Implications of Utility Analysis Adjustments for Estimates of Human Resource Intervention Value

Michael C. Sturman, Louisiana State University

This paper argues that most utility analysis (UA) applications are flawed because they employ an overly simplistic formula. This piece reviews the literature on UA adjustments and demonstrates that the adjustments have a large impact on resulting estimates. These results imply that human resources programs do not invariably yield positive returns; rather, intervention success is contingent on program-specific, organizational, and environmental factors.

Utility Analysis (UA) allows human resource (HR) decision makers to produce bottom line figures, thus purportedly (a) aiding managers deciding whether to invest in HR management systems, such as selection systems or training programs, and (b) adding credibility to the perceived "soft" decisions commonly associated with HR (Cascio, 2000). The usefulness of UA as a managerial decision aid, though, is questionable. Many have cautioned that UA produces inflated estimates (Becker, 1989; Boudreau, 1983a; Cascio, 1993; Latham & Whyte, 1994; Murphy, 1986), which may make managers skeptical of its results and the methods used to obtain them (Hunter, Schmidt, & Coggin, 1988; Wintermantel & Mattimore, 1997). Although there are notable instances where UA helps inform managerial evaluations (e.g., Florin-Thuma & Boudreau, 1987; Morrow, Jarrett, & Rupinski, 1997), some have argued that other demonstrations with astronomical estimates strain the credibility of UA specifically, and I/O Psychology in general (Ashe, 1990; Cascio, 1993).

Even proponents of UA admit that practitioners view UA predictions skeptically (Cronshaw, 1997). However, it is unclear whether adverse reactions to UA signify an aversion to quantitative HR decision aids in general or simply reflect a disbelief in the results of an inaccurate decision aid. While recent experiments (Latham &Whyte, 1994; Whyte & Latham, 1997) show that managers do not find UA to be useful, their analyses omitted potentially useful modifications to the equation which may have made perceptions of the results more positive. Latham and Whyte (1994; Whyte & Latham, 1997) use the basic Brogden-Cronbach- Glesser model. Whereas this model represents an important first step in UA research (i.e., Brogden, 1949; Cronbach & Glesser, 1965), the formula relies on a number of assumptions. These assumptions are known to limit UA's accuracy and inflate its estimates; however, researchers have produced modifications to the formula to help correct many of these problems (e.g., Cascio, 1993).

To help improve the credibility, and arguably the validity of UA estimates, researchers have made a number of advances in UA calculation in the form of adjustments to the basic UA formula (e.g., Boudreau, 1983a; Boudreau & Berger, 1985; De Corte, 1994; Murphy, 1986); however, no research has yet examined the implications of all of these adjustments simultaneously. This paper provides a review of published, algebraic UA adjustments and demonstrates the implications of these adjustments for a wide range of plausible parameter values. Although there have been descriptive reviews of the types of adjustments available (e.g., Cascio, 1993), and there have been comparisons of utility models (e.g., Judiesch, Schmidt, & Hunter, 1993; Raju, Burke, & Normand, 1990; Raju, Burke, Normand, & Lezotte, 1993), no research has yet compared the magnitude of these adjustments' impacts.

Through this review and comparison, this paper contributes to the UA and selection literatures in two ways. First, this paper investigates the practical implications of applying these adjustments simultaneously, thus providing guidance for future UA applications as to which UA modifications have the most prominent effects. This contribution should be of benefit both for practitioners using UA to estimate values and for researchers investigating how decision-makers react to UA. Second, this paper demonstrates how the impact of generalziable behavioral findings (e.g., the validity of a type of selection device) is dependent on a variety of organizational-specific and environmental factors.

