderlic Personality Test are correlated .29 with the overall measure of performance.

Using the basic UA formula (Boudreau, 1991a; Cascio, 2000),

$$\Delta U = N_h * Z_x * r * SD_y - [C_i + (C_a * N_a)]$$

$\Delta U$	= Utility change from selection device
$N_h$	= Number of people to be hired
$Z_x$	= Average Z-score of the predictor of hired employees
$R$	= Correlation between the predictor and criterion
$SD_y$	= Dollar value of a standard deviation in the criterion
$C_i$	= Cost of acquiring or setting up the test
$C_a$	= Cost of administering the test to a single applicant
$N_a$	= Number of applicants

the utility of the Test of Learning Ability over random selection is \$75,961, and the utility of the Wonderlic Personnel Test is \$76,290. This analysis, though, does not give an estimate of the utility of using both devices. For the reasons discussed above, simply adding the two utility values together would likely highly overestimate the total utility. However, by first calculating the  $r$  of both selection devices using available statistical methods (e.g., Nunnally & Bernstein, 1994), the combined  $r$  is .38. The rest of the values in the basic utility formula stay the same, except for the costs, which equal the costs associated with the Test of Learning Ability plus the costs of the Wonderlic Personnel Test. Thus, the utility in the first year of using both devices, over a policy of random selection, is \$98,732. These two analyses yield large estimates for the utility of the selection devices.

Note that it is possible that the above analyses are inappropriately large because of a number of UA's implicit assumptions. Specifically, accounting for economic factors-like taxes and a discount rate (Boudreau, 1983), and employee turnover-may lead to inappropriately large estimates (Sturman, 2000). Values were chosen for these factors that reflected realistic estimates. The turnover rate (20/296) was

based on company data. A tax rate (45%) was chosen to be comparable to past UA calculations (e.g., Burke & Frederick, 1986; Mathieu & Leonard, 1987; Rich & Boudreau, 1987). A discount rate (10%) was set to reflect a realistic return for an alternative investment. I examined financial measures that reflect alternative investments (e.g., cost of equity, long-term corporate bond rates, as reported in the *Wall Street Journal*). Values varied from 5% to 15%, so a value of 10% seemed a reasonable estimate to use in this case for illustrative purposes.

Yet even after modifying the UA estimates by accounting for the turnover rate, discount rate and tax rate, the estimated utility values are still very large: the utility of using the Test of Learning Ability is \$1,106,810; the utility of the Wonderlic Personnel Test is \$1,108,169; and the utility of using both tests becomes \$1,447,468. This analysis demonstrates that the utility of using two devices is significantly different from the sum of using both devices independently (i.e., \$1.4 million vs. 2.2 million).

### **Utility of Selection Devices over Current Battery**

I now begin to illustrate the application of this paper's methods to the administrative assistant selection problem. Before being able to evaluate the utility of the new selection devices, the validity of the current battery of selection devices must be calculated. Using the overall measure of performance as the criterion, the validity of the current selection tests is .33. When the validity of adding any device to the selection battery is computed, the validity of the current selection method will be subtracted from the validity of the potential new (larger) package.

The validity of the selection battery with the current devices and the addition of the Test of Learning Validity is .34. Thus, the incremental validity of the test of Learning Ability is .01. Similarly calculated, the incremental validity of the Wonderlic Personnel Test is .05. The validity of the selection battery including all 8 devices is .40. Thus, the incremental validity of adding both new tests is .06.

With the incremental rs calculated, determining the utility of the new devices follows the basic methods of utility analysis. For the first year, the utility of the Test of Learning Ability is \$1,034, for the

Wonderlic Personnel Test \$12,066, and for the combination, \$15,776. Accounting for value over 10 years, and including a 10% discount rate and 45% tax rate, the values for the Test of Learning Ability, Wonderlic Personnel Test and the combination of the two become \$32,293, \$187,155 and \$257,824, respectively. These values are significantly less than the estimates of utility devices over random selection.

