June 1995

Utility Analysis for Multiple Selection Devices and Multiple Outcomes

Michael C. Sturman
Cornell University, mcs5@cornell.edu

Timothy A. Judge
Cornell University

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 leads to overestimates of a device's value. This paper generates a new utility model that accounts for multiple selection devices and multiple criteria. The model is illustrated using previous utility analysis work and an actual case of secretarial employees with eight predictors and nine criteria. A final example also is provided which 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
research, model, organization, information, psychology, applied, study, effect, human, employ, work, criteria, utility

Comments

Suggested Citation

Sturman, M. C. & Judge, T. A. (1995). *Utility analysis for multiple selection devices and multiple outcomes* (CAHRS Working Paper #95-12). Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.
http://digitalcommons.ilr.cornell.edu/cahrswp/202

CAHRS

Center for Advanced Human Resource Studies

CAHRS / Cornell University
187 Ives Hall
Ithaca, NY 14853-3901 USA
Tel. 607 255-9358
www.ilr.cornell.edu/CAHRS/

WORKING PAPER SERIES

Utility Analysis for Multiple Selection Devices and Multiple Outcomes

Michael C. Sturman
Timothy A. Judge

Working Paper 95 – 12

ilr School of Industrial and Labor Relations
Advancing the World of Work

CORNELL UNIVERSITY

Utility Analysis for Multiple Selection Devices and Multiple Outcomes

Michael C. Sturman and Timothy A. Judge
Center for Advanced Human Resource Studies
Cornell University
393 Ives Hall
Ithaca, New York 14853-3901
(607) 255-4470

Working Paper #95-12

www.ilr.cornell.edu/cahrs

Running Head: UTILITY ANALYSIS

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 leads to overestimates of a device's value. This paper generates a new utility model that accounts for multiple selection devices and multiple criteria. The model is illustrated using previous utility analysis work and an actual case of secretarial employees with eight predictors and nine criteria. A final example also is provided which 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.

Utility Analysis for Multiple Selection Devices and Multiple Outcomes

Utility Analysis (UA) has evolved into a complex tool for estimating the value of human resource interventions. UA allows human resource decision makers to produce bottom line figures, thus adding credibility to the perceived "soft" decisions commonly associated with human resources. UA has had to overcome many limitations of its application over the years, especially with regard to the credibility of its results. These limitations have been partially addressed through developments of SD_y calculation (e.g., Cascio & Ramos, 1986; Schmidt & Hunter, 1983; Schmidt, Hunter, McKenzie, & Muldrow, 1979), the addition of finance variables (Boudreau, 1983), techniques such as breakeven analysis (Boudreau, 1984), accounting for employee flows over time (Boudreau & Berger, 1985), relaxing the assumption of top-down hiring (Hogarth & Einhorn, 1976; Murphy, 1986), determining the effects of recruiting efforts (Rynes & Boudreau, 1986), and allowing poor performers to be dismissed (De Corte, 1994). With these developments have come added complexity; yet, even in its current state, limitations still exist which cause UA to produce imperfect estimates of the expected utility of human resource staffing policies. This paper will address some of these limitations and present a modified UA that may prove to be more accurate.

Revising the Traditional Utility Analysis Formula

There is a large body of literature which reveals that selection devices can predict multiple criteria relevant to employment. Although the majority of validity studies use performance ratings as the criterion, a meta-analysis of various selection devices performed by Schmitt, Gooding, Noe and Kirsch (1984) shows that many studies have addressed the effect of selection measures on turnover, achievement/ grades, productivity, status changes, wages, and work samples. Similarly, an organization may wish to hire employees based on needed levels of competencies, rather than based on a more abstract notion of general performance. Because current UA models only concern themselves with a single criterion variable, traditional UA estimates may be inaccurate. 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, new UA models need to be constructed to account for multiple criteria.