The Basic Utility Model, its Assumptions, and its Use

The basic utility formula, emerging from the work by Brogden (1949) and Cronbach and Glesser (1965), is as follows:

$$\Delta U = T \times N_h \times \overline{Z_x} \times r \times SD_y - C$$

$\Delta U =$ Utility change from selection device

$N_h =$ Number of people to be hired

$T =$ Average tenure of those hired

$\overline{Z_x} =$ Average Z-score of the predictor of hired employees

$r =$ Correlation between a predictor and criterion

$SD_y =$ Dollar value of a standard deviation in the criterion

$C =$ Cos of acquiring and administering a selection battery

The usefulness of this UA model has been challenged both with respect to the estimation method and its application by decision makers. The equation subsumes a number of critical assumptions, each of which may affect the utility estimate yielded by the formula. Researchers have proposed algebraic modifications to the UA model to include the use of finance variables (Boudreau, 1983a), account for employee flows over time (Boudreau & Berger, 1985), relax the assumption of top-down hiring (Hogarth & Einhorn, 1976; Murphy, 1986), recognize that organizations use multiple selection devices rather than a policy of random selection (Brogden, 1949; Burke & Frederick, 1986; Ruju & Burke, 1986), and allow for the possibility that organizations will dismiss poor performers (De Corte, 1994). Although other research has also noted a number of other phenomenon that may also effect utility estimates—labor makets (Becker, 1989), validity changes over time (Henry & Hulin, 1997; Hulin, Henry, & Noon, 1990), and learning curves (Cascio, 1993)—this paper focuses on evaluating the implications of the algebraic modifications that have been published in the UA literature. As these adjustments require greater complexity, it is important to evaluate their relative effects. Although adding other variables may also have sizable effects, the UA literature has not yet benefited from a systematic evaluation of the set of techniques proposed to improve the basic UA formula.

Adjustments to the Basic Utility Formula

Despite concerns regarding its formula and use, the basic utility formula provides an essential foundation upon which further UA research has been built. Essentially, this utility formula reveals that the value ascribed to HR programs is attributable to three factors: costs, quantity of employees affected, and quality of employees affected (Boudreau & Berger, 1985). This section reviews algebraic modifications to the basic utility formula.

Note that this paper does not address methods of estimating UA parameters. For example, accurate UA estimates depend in part on accurate estimates of selection device validity (r). If r is estimated from range restricted data, then steps should be taken to correct the estimate and produce an accurate value (see Raju, Burke, & Maurer, 1995). Similarly, there has been substantial debate on how the dollar value of employee performance should be measured (e.g., Judiesch et al., 1993; Raju et al., 1990; Raju et al., 1993). This paper views such techniques as methods helping in the estimation of important parameters, which certainly may have large implications for final UA estimates. Those employing UA should strive to obtain accurate estimates of the necessary parameters; however, this paper focuses on the effects of algebraic modifications to the basic UA model.

Adjustments Affecting Estimates of Costs

Economic factors. As UA supposedly produces estimates that allow managers to evaluate the value of a HR investment, one of the earliest adjustments to the basic UA formula was to adjust its result to reflect variables commonly used when evaluating other investments. Boudreau (1983a) recognized that UA models addressed economic and financial consequences of HR decisions, but the original UA formula failed to incorporate certain financial and economic considerations.

The adjustments introduced by Boudreau (1983a) included three different processes. First, he differentiated between increases in performance and increases in total value. Performance increases may cause additional costs, such as increased compensation, higher costs of goods sold, etc. Thus, Boudreau (1983a) introduced variable costs into the utility formula to reflect the added costs associated with performance gains. Second, Boudreau noted that taxes play a role for affecting the gains and losses associated with any investment. He thus incorporated tax rate considerations into the basic utility formula. Third, Boudreau modified the UA formula to determine the present value of an investment. By incorporating a discount factor into the equation, the UA formula computes the present value of a program and can be more comparable to other investment information. Accounting for such economic factors often dramatically reduces the estimates produced by the UA formula (Boudreau, 1983a; 1991).

Adjustments Affecting Estimates of Quantity

Employee flows. Employee flows into, through, and out of an organization over time influence the value of a HR intervention (Boudreau & Berger, 1985). These effects may occur over multiple years; thus, considering the utility of an intervention only in its first year may understate its utility (Boudreau, 1983b). This occurs when a program affects a workforce for more than one year or if an organization implements a program for multiple time periods. Thus, a program affects new portions of a workforce, and the benefit for multiple groups may continue over multiple time periods.