### **Utility of Selection Devices over Current Battery and for Multiple Criteria**

The third set of utility equations involves evaluating the utility of selection devices for multiple predictors and multiple criteria. Three examples will be used to illustrate the results. Normally, the criteria would be weighted on the basis of job analysis results. The first example will be for a situation where each of the criteria (administrative skills, ability to handle stress, adaptability to change, customer service, attention to detail, writing ability, computer skills, numerical ability and proofreading ability) is weighted equally. Given the methods described in this paper for maximizing the predictability of multiple selection devices for multiple criteria, it is possible to determine the validity of the current battery of tests ( $r = .38$ ), the change in  $r$  by adding the Test of Learning Ability (change in  $r = .00$ ), the change in  $r$  by adding the Wonderlic Personality Test (change in  $r = .03$ ), and the change in  $r$  by adding both devices (change in  $r = .04$ ).

Once again, given the values for  $r$ , calculating the utility is relatively simple. The utility of the Test of Learning Ability is -\$ 1,642; the utility for the Wonderlic Personnel Test is \$6,714; and the utility for adding both devices is \$7,748. Accounting for value over time, the discount rate and the tax rate, the utilities are -\$6,082, \$110,404 and \$142,697, respectively.

The second and third examples demonstrate some of the flexibility in this method of utility calculation. Given the same data and validity values, it is possible to re-weight the criteria in accordance with job analysis data. The second example reflects a need for administrative assistants in a faster-paced, less organized office. The criteria are weighted as follows. The most important skills for the position are

ability to handle stress, adaptability to change and attention to detail. These skills are twice as important as the next set of skills: computer skills and numerical skills. Finally, these two skills are twice as important as the remaining set of skills: administrative skills, customer service, written communication and proofreading. These descriptions can be turned into numerical weights (four for the most important skills, two for the next set and one for the final set). Using these weights and the equations derived in this paper, the correlation for the current test battery is .37, the incremental validity of the Test of Learning Ability is .003, the incremental validity of the Wonderlic Personnel Test is .02 and the incremental validity of adding both new tests is .03.

Given these  $r$ s, the utility can be easily calculated. The utility for implementing the Test of Learning Ability in the first year is -\$840; for the Wonderlic Personnel Test, \$4,038; and for using both tests, \$5,072. The utility of the tests over time, and accounting for the discount rate (10%) and tax rate (45%), are \$5,431, \$72,028 and \$104,321, respectively. Note that despite the utility values in the first year, the Test of Learning Ability yields positive utility and the combination of both tests yields the highest utility over time.

The third example utilizes weights that are the reciprocal of those in the second example. In other words, administrative skills, customer service, written communication and proofreading are weighted the highest (weight = 4); computer skills and numerical skills are weighted in the middle (weight = 2); and ability to handle stress, adaptability to change and attention to detail are weighted as needed but least important (weight = 1). For these weights, the correlation coefficients change somewhat. The multiple  $r$  of the current selection battery is .376. The incremental validity of the Test of Learning Ability is .003; the incremental validity of the Wonderlic Personality Test is .04; and the incremental validity of adding both devices is .04.

Once again, given these  $r$ s, the utility can be calculated. The utility for implementing the Test of Learning Ability in the first year is -\$840; for the Wonderlic Personnel Test, \$9,390; and for using both tests, \$7,748. The utilities of the three options over time, and accounting for the discount rate (10%) and tax rate (45%), are \$5,431, \$148,779, and \$142,697, respectively. In these calculations the Test of

Learning Ability does not add value over the combination of the current selection battery and the Wonderlic Personnel Test.

### **Comprehensive Model**

So far, this paper has concentrated on examples of the methods proposed here to make the utility estimate more accurate. It is important to note, however, that many other proposed algebraic changes to the UA model also have substantial effects of the magnitude of estimates (Sturman, 2000). Although the prior illustrations have not included these methods for the sake of simplifying the examples, these adjustments are included here to demonstrate how the total UA model might work.

Specifically, two advances of the utility model are now included: one, where the assumption of top-down hiring is relaxed (Hogarth & Einhorn, 1976; Murphy, 1986); and two, where the company can dismiss those employees who are later judged to perform at an unacceptable level (De Corte, 1994). Note that economic variables (Boudreau, 1983) and employee flows (Boudreau & Berger, 1985) have already been incorporated.