Perhaps a more central weakness of present formulations of utility is that while it measures the utility of a single device over random selection, most organizations' gather information from multiple sources in making staffing decisions. The vast majority of organizations use reference checks (97%) and interviews (81%) when making selection decisions; other commonly used devices include application blanks and various testing programs (Gatewood & Feild, 1994). Thus, very few organizations limit their selection process by using only one, let alone no, selection procedure.

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 inappropriate 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 B-C-G 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 UA formula, as currently developed, uses a correlation coefficient, based on the correlation between a single predictor, $\underline{p}$, and a single criterion, $\underline{c}$. By expanding the analysis to a set of predictors and criteria, represented by the vectors $\underline{p}$ and $\underline{c}$ the computations needed to calculate a single correlation value become more complex. The analysis becomes even more complex given that the practitioner would want the predictors weighted to yield maximum potential predictive power, but the performance criteria would determined independent of the utility analysis, such as from a job analysis.

We can therefore see the new correlation coefficient being based on the correlation between two linear combinations: $\underline{u}$, which equals the predictors, $\underline{p}$, times a set of weights, $\underline{a}$ and, $\underline{v}$ which equals the criteria, $\underline{c}$ times a set of weights, $\underline{b}$.

$$\underline{u} = \underline{a} \times \underline{p}$$

$$\underline{v} = \underline{b} \times \underline{u}$$

The set of weights, $\underline{a}$ will be mathematically derived, but the set of criteria weights, $\underline{b}$ will be predetermined. The problem thus becomes calculating the correlation between the linear combinations $\underline{u}$ and $\underline{v}$. The solution to the problem is based on the principles behind calculating the correlation of sums (Nunnally, 1978). One might also note that this problem is similar to the calculation of canonical correlations (Johnson & Wichern, 1992); however, because the set of weights $\underline{b}$ are not being modified to yield a maximal correlation, the resulting solution will differ.

Figure 1 shows the formulas and necessary covariance matrix for calculating the correlation value. The derivation is shown in the Appendix.

FIGURE 1: Needed Information and Matrix Calculation

Predictors = $[p_1, p_2, \ldots, p_n)$

Predictor Weights $\mathbf{a}$ = $[w_{p1}, w_{p2}, \ldots, w_{pn}]$ To be maximized

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

Criteria Weights $\mathbf{b}$ = $[w_{c1}, w_{c2}, \ldots, w_{cm}]$ Predetermined by Organization

$$
\begin{array}{ll|c|c|}
 & & \textbf{Predictors} & \textbf{Criteria} \\
 & & 1\ 2\ \ldots\ n & 1\ 2\ \ldots\ m \\
\hline
\textbf{Predictors} & \begin{matrix}1\\2\\n\end{matrix} & \Sigma_{1\,1} & \Sigma_{2\,1} \\
\hline
\textbf{Criteria} & \begin{matrix}1\\2\\m\end{matrix} & \Sigma_{1\,2} & \Sigma_{2\,2} \\
\hline
\end{array}
$$

$$\mathbf{r_{xy}} = \frac{a' \, \Sigma_{1\,2} \; b}{\sqrt{a' \, \Sigma_{1\,1} \; a} \; \sqrt{b' \, \Sigma_{22} \; b}}$$

$$\mathbf{a} = \frac{\Sigma_{1\,1}^{-1} \; \Sigma_{1\,2} \; b}{\sqrt{b' \, \Sigma_{22} \; b}}$$

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 formulas of UA, 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 formulae 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, 1991).

To illustrate the use of this new model, this paper will present two sets of examples of computing the expected utility of new selection devices. The first example will apply the advances described here to previous work on utility analysis. The second example will be an actual implementation of the revised utility formula for the selection of secretarial workers. The examples also will show how the new utility model can yield results quite different from the traditional model.

Application of Revised Utility Model to Previous UA Research

With the equations for determining the utility of multiple devices for multiple criteria defined, we will now turn to illustrating its use. Specifically, this section of the paper will reapply the new UA model to previous research where simpler UA models were used.