Employee flows affect the quantity of employees influenced by a HR intervention (Boudreau, 1983b). A simple adjustment involves multiplying the benefit by the average tenure of selected employees (Schmidt, Hunter, McKenzie, & Muldrow, 1979), which was included above as part of the basic utility model; however, this approach misses aspects of what occurs over time, such as additional costs and benefits associated with replacing those who leave the organization. It also ignores the time value of money, for which the economic adjustments account.

Specific derivation and application of these methods are described in detail in Boudreau (1983b), but essentially require that each cohort of employees be tracked over time. Whereas tracking employee cohorts in multiple time periods adds complexity, it makes the UA estimate more precise (Boudreau & Berger, 1985).

Adjustments Affecting Estimates of Quality

Employee flows. In addition to effects on quantity, employee flows may affect the average quality of employees in an organization. Boudreau and Berger (1985) illustrate how the quality of those staying with an organization (versus that of those leaving) affects the utility of a selection system. As for the adjustment for quantity described above, accounting for employee quality over time necessitates tracking cohorts over the requisite time periods. This allows the UA formula to account for such phenomena as the relationship between turnover and performance, which affects average employee performance levels over time. Granted, such calculations necessitate more work and greater complexity, but they describe more specifically employee performance levels that result from a HR intervention.

The Boudreau and Berger (1985) model contributes to the process of UA calculations by providing a framework for considering costs, quantity, and quality over time. It has no specific algebraic changes regarding quality, but provides a method that allows the UA calculator to consider such relationships as that between turnover and performance (e.g., Williams & Livingstone, 1994). Others have built upon this framework to consider other changes to UA calculations, such as the effect of a probationary period for employees (De Corte, 1994).

Effect of a probationary period. Another refinement to the utility model involves determining the effect of providing a probationary period for new hires. In a probationary period, an organization dismisses those who do not perform adequately after a given time period (e.g., one year). To reflect this change, the utility equation must estimate the performance difference between initially hired employees and employees who survive the probationary period (De Corte, 1994).

Because organizations dismiss lower performers, the average performance of a given cohort increases after the probationary period ends. As described by De Corte (1994), the change in average performance depends on the validity of the selection device and the cutoff for determining what constitutes unsuccessful performance. De Corte (1994) also discusses the importance of accounting for such costs as training cost per selectee and separation cost per unsuccessful employee, both of which would impact the cost estimate of a hiring program.

The effect of De Corte's adjustments are to increase the estimated utility of a selection device; however, this increase occurs for the intervention in question and the current policy (or random selection). Including a probationary period thus reduces the utility gain attributable to a selection intervention. De Corte (1994) explains the specific utility model changes involved in detail, although they can be seen as a special case of the Boudreau and Berger (1985) model. It is, therefore, relatively easy to incorporate this adjustment using the Boudreau and Berger (1985) framework.

Multiple selection devices. Another assumption of the basic utility model is that it provides the utility gain of a selection battery over a policy of random selection (Boudreau, 1991; Cascio, 2000). Utility researchers (e.g., Brogden, 1949; Burke & Frederick, 1986; Ruju & Burke, 1986) have long noted this assumption to be weak. An examination of organizational practices shows a majority of organizations use multiple selection devices, such as reference checks, interviews, application blanks, work samples, cognitive ability tests, and assessment centers (Gatewood & Feild, 1994). Although some of these devices have low validity, the validity of all of these methods is greater than zero (Gatewood & Feild, 1994). Thus, comparing a new selection device to a policy of random hiring overestimates the value of the selection device.

Incorporating the effects of multiple selection devices into UA calculations is not new. Brogden (1949) pointed out that the formula for estimating the payoff of a selection device may be modified to describe the difference between two selection batteries. This same approach continues to be recommended as a means to compare the utility of an existing selection program with another selection program (Raju & Burke, 1986), and has been applied to make such demonstrations (Boudreau, 1991; Burke & Frederick, 1986). To properly utilize the UA formula, the validity of the selection battery including the new selection device must be computed and compared to the validity of the selection battery without the new method (or the battery it is replacing). The value of r that must then be used in the analysis equals the difference between the predictive validity of the new selection battery (R_{New}) and the predictive validity of the old selection battery (R_{Old}).