When job candidates choose to reject a job offer, the employer must make an offer to employees that the selection devices predicted to be less qualified. This results in a decrease in the average predicted score of the new employees, or as expressed in standardized units, a decrease in  $Z_x$ . The logic and derivation of the effect of this is provided in Murphy (1986). It should be noted, though, that an assumption needs to be made regarding the relationship between the quality of applicants and the probability of their accepting offers. Murphy (1986) provides three cases: where jobs are rejected at random, where jobs are rejected by the top applicants and where the probability of job acceptance is negatively correlated with quality. We agree with Murphy (1986) and others (Hogarth & Einhorn, 1976) that the most reasonable assumption is a negative correlation between ability and probability to accept offers. While the exact value of this correlation is unknown for our sample, I will use the value of  $-.20$ , because it both seems reasonable and because Murphy (1986) used it in his example.

It is also necessary to know the proportion of applicants rejecting offers (Murphy, 1986). Again, there are no data available for our sample; however, Murphy (1986) did cite a college recruiting report indicating that less than 65% of the job offers made in technical and engineering fields are accepted; and for non-technical areas, the acceptance rate is less than 75%. Thus, for our clerical sample, assuming an acceptance rate of 70% seems reasonable. Murphy (1986) provides methods for estimating the average score of those hired (in standardized units, i.e.,  $Z_x$ ) in the situation where a correlation exists between ability and predicted performance.

Another refinement to the utility model involves determining the effect of providing a one-year probationary period for new hires: after one year, those who do not perform adequately are dismissed. Adding this refinement yields a basic change to the utility model. The utility equation must estimate the performance difference between initially hired employees and employees who survive the probationary period (De Corte, 1994). Because poor performing employees are dismissed after one year, the average performance of employees is higher in later years than it is in the first year.

The specific changes involved in the utility model are explained in detail by De Corte (1994). It should be noted, though, that when these changes are included with the changes proposed in this paper, the method for computing the value of additional selection devices must be altered somewhat. Because of  $Z_x$ , changes between the first and latter years, and because the extent of this change depends on the validity of the selection process, simply finding the change in  $r$  and inserting this value into a single utility equation will not work. This can be easily remedied, though, by computing the utility of each selection alternative independently and then subtracting the appropriate values to yield the change in utility.

Based on De Corte's (1994) methods, it is necessary to determine a cutoff score that will be used as the performance threshold for probationary employees. For the illustration, I will assume that performance must be no less than one standard deviation below average. Those employees who score below this threshold would be dismissed after one year. Other values are certainly plausible, but for the sake of illustration the above cutoff value will be used. Because of this cutoff,  $Z_y$  (not  $Z_x$ ) will be 0.29 higher for those who survive the probationary period. Thus, in year two and beyond,  $Z_y$  will be 0.29 for

the baseline group, 0.30 when using the Test of Learning Ability, 0.36 when using the Wonderlic Personnel Test and 0.37 when using both new selection devices.

The utility of the three options over 10 years, and accounting for the discount rate (10%) and tax rate (45%), are \$3,144, \$109,064, and \$118,358, respectively. Granted, the inclusion of all the algebraic adjustments makes the computation of utility significantly more complex. Nonetheless, this full model represents perhaps the most accurate utility computation possible given the available data. Practitioners employing this technique could further refine the model by using actual data from their companies rather than the assumed values included here.

### **Discussion and Implications**

Utility Analysis for multiple methods and multiple criteria introduced in this study yield significantly different results than would be obtained had simple UA procedures been followed. Using incremental validity instead of assuming a policy of random selection had the largest impact on estimated utility scores. Results of varying the criteria weighting schemes did not yield notably different results, but this may be attributable to the subcriteria being related. If less related criteria were examined, such as absenteeism or turnover, the effects could have been larger. Additionally, these results still more accurately reflect the true utility of selection devices and provide managers with a means to select people for specific positions within a more general type of job.

It should be noted that when devices are considered over a policy of random selection, the utility of using multiple devices almost invariably would be less than the sum of using each device individually. However, when the devices are considered over a policy that includes a number of selection devices, the opposite may be true. This occurs because the intercorrelation between the new devices may add explanatory power, even if the simple correlation between the device and the criteria is not large. A good example of this is presented by Nunnally and Bernstein (1994). In their example,

$r_{y1}=.60$ ,  $r_{y2}=.00$  and  $r_{12}=.50$ . Given these values, the validity of using both devices is .69, which is greater than the sum of  $r_{y1}$  and  $r_{y2}$ .