To do this, we rely on a review of UA found in Boudreau (1991a) which includes an appendix containing summaries of 42 utility analyses, 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 is compared to the utility of an interview. However, these studies only investigated the value of the new selection device over the interview and do not provide information of the incremental utility achieved by implementing the new device in conjunction with the interview. Thus, we will apply the utility model derived in the beginning of this paper to obtain a more realistic estimation of the utility of implementing the new selection devices.

Before presenting our 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. Second, information on the correlation between the interview and the new selection device was not provided by the studies. Thus, the analysis will be performed with an assumed intercorrelation. In order to be accurate and fair, the analysis will be performed for intercorrelation values of 0, 0.25, 0.50, and 0.75. This method of performing UA using different reasonable values of a specific variable also has been used by other researchers to better estimate the value of selection devices or to illustrate a UA technique (e.g., Cascio & Silbey, 1979; Murphy, 1986; Rich & Boudreau, 1987). Although it is our belief that the intercorrelations will generally be between 0 and 0.50, all of these values will be shown to include a broad range of possible effects. Information on selection devices, validities, and utilities for each study are shown in Table 1.

TABLE 1: Validity and Utility of Multiple Selection Devices

	Validity				Utility			
Study/ Devices Used	Device 1 (Interview)	Device 2	Inter-correlation	Combined	Device 1 (Interview)	Device 2	Combined	Adding Device 2
Cascio & Silbey	0.25	0.35	0.00	0.43	$350,357	$504,211	$575,733	$225,376
(1979)	0.25	0.35	0.25	0.39	$350,357	$504,211	$509,476	$159,119
(Assessment Center)	0.25	0.35	0.50	0.36	$350,357	$504,211	$459,783	$109,426
	0.25	0.35	0.75	0.35	$350,357	$504,211	$441,611	$91,254
Schmidt, Hunter,	0.14	0.76	0.00	0.77	$6,849,022	$38,755,422	$38,911,801	$32,062,779
McKenzie &	0.14	0.76	0.25	0.76	$6,849,022	$38,755,422	$38,396,983	$31,547,961
Muldrow (1979)	0.14	0.76	0.50	0.81	$6,849,022	$38,755,422	$40,971,076	$34,122,054
(PAT Test)	0.14	0.76	0.75	NA	$6,849,022	$38,755,422	NA	NA
Ledvinka, Simonet,	0.14	0.36	0.00	0.39	$14,881	$37,162	$40,351	$25,470
Neiner & Kruse	0.14	0.36	0.25	0.36	$14,881	$37,162	$37,162	$22,281
(1983)	0.14	0.36	0.50	0.36	$14,881	$37,162	$37,162	$22,281
(JEPS Test)	0.14	0.36	0.75	0.41	$14,881	$37,162	$42,477	$27,596
Schmidt, Mack &	0.14	0.51	0.00	0.53	$644,384	$2,960,542	$2,853,740	$2,209,356
Hunter (1984)	0.14	0.51	0.25	0.51	$644,384	$2,960,542	$2,728,542	$2,084,158
(PACE Test)	0.14	0.51	0.50	0.53	$644,384	$2,960,542	$2,853,740	$2,209,356
	0.14	0.51	0.75	0.63	$644,384	$2,960,542	$3,479,728	$2,835,344
Burke & Frederick	0.16	0.59	0.00	0.61	$42,098	$115,727	$98,432	$56,334
(1986)	0.16	0.59	0.25	0.59	$42,098	$115,727	$89,980	$47,882
(Assessment Center)	0.16	0.59	0.50	0.61	$42,098	$115,727	$98,432	$56,334
	0.16	0.59	0.75	0.73	$42,098	$115,727	$149,310	$107,212
Cascio & Ramos	0.14	0.39	0.00	0.41	$7,335,605	$20,830,312	$21,101,212	$13,765,607
(1986)	0.14	0.39	0.25	0.39	$7,335,605	$20,830,312	$19,903,804	$12,568,199
(Assessment Center)	0.14	0.39	0.50	0.39	$7,335,605	$20,830,312	$19,903,804	$12,568,199
	0.14	0.39	0.75	0.45	$7,335,605	$20,830,312	$23,496,028	$16,160,423
Rich & Boudreau	0.14	0.73	0.00	0.74	$431,744	$3,198,258	$3,024,175	$2,592,431
(1987)	0.14	0.73	0.25	0.73	$431,744	$3,198,258	$2,972,715	$2,540,971
(PAT Test)	0.14	0.73	0.50	0.77	$431,744	$3,198,258	$3,165,209	$2,733,465
	0.14	0.73	0.75	0.95	$431,744	$3,198,258	$4,011,408	$3,579,664