Note, however, that simply subtracting the correlation values (e.g., Boudreau, 1991; Raju & Burke, 1986) is only a partial adjustment to the basic utility equation. Instead, this adjustment should be incorporated with the other adjustments for cost, quantity, and quality because the validity adjustment may have effects for other UA parameters. For example, a different validity score may also affect the UA modifications attributable to employee flows and a probationary period. Computing utility using R_{New} and R_{Old}, and then computing the difference in utility values is therefore recommended.

Departures from top-down hiring. UA estimates from the basic model are calculated with the assumption that candidates are hired in a top-down manner. That is, those with the highest predicted performance scores are hired until all the desired number of positions have been filled. This assumption is unwarranted because some high-scoring applicants may reject offers (Hogarth & Einhorn, 1976; Murphy, 1986). Additionally, some high scoring job applicants may not receive offers because of specific organizational selection practices, such as veteran preference policies or efforts to enhance racial and/or gender composition. When the top job candidates are not hired, the employer must make offers 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 $\overline{Z_y}$.

Murphy (1986) provides formulae to determine the effect of job offer rejections on resulting utility estimates. He provides three cases: where applicants reject job offers at random, where the top applicants reject job offers, and where the probability of job acceptance is negatively correlated with quality. Murphy (1986) and others (Hogarth & Einhorn, 1976) argue that the most reasonable assumption is a negative correlation between ability and probability to accept offers. The calculation thus requires information (or an assumption) about (a) the relationship between the quality of the applicants and the probability of accepting a job, and (b) the proportion of job offers being rejected. Note that if multiple adjustments are being applied, and particularly if considering an old selection battery rather than random hiring, the utility estimate requires incorporating the Murphy (1986) adjustment for both the new selection battery and the old selection battery.

The Present Study

Current UA research suggests that (a) the basic UA formula has flaws, and (b) improvements exist to modify the formula. This paper focuses on the effects that five algebraic UA adjustments—economic variables, employee flows, probationary periods, existing selection batteries, and rejected job offers— have on utility estimates. To date, no study has systematically examined the effects of multiple UA adjustments applied simultaneously.

By themselves, each of the UA adjustments relaxes an assumption of the basic UA model. Each UA adjustment should yield a lower result than produced by the basic UA model. This should occur either by directly reducing the utility estimate or by increasing the utility estimate for both the intervention in question (e.g., a new selection battery) and the previous condition (e.g., the current selection battery or a policy of random selection). However, the magnitude of such changes will also depend on the many factors entering into the utility equation. Even in cases where there is high validity, combinations of organizational practices and environmental circumstances may make such interventions a poor investment. Thus, although there may be generalizable findings of device validity, or an intervention may be heralded as a “best practice” (e.g., Huselid, 1995; Pfeffer, 1994; Terpstra & Rozell, 1993), the modified utility equations suggest that the value of such practices will depend on a variety of situational parameters.

Method

Computer simulation was used to examine the effects of various combinations of adjustments in circumstances where the parameters required by the utility model varied. The simulation generates parameters required by the basic UA model and the five algebraic adjustments described above. The simulator yielded data for two sets of analyses. The first set, a usefulness analysis, examines the relative magnitude of each of the adjustments. The second set provides a better illustration of the implications by applying the adjustments to the UA example used by Latham and Whyte (1994; Whyte & Latham, 1997).

Parameter values were chosen to represent a wide range of realistic levels, shown in Table 1. The values are based on the type of values used in the initial description of UA adjustments (e.g., Boudreau, 1983a; Murphy, 1986; De Corte, 1994), the range of values from previous UA applications (e.g., see Boudreau, 1991), results of meta-analyses (e.g., Williams & Livingstone, 1994), and levels reported in HR texts (e.g., Gatewood & Feild, 1994). Note that the intent of this paper is not to provide a literature review of UA parameters, but to reveal the implications of the UA adjustments for a range of realistic scenarios. The values employed by the simulator cover a wide range of values, and thus yield diverse realistic environments for the simulation. For example, in the Boudreau (1991) review of UA studies, the number of people hired in these studies ranged from 1 to 225,731; however, 95% of the studies had values between 1 and 1100. Thus, this range of values was used to represent what would likely be observed in other UA applications.