This paper was intended to apply existing statistical knowledge on the correlation between multiple predictors and multiple outcomes to help UA more accurately measure the return of human resource interventions. It is not the purpose of this paper to draw conclusions regarding the general utility of specific selection methods. In fact, we would argue that the calculated results of this study have very limited generalizability to other situations where the Test of Learning Ability or the Wonderlic Personnel Test is being considered. Indeed, this paper contends that the utility of new selection devices depends on the current firm-specific practices. What are generalizable are the methods for calculating utility. These methods should yield a more accurate estimate of the dollar return of a proposed selection device.

The results do allow some generalizations regarding the implementation of selection devices and the implications of the proposed changes for UA implementation. Selection devices should not be considered in isolation, but rather as part of an overall battery of selection devices. A highly valid selection device may not necessarily add utility to the selection process, even if the cost is relatively low. This is because the new device may not add sufficient information to the decision model. Conversely, a selection device with a low validity, even a measure with a simple correlation with a criterion of 0.00, may add substantial validity to the selection process because of the intercorrelations with the current selection devices. The techniques exist with which to compute this incremental validity. Computing utilities without these modifications, or implementing a utility study without collecting this necessary information, will probably lead to gross overestimates of the value of the selection process.

This paper shows that the practice of traditional UA of comparing a new selection device to a policy of random selection leads to overestimated utility results, and thus this paper paints a somewhat more conservative picture of the value of a single selection method. However, the results do not suggest that selection devices are not valuable. Indeed, the majority of the examples (see Tables 1 and 3) have

positive returns for adding new selection devices. Rather, this paper shows that estimates generated using the basic utility model are too large.

### **Application of the Improved Computation Procedures For Decision Makers**

With the apparent lack of use of UA by managers, and with the added complexity of this paper's procedure for estimating selection device validity, it may seem less likely that managers will take advantage of UA procedures. However, it is possible to encourage use of this model by reducing the complexity of applying it through use of a computer program.

Obviously, the technology for the calculations exists. All the calculations for this paper were performed on a spreadsheet program or with a matrix calculation program. Although we do not believe every manager will customize applications (or learn matrix algebra) to perform the needed calculations, the tools do exist to begin such a programming task. Additionally, a well-designed user interface should help facilitate the use of UA. Through a computer program, all the necessary information could be collected in a format that is easy to understand, reasonable default values could be included for when there is missing data and the calculations would be performed for the user. Thus, a manager could use this model without actually having to know the extent of its complexity.

Some research suggests that UA, in its current form, may have a limited impact on managerial decision making (Latham & Whyte, 1994; Whyte & Latham, 1997). While one of the reasons for the limited impact may be the unrealistic nature of the assumptions underlying UA (Cascio, 1992), another reason may lie in its complexity (Latham & Whyte, 1994). Thus, while the revisions of the UA formula presented in this paper make the algorithm more realistic, they also make it more complex. As a result, without the introduction of useful computerized decision aids, it is not clear that any improvements to the basic formula, or even the basic utility model itself, will ultimately result in widespread use.

## **Future Research**

Research on selection devices needs to pay more attention to the intercorrelation between the new device and existing devices. Although there are a myriad of studies showing the correlation between single selection devices (e.g., biodata, personality tests, cognitive ability tests and honesty tests) and performance, this paper demonstrates that this simple correlation is not sufficient information to yield an accurate utility estimate. Utility researchers need to know the variance explained beyond that of current methods: the assumption of random selection is simply not valid.

Future research could also start where this paper ends. The current paper involves determining if new selection devices add utility to a current battery when all the tests are administered simultaneously. However, with more expensive selection devices (e.g., assessment centres, work samples and interviews), an employer may wish to limit the number of applicants who are screened through the device. Indeed, companies frequently use second interviews after screening an initial set of applicants. With current methods, computing the optimal selection strategy, in terms of predictor cutoffs and multiple batteries, would involve extensive guesswork or trial-and-error permutations of current and potential procedures. It would be valuable for future research to investigate the combination of optimization strategies for multiple devices to help human resource decision makers effectively develop a multi-stage selection procedure.