Note: NA = Not available because this intercorrelation is not possible given the correlations of the two devices.

The results of the new utility analysis lead to a number of conclusions. Most notably, increases in validity from combining two devices were almost always less than the sum of the validities of the two selection devices analyzed separately. The only exception is the case where the assumed intercorrelation makes the combined validity close to 1.0 (e.g. Rich & Boudreau, 1987; combined validity = .95). For all other cases, including the Rich and Boudreau case at more reasonable intercorrelations (not .75), 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 84%. 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 either stayed the same or dropped for 16 of the 27 calculations. The majority of the increases occurred when the intercorrelation was high (.75); only four increases were calculated for circumstances of intercorrelation values of a magnitude similar to those values commonly reported in the selection literature (.00, .25, and .50).

As mentioned earlier, the more 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. The methods employed here, though, 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.

Application of Revised Utility Model to a Selection Program

While the demonstration of this revised approach to UA using previous research illustrates the general differences between it and the traditional method, an actual example of its application in a situation where there are more than two predictors, multiple criteria, and all the necessary intercorrelation data, would being useful. We will now describe such an application considering various selection tests and multiple dimensions of job performance.

The application is for secretarial positions at a specific 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), will be 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. The majority of 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); however, the use of mental ability tests is increasing (Gatewood & Feild, 1994). Thus, it is quite plausible that a company may be considering adding a mental ability test to its current battery of work sample tests. It should also be noted that this decision situation is merely an example designed to illustrate the new UA approach. Although the validity coefficients and company information are based on real data, the scenario is just one of many potential situations for which this UA approach is applicable.

Secretaries are evaluated on 9 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 secretary received an overall performance rating. A validity study was conducted on 296 current secretarial workers 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.

TABLE 2: Correlations of Predictors and Criteria

		Predictors								Criteria								
	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Predictors	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**	—									
Criteria	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.

A utility analysis, using the techniques prescribed in this paper for computing r will be compared to a traditional utility analysis. The other aspects of the utility equation (e.g., N, SD_y, Z_x, & C) will be the same for both calculations. The number of applicants and number selected are based on estimates from the company which are representative of a hiring situation they might face.

The estimated number of new secretaries needed each year is 20. The average tenure for a secretary 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,, score of 2.23. The Wonderlic Personnel Test (Wonderlic, 1983) has an initial cost of $130, which includes 50 tests, manual, and a score key. One hundred additional tests can be purchased for $158. The Test of Learning Ability (Richardson, Bellows, Henry & Co., 1989) costs $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 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, SD_y was based on the 40% rule (Schmidt & Hunter, 1983). Although there are other methods of estimating SD_y (Cascio, 1991), the 40% rule seems adequate for this study. Thus, based on a reported starting pay of secretaries at the company, SD_y equals roughly $6,000.

Utility Evaluations

To demonstrate the implications of the methods described in this paper, a number of utility analyses are performed. Note that it is not the intent of this study to provide a detailed description of the types of tests used and the implications of their validities. Rather, this study is intended to demonstrate an application of a revised utility analysis technique and to show how it differs from the current approach.

First, a conventional analysis is performed, 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 will be used in a basic utility analysis. In both of these cases, the correlation between the selection devices and the overall measure of performance will be used.