The simulation can compute a utility estimate based on any combination of desired adjustments, specific parameter values, or randomly selected values from the ranges reported in Table 1. First, the simulator generates the values for each parameter. Each value is generated independently. Thus, the expected correlation between any two variables is zero; however, a large number of simulations are used to provide a comprehensive array of value combinations. Each of the values is randomly generated from a uniform distribution, except for the number of people selected and the cost per applicant. Examination of the Boudreau (1991) review of UA studies showed that these variables were skewed (i.e., there were mostly lower values, but a significant tail of the distribution extending to large values, resembling distributions of personal income). Thus, to reflect these distributions, uniform variables were first generated, and then the exponential of those values were used. Uniform distributions were employed to ensure representation of all values throughout the distribution. Because the distributions were chosen to represent a range of likely values, there was no reason to force the simulation to choose values predominantly from the center of the distribution as would have occurred had a normal distribution been used.

Table 1. Parameter Values for Utility Analysis Calculations

Parameter	*Needed by adjustment*	*Range of Values*
Number of years:	0	1 to 10
Number of people selected:	0	1 to 1100
Applicants per hire:	0	1 to 50
SDy:	0	\$4,000 to \$40,000
Cost per applicant:	0	\$0 to \$1100
Validity of new battery:	0	0.05 to 0.38 above the old (from 0.10 to 0.77)
Discount rate:	1	0.01 to 0.11
Tax rate:	1	0.30 to 0.63
Variable cost rate:	1	−0.02 to −0.35
Probability of turnover:	2	0.00 to 0.33
Correlation between performance and turnover:	2	0.00 to −0.50
Stability (correlation) of performance over time:	3	0.50 to 1.00
Cutoff is a probationary period (standardized performance score):	4	0 to −2
Validity of old battery:	5	0.05 to 0.38
Percent accepting initial offer:	6	20% to 70%
Correlation between probability of accepting initial job offer and performance:	6	0.00 to −0.50

Notes: (0) Basic UA formula; (1) Economic adjustments; (2) Employee flows; (3) Temporal validity; (4) Probationary period; (5) Multiple selection devices; (6) Deviations from top-down hiring. All distributions are uniform, except for number of people selected and cost per applicant, which are exponential values from an originally uniform distribution.

Once the parameters have been generated, the simulator computes the utility estimate for the basic utility model and for any specified combination of adjustments. If the results from multiple adjustments are desired (e.g., the researcher wanted to see the result of applying (a) the economic adjustments, (b) the adjustment for employee flows, and (c) the result of applying both economic and employee flow adjustments), the program computes the multiple utility estimates for each set of parameters. The user of the simulator also specifies how many sets of parameters to generate.

Usefulness Analysis

To investigate the relative contribution of the various adjustments on resulting utility estimates, and to inform the literature of the practical implications of these adjustments, we performed a form of *usefulness analysis* (Darlington, 1968). A usefulness analysis generally examines a predictor's contribution to explaining unique variance in a criterion beyond another predictor's (or set of predictors') contribution. Although generally applied to multiple regression analysis, it provides a useful framework for examining the relative strengths of the various adjustments. The method reveals which adjustments have the largest impact, both initially and after other adjustments have been applied. The results of the procedure provide a logic to apply when selecting adjustments if not all adjustments can or need to be used: in order of the relative practical significance of each UA adjustment.

The procedure entailed using each adjustment in the UA calculation and noting the strength of the effect. Because the distributions of the effects were heavily skewed, it made more sense to examine the median effect as a measure of adjustment strength rather than mean effect, which could be biased by a few extreme values. After applying all five of the adjustments independently, the adjustment with the largest median effect would then be applied in addition to each of the remaining four individual adjustments. Of these four combinations of two adjustments, the added adjustment with the largest median effect is again selected, and the process repeats with combinations of three adjustments. The process continues until the relative strength of each adjustment, after applying the various larger-effect adjustments, has been estimated.

The simulator generated 10,000 UA scenarios. For each scenario, the simulator generated values for each of the variables listed in Table 1. Each variable was generated independently of the other variables, and each scenario's variables were generated independently of the previous scenarios. Using the data for each scenario, utility estimates were computed using the five adjustments applied independently. Table 2 presents the median and mean effects.