Finally, researchers should also pay more heed to the actual use of UA. Research is needed into the actual decision-making processes used by managers and how UA can affect such decisions. In both studies by Latham and Whyte (1994; Whyte and Latham, 1997), the subjects had no prior knowledge of UA. While this is not unrealistic, it demonstrates the need to improve managerial education on the techniques of human resource cost benefit analysis.

Overall, research on UA has much room to grow. Research is needed to increase the accuracy of the models, to make the models easier to use and to determine the effect of UA on decision makers.

Although the third objective perhaps represents the ultimate goal of UA research, advances in the first two areas are needed before the third goal is even feasible.

Predictors =  $[p_1, p_2, \dots, p_n]$

Predictor Weights:  $a = [w_{p1}, w_{p2}, \dots, w_{pn}]$  to be maximized

Criteria =  $[c_1, c_2, \dots, c_m]$

Criteria Weights:  $b = [w_{c1}, w_{c2}, \dots, w_{cm}]$  predetermined by company or practitioner

		Predictors				Criteria			
		1	2	...	n	1	2	...	m
Predictors	1	$\sum_{11}$				$\sum_{21}$			
	2								
	...								
	n								
Criteria	1	$\sum_{12}$				$\sum_{22}$			
	2								
	...								
	n								

$$r_{xy} = \frac{a' \sum_{12} b}{\sqrt{a' \sum_{11} a} \sqrt{b' \sum_{22} b}}$$

$$a = \frac{\sum_{11}^{-1} \sum_{12} b}{\sqrt{b' \sum_{22} b}}$$

**Figure 1.** Needed information and matrix calculations.

**Table 1.** Validity and utility of multiple selection devices.

Study/ Devices Used	Validity				Utility			
	Device 1 (Interview)	Inter- Device 2	Device 1 correlation	Combined	Adding (Interview)	Device 2	Combined	Device 2
Cascio & Silbey (1979) (Assessment Centre)	0.25	0.35	0.00	0.43	\$350,357	\$504,211	\$575,733	\$225,376
	0.25	0.35	0.25	0.39	\$350,357	\$504,211	\$509,476	\$159,119
	0.25	0.35	0.50	0.36	\$350,357	\$504,211	\$459,783	\$109,426
Schmidt, Hunter, McKenzie & Muldrow (1979) (PAT Test)	0.14	0.76	0.00	0.76	\$6,849,022	\$38,755,422	\$38,911,801	\$32,062,779
	0.14	0.76	0.25	0.76	\$6,849,022	\$38,755,422	\$38,396,983	\$31,547,961
	0.14	0.76	0.50	0.81	\$6,849,022	\$38,755,422	\$40,971,076	\$34,122,054
Ledvinka, Simonet, Neiner & Kruse (1983) (JEPS Test)	0.14	0.36	0.00	0.39	\$14,881	\$37,162	\$40,351	\$25,470
	0.14	0.36	0.25	0.36	\$14,881	\$37,162	\$37,162	\$22,281
	0.14	0.36	0.50	0.36	\$14,881	\$37,162	\$37,162	\$22,281
Schmidt, Mack & Hunter (1984) (PACE Test)	0.14	0.51	0.00	0.53	\$644,384	\$2,960,542	\$2,853,740	\$2,209,356
	0.14	0.51	0.25	0.51	\$644,384	\$2,960,542	\$2,728,542	\$2,084,158
	0.14	0.51	0.50	0.53	\$644,384	\$2,960,542	\$2,853,740	\$2,209,356
Burke & Frederick (1986) (Assessment Centre)	0.16	0.59	0.00	0.61	\$42,098	\$115,727	\$98,432	\$56,334
	0.16	0.59	0.25	0.59	\$42,098	\$115,727	\$89,980	\$47,882
	0.16	0.59	0.50	0.61	\$42,098	\$115,727	\$98,432	\$56,334
Cascio & Ramos (1986) (Assessment Centre)	0.14	0.39	0.00	0.41	\$7,335,605	\$20,830,312	\$21,101,212	\$13,765,607
	0.14	0.39	0.25	0.39	\$7,335,605	\$20,830,312	\$19,903,804	\$12,568,199
	0.14	0.39	0.50	0.39	\$7,335,605	\$20,830,312	\$19,903,804	\$12,568,199
Rich & Boudreau (1987) (PAT Test)	0.14	0.73	0.00	0.74	\$431,744	\$3,198,258	\$3,024,175	\$2,592,431
	0.14	0.73	0.25	0.73	\$431,744	\$3,198,258	\$2,972,715	\$2,540,971
	0.14	0.73	0.50	0.77	\$431,744	\$3,198,258	\$3,165,209	\$2,733,465