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

The third UA will employ all the methods proposed in this paper. The utility of the two selection devices, both separately and in conjunction with each other, will be evaluated in addition to the current battery of six tests, and the individual facets of performance will be used to weight the tests. Additionally, examples will be 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 UA will be performed which combines the advances of this paper with other researchers' work refining the utility model. The basic example will be the same as the second case of the third analysis, but the UA will also include 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.

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
	Wond	$76,290	$41,959	$1,108,169
	Both	$98,732	$54,302	$1,447,468
Over current battery of tests; criterion is overall measure of performance.	TLA	$1,034	$568	$32,293
	Wond	$12,066	$6,636	$187,155
	Both	$15,776	$8,677	$257,824
Over current better of tests; multiple criteria; equal weighting of each criterion	TLA	-$1,642	-$903	-$6,082
	Wond	$6,714	$3,693	$110,404
	Both	$7,748	$4,261	$142,697
Over current batter of tests; multiple criteria; criteria weighted for fast-paced office environment	TLA	-$840	-$462	$5,431
	Wond	$4,306	$2,368	$75,866
	Both	$4,001	$2,201	$88,971
Over current battery of tests; multiple criteria; criteria weighted opposite of that in previous example.	TLA	-$840	-$462	$5,431
	Wond	$8,587	$4,723	$137,267
	Both	$8,283	$4,555	$150,372
Over current battery of tests; multiple criteria; criteria weighted 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
	Wond	$7,699	$4,235	$109,064
	Both	$7,275	$4,001	$118,358

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

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, 1991)

$$DU = N_h * Z_x * r * SD_y - [Ci + (C * N_a)]$$

ΔU = 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

Na = 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, 1978, p. 177), 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. Even if we modify these estimates by accounting for the turnover rate (20/296) of the secretarial employees, using an estimated discount rate (10%), and using a tax rate (45%), 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. Yet, even this simplistic analysis demonstrates that the utility of using two devices is significantly different from the sum of using both devices independently.

Utility of Selection Devices Over Current Battery

We will now begin to apply the methods proposed in this paper to the secretarial 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 .07.

With the incremental r's 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) are 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 ($\Delta r = .00$), the change in r by adding the Wonderlic Personality Test ($\Delta r = .03$), and the change in r by adding both devices ($\Delta r = .04$).

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 secretaries in a faster-pace, 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 utility 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.

Comprehensive Model

So far, this paper has concentrated on examples of the methods proposed here to make the utility estimate more accurate. We have included, though, some advances from other research (e.g., effect of flows over time, discount rate, tax rate adjustments). However, other research has developed refinements of the utility model that should also make the estimates more accurate. Although we have not included these methods in the preceding sections for the sake of simplifying our examples, we will now included these advances here to demonstrate how the total UA model might work.

Specifically, we will include two advances of the utility model: 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 are judged to perform at an unacceptable level (De Corte, 1994).

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 in other words, 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, and we are unaware of any research investigating this matter, we 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, and no research specifically addressing this issue. 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 a acceptance rate of 70% seems reasonable, and is perhaps a little conservative. Murphy (1986) provides methods for estimating the average Z-score of those hired 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 $\underline{r}$ 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 our case, we will assume that performance must be no less than one standard deviation below average. Those employees who score below this threshold will 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.

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 utility analysis 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 criteria weighting schemes did not yield as large a difference, 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 will almost invariably be less than the sum of using each device individually. However, when the devices are considered over a policy which includes a number of selection devices, the opposite may be true. This is true 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 (1978, pp. 177-178). In Nunnally's 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 produce a method to 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 no generalizability to other situations where the Test of Learning Ability or the Wonderlic Personnel Test are being considered. Indeed, this paper contends that the utility of new selection devices depend on the current firm-specific practices. What are generalizable are the methods for calculating utility. These methods should yield, for any realistic hiring situation, more accurate estimates of the dollar return of a proposed selection device.

The results do allow us to make some generalizations regarding the implementation of selection devices and the implications of the proposed changes for utility analysis 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, we do not purport 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. This paper merely shows that estimates generated using the B-C-G utility analysis model are too high in realistic decision making contexts.