**Table 2.** Correlations of predictors and criteria.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
<b>Predictors</b>																	
1. R.D. Craig Typing Test	-																
2. SRA Checking test	.41**	-															
3. SRA Coding test	.33**	.53**	-														
4. SRA filing test	.55**	.53**	.51**	-													
5. SRA Grammar test	.30**	.42**	.25**	.38**	-												
6. SRA Punctuation test	.28**	.42**	.24**	.39**	.60**	-											
7. Test of Learning Ability	.49**	.36**	.28**	.43**	.55**	.53**	-										
8. Wonderlic Personnel Test	.11	.35**	.13*	.19**	.41**	.41**	.17**	-									
<b>Criteria</b>																	
9. Administrative skills	.20**	.20**	.11	.22**	.24**	.27**	.18**	.35**	-								
10. Ability to handle stress	.16**	.17**	.11	.16**	.17**	.21**	.20**	.18**	.48**	-							
11. Adaptability to change	.25**	.21**	.19**	.23**	.21**	.20**	.21**	.20**	.54**	.73**	-						
12. Customer service	.12*	.11	.11	.11	.24**	.19*	.21**	.20**	.58**	.48**	.54**	-					
13. Attention to detail	.22**	.19**	.11	.19**	.13*	.16**	.21**	.10	.59**	.37**	.37**	.48**	-				
14. Writing ability	.24**	.24**	.18**	.25**	.27**	.29**	.30**	.27**	.52**	.27**	.34**	.42**	.45**	-			
15. Computer skills	.21**	.21**	.22**	.19**	.14*	.25**	.19**	.24**	.56**	.31**	.40**	.37**	.49**	.49**	-		
16. Numerical ability	.25**	.18**	.20**	.20**	.10	.16**	.13*	.09	.52**	.38**	.37**	.33**	.48**	.40**	.63**	-	
17. Proofreading ability	.19**	.19**	.15**	.16**	.11	.27**	.19**	.17**	.58**	.34**	.42**	.48**	.71**	.58**	.55**	.54**	-
18. Overall measure of performance	.22**	.19**	.19**	.21**	.23**	.27**	.29**	.29**	.77**	.63**	.67**	.67**	.65**	.55**	.60**	.55**	.71**

Notes: \*  $p < .05$ ; \*\*  $p < .01$ ;  $N = 296$ .

**Table 3.** Summary of utility analysis results.

Utility Analysis Comparison	Device(s) Considered	Utility in Year 1 (Basic)	Utility in Year 1 (10% Discount Rate; 45% Tax Rate)	Utility Over 10 Years (10% Discount Rate; 45% Tax Rate; Turnover = 20/296)
Over policy of random selection	TLA	\$75,961	\$41,779	\$1,106,810
	Word	\$76,200	\$41,959	\$1,108,169
	Both	\$98,732	\$54,302	\$1,417,468
Over current battery of tests; criterion is overall measure of performance.	TLA	\$1,034	\$568	\$32,293
	Word	\$12,066	\$6,636	\$187,155
	Both	\$15,776	\$8,677	\$257,824
Over current battery of tests; multiple criteria; equal weighting of each criterion	TLA	-\$1,642	-\$903	-\$6,082
	Word	\$6,714	\$3,693	\$110,404
	Both	\$7,748	\$4,261	\$142,697
Over current battery of tests; multiple criteria; criteria weighed for fast-paced office environment	TLA	-\$840	-\$462	\$5,431
	Word	\$4,306	\$2,368	\$75,866
	Both	\$4,001	\$2,201	\$88,971
Over current battery of tests; multiple criteria; criteria weighed opposite of that in previous example.	TLA	-\$840	-\$462	\$5,431
	Word	\$8,887	\$4,723	\$137,267
	Both	\$8,283	\$4,555	\$150,372
Over current battery of tests; multiple criteria; criteria weighed for fast-paced office environment; 30% of job offers are initially rejected; -0.20 correlation between performance and probability of accepting job; one year probationary period, after which all employees not performing at least one half of standard deviation below average are dismissed	TLA	-\$912	-\$501	\$3,144
	Word	\$7,699	\$4,235	\$109,064
	Both	\$7,275	\$4,001	\$118,358

Note: TLA = Test of Learning Ability; Word = Wonderlic Personnel Test; Both = Use of both the Test of Learning Ability and the Wonderlic Personnel

## Appendix

### Calculation of the overall correlation

The correlation of any two linear combinations, x and y, can be obtained as follows:

$$R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

When considering standardized variables,

$$R_{xy} = \frac{R_{xy}}{\sqrt{R_x} \sqrt{R_y}}$$

If weights are attached to the x and y variables,

$$R_{xy} = \frac{\sum w_i \sum w_j r_{ij}}{\sqrt{\sum \sum w_i w_i r_{ii}} \sqrt{\sum \sum w_j w_j r_{jj}}}$$

This expression can be simplified by putting it in matrix notation:

$$R_{xy} = \frac{a' \sum_{12} b}{\sqrt{a' \sum_{11} a} \sqrt{b' \sum_{22} b}}$$

This can be further simplified as follows: let  $c = \sum_{11}^{-1/2} a$ ,  $d = \frac{\sum_{11}^{-1/2} \sum_{12} b}{\sqrt{b' \sum_{22} b}}$

$$R_{xy} = \frac{c' \sum_{11}^{-1/2} \sum_{12} b}{\sqrt{c' c} \sqrt{b' \sum_{22} b}}$$

$$R_{xy} = \frac{c' d}{\sqrt{c' c}}$$

R is maximized when the cross product of the two vectors are equal; that is, the product is maximized when  $c = d$

$$R_{xy} = \frac{d' d}{\sqrt{d' d}}$$

$$R_{xy} = \frac{\sum d_i^2}{\sqrt{\sum d_i^2}} = \sqrt{\sum d_i^2} = |d|$$

From above:  $c = \sum_{11}^{-1/2} a$  ;  $\sum_{11}^{-1/2} c = a$

Therefore (because  $c = d$ )

$$a = \frac{\sum_{11}^{-1/2} \sum_{11}^{-1/2} \sum_{12} b}{\sqrt{b' \sum_{22} b}}$$

$$a = \frac{\sum_{11}^{-1} \sum_{12} b}{\sqrt{b' \sum_{22} b}}$$

## References

- Borman, W. C., & Motowidlo, S. J., 1993, Expanding the criterion domain to include elements in contextual performance. In N. Schmitt & W. C. Borman and Associates (Eds.) *Personnel selection in organizations* (pp. 71-98), San Francisco, California, Jossey-Bass.
- Boudreau, J. W., 1983, Economic considerations in estimating the utility of human resource productivity improvements. *Personnel Psychology*. 36. 551-557.
- Boudreau, J. W., 1991, Utility analysis for decisions in human resource management. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, 2nd ed., pp. 621-745), Consulting Psychologist Press, Palo Alto, California
- Boudreau, J. W., 1996, Motivational impact of utility analysis and HR measurement. *Journal of Human Resource Costing and Accounting* 1, 73-84.
- Boudreau, J. W., & Berger, C. J., 1985, Decision-theoretic utility analysis applied to external employee movement. *Journal of Applied Psychology*. 70, 581-612.
- Brief, A. P., & Motowidlo, S. J., 1986, Prosocial organizational behaviors. *Academy of Management Review*. 11. 710-725.
- Brogden, H. E., 1949. When testing pays off. *Personnel Psychology*, 171-183.
- Bureau of National Affairs, 1988, *Recruiting and Selection Procedures*, Bureau of National Affairs, Washington, D.C.
- Burke, M. J., & Frederick, J. T., 1984. Two modified procedures for estimating standard deviations in utility analysis. *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. R, 1990, Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and*