Application of the New Utility Model for Decision Makers

With the apparent lack of use of UA by managers, and with the added complexity of this new model, it may seem less likely that managers will take advantage of this more accurate tool. 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 exist. 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 utility analysis. 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 utility analysis, in its current form, may have a limited impact on managerial decision making (Latham & Whyte, 1994). While one of the reasons for the limited impact may be the unrealistic nature of the assumptions underlying utility analysis (Cascio, 1992), another reason may lie in its complexity (Latham & Whyte, 1994). Thus, while the revisions of the utility analysis formula presented in this paper make the algorithm more realistic, they also make it more complex. As a result, it is not clear that these improvements in the basic formula will ultimately result in greater applicability of utility analysis.

Future Research

Research on selection devices needs to pay more attention to the intercorrelation between the new device and existing devices. Although there are myriad studies showing the correlation between single selection devices (e.g. biodata, personality tests, cognitive ability tests, honesty tests, etc.) 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 centers, work samples, 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 of trial-and-error permutations of current and potential procedures. Thus, future research investigating the combination of optimization strategies for multiple devices will further help human resource decision makers effectively develop a selection procedure.

Overall, research on utility analysis 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 goal is perhaps the most needed work, advances in the first two areas are needed before the third goal is even feasible.

References

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

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

Boudreau, J. W. (1991a). 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). Palo Alto, CA: Consulting Psychologist Press.

Boudreau, J. W. (1991b). Staffing planning at COMPUTERCO: A personal computer analysis. Working paper #91-12, Center for Advanced Human Resource Studies, Cornell University.

Boudreau, J. W. (1991c). EXTMOV: A spreadsheet program for quantifying external employee movement decisions. Working paper #91-13. Center for Advanced Human Resource Studies, Cornell University.

Boudreau, J. W., & Berger, C. J. (1985). Decision-theoretic utility analysis applied to external employee movement. Journal of Applied Psychology, 70, 581612.

Bureau of National Affairs. (1988). Recruiting and Selection Procedures. Washington, D.C.: Bureau of National Affairs.

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

Cascio, W. F. (1991). Costing human resources: The financial impact of behavior in organizations. Boston, MA: Kent.

Cascio, W. F. (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, CA: Jossey-Bass.

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 71 20-28.

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

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. (1994). Human resource selection. Chicago, IL: Dryden Press.

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., & Wichern, D. W. (1992). Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice Hall.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt.

Latham, G. P., & 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 TEPS at Life of Georgia. Unpublished technical report.

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.

Nunnally, J. C. 1978. Psychometric theory. New York: McGraw-Hill.

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.

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

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

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.

Thomas; G. B. jr., & Finey, R. L. (1986). Calculus and Analytic Geometry (6th ed.), Reading, MA: Addison-Wesley.

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

E. F. Wonderlic and Associates. (1983). Wonderlic Personnel Test.

APPENDIX A: Calculation of the Overall Correlation

The correlation of any two linear combination, x and y, can be obtained as follows: $R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

When considering standarized variables, $R_{xy} = \frac{R_{xy}}{\sqrt{R_x}\sqrt{R_y}}$

If weights are attached to the x and y variables, $R_{xy} = \frac{\Sigma w_i \ \Sigma w_j \ r_{ij}}{\sqrt{\Sigma \Sigma w_i \ w_i \ r_{ii}} \quad \sqrt{\Sigma \Sigma w_j \ w_j \ r_{jj}}}$

This expression can be simplified by putting it in matix motation $R_{xy} = \frac{a' \Sigma_{12} \ b}{\sqrt{a' \Sigma_{11} \ a} \ \sqrt{b' \Sigma_{22} \ b}}$

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

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

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

r is maximized when the cross product of the two vectors are equal; that is, the product is maximized when **c** = **d** (Thomas & Finey, 1986).

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

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

From Above: $\quad \mathbf{c} = \Sigma_{11}^{1/2} a$

$$\Sigma_{11}^{-1/2} \ c = a$$

Because **c** = **d**:

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

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