- organizational psychology* (Vol. 1, pp. 687-732), Palo Alto, California, Consulting Psychologists Press.
- Cascio, W. F., 1996, The role of utility analysis in the strategic management of organizations. *Journal of Human Resource Costing and Accounting*. 1. 85-95.
- Cascio, W. F., 2000, *Costing human resources: The financial impact of behavior in organizations* (4th edition). Boston, MA, Kent Publishing.
- Cascio, W. R., 1992, Assessing the utility of selection decisions: Theoretical and practical considerations. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 310-340), San Francisco, California, Jossey-Bass.
- Cascio, W. R., & Ramos, R., 1986, Development and application of a new method for assessing job performance in behavioral/economic terms. *Journal of Applied Psychology*. 71, 20-28.
- Cascio, W. R., & Silbey, V., 1979, Utility of the assessment center as a selection device. *Journal of Applied Psychology*. 64, 107-118.
- R. D. Craig Assessments, 1990, R. D. Craig Typing Test.
- Cronbach, L. J., & Glesser, G C., 1965, *Psychological tests and personnel decision*. Urbana, Illinois, University of Illinois Press.
- De Corte, W., 1994. Utility analysis for the one-cohort selection-retention decision with a probationary period. *Journal of Applied Psychology*. 79, 402-411.
- Gatewood, R. D., & Feild, H. S., 2001, *Human resource selection* (5th edition). Chicago, Illinois: Dryden Press.
- Heneman, H. G., Heneman, R. L., & Judge, T. A., 1997, *Staffing Organizations* (2nd edition), Chicago, Illinois, Irwin.
- Hogarth, R. M., & Einhorn, H. J., 1976, Optimal strategies for personnel selection when candidates can reject job offers. *Journal of Business*. 49, 479-495.
- Johnson, R. A., & Wiehern, D. W., 1998, *Applied multivariate statistical analysis* (4th edition), Englewood Cliffs, New Jersey, Prentice Hall.

Latham, G. R., & Whyte, G., 1994, The futility of utility analysis. *Personnel Psychology*. 47, 31-46.

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

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.

Milkovich, G. T., & Newman, J. M., 2002, *Compensation* (7th edition). New York, Irwin-McGraw Hill.

Motowidlo, S. J., & Van Scotter, J. R., 1994, Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*. 79, 475-480.

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

Murphy, K. R., 1989, Dimensions of job performance. In R. Dillon & J. Pellingrino (Eds.), *Testing: Applied and theoretical perspectives* (pp. 218-247), New York, Praeger.

Nunnally, J. C., & Bernstein, I. H., 1994, *Psychometric theory* (3rd edition). New York, McGraw-Hill.

Organ, D. W., 1988, *Organizational citizenship behavior: The good soldier syndrome*, Lexington, Massachusetts, Lexington Books.

Raju, N. S., & Burke, M. J., 1986, Utility analysis. In R. A. Burk (Ed.), *Performance assessment: Methods and applications* (pp. 186-202). Baltimore, Maryland, Johns Hopkins University Press.

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.

Richardson, Bellows, Henry, & Co., 1989, *Test of Learning Ability*.

Robinson, S. L., & Bennett, R. J., 1995, A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal*. 38. 555-572.

Rotundo, M., & Sackett, P. R., In press, The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy capturing approach. *Journal of Applied Psychology*.

- Rynes, S. L., & Boudreau, J. W., 1986, College recruiting in large organizations: Practice, evaluation, and research implications. *Personnel Psychology*. 39. 729-757.
- Schmidt, F. L., & Hunter, J. E., 1983, Individual differences in productivity: An empirical test of estimates derived from studies on 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*. 73, 46-57.
- 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.
- SRA/London House Office Skills Tests, 1977, *Checking*.
- SRA/London House Office Skills Tests, 1977, *Coding*.
- SRA/London House Office Skills Tests, 1977, *Filing*.
- SRA/London House Office Skills Tests, 1977, *Grammar*.
- SRA/London House Office Skills Tests, 1977, *Punctuation*.
- Sturman, M. C., 2000, Implications of utility analysis adjustments for estimates of human resource intervention value. *Journal of Management*. 26, 281-299.
- Whyte, G., & Latham, G R, 1997, The futility of utility analysis revisited: When even an expert fails. *Personnel Psychology*. 50. 601-610.
- E. F. Wonderlic and Associates. (1983). *Wonderlic Personnel Test*.