In column one of Table 2, all adjustments are compared to the estimate of the basic UA formula. Accounting for economic variables had the largest median (and mean) effect. Thus, this analysis suggests that, for a variety of organizational circumstances, the Boudreau (1983a) alternations have the greatest practical significance on resulting utility estimates. The simulator then generated utility estimates (still using the same 10,000 cases) using the economic variable adjustment in addition to each of the four other adjustments. The second column of Table 2 depicts the effect of implementing the remaining four adjustments both in comparison to the basic utility model and beyond that accounted for by the economic variable adjustments. The adjustment for considering the implications of multiple selection devices had the largest median effect and thus was selected as the next most "useful" adjustment. The procedure was repeated, and the importance of the other variables were ranked, with the (a) economic adjustments and (b) inclusion of multiple devices followed by (c) departures from top-down hiring, (d) a probationary period, and (e) employee flows.

After applying all five adjustments, the median effect size of the total set of adjustments was 291%, and the mean effect was 298%. This suggests that, although the majority of resulting utility estimates for the simulated HR interventions are still positive, the modifications to the basic UA model have sizable and noteworthy practical effects on the resulting estimates.

Table 2. Usefulness Analysis of Utility Analysis Adjustments: Net and Relative Effects

	Number of Modifications Being Applied								
	1	*2*		*3*		*4*		*5*	
		a	*b*	*a*	*b*	*a*	*b*	*a*	*b*
Modification Being Applied	*Net Effect*	*Net Effect*	*Over Previous Modification*	*Net Effect*	*Over Previous Modification*	*Net Effect*	*Over Previous Modification*	*Net Effect*	*Over Previous Modification*
Economic variables	−64%								
	−64%								
Multiple Devices	−53%	−84%	−53%						
	−59%	−84%	−53%						
Deviations from	−23%	−73%	−23%	−89%	−25%				
top-down hiring	−38%	−76%	−34%	−92%	−37%				
Effect of a	−22%	−72%	−21%	−87%	−17%	−91%	−8%		
probationary period	−25%	−72%	−24%	−88%	−21%	−94%	−10%		
The effect of	−1%	−65%	−1%	−84%	−1%	−90%	−2%	−91%	0%
employee flows	−9%	−65%	−7%	−87%	−8%	−96%	−1%	−98%	−1%

The first row in each group shows the median effect of adding the specified utility analysis modification, beyond those already added, signified by the grayed areas; the second row is the mean. Modifications were added in order of the size of the median effect over the previous modification. Column one shows the effect of implementing each modification by itself. Columns 2a, 3a, 4a, and 5a show the effect of implementing the combination of modifications over an unmodified utility analysis. Columns 2b, 3b, 4b, and 5b show the effect of the modification over the previously applied set of modifications.

Note that the minimum total effect of applying all of the adjustments simultaneously (i.e., the largest utility estimate provided after all five modifications were applied) was to decrease the estimate by 71%. Furthermore, the modified estimates were negative 16% of the time. This supports the view that the utility adjustments substantially decrease the utility estimate. Of greater importance, these results suggest that while valid selection devices may often lead to increased organizational utility, this result is not necessarily generalizable. Rather, the estimated value of a selection battery is contingent on organizational and environmental characteristics.

Illustrating the Implications: The Latham and Whyte Example

The usefulness analysis (a) demonstrates that these adjustments, on average, have large effects on resulting utility estimates, and (b) presents information that may inform UA users on which adjustments have the largest effects. However, it would be desirable to provide the UA with more intervention-specific and organizational- specific data to yield a distribution of utility estimates specific to a given intervention. Although the usefulness analysis shows the expected value of changes to utility estimates to be large, it is tenuous to generalize these expected value changes to specific UA examples. Indeed, although these results inform us that HR interventions do not universally have high utility, neither do they say that all interventions have poor utility. Rather, it depends on a variety of situational parameters. Thus, to help illustrate the effects of the UA adjustments on an actual UA example, this paper examines the UA case employed by Latham and Whyte (1994; Whyte & Latham, 1997) in their experiments.

Although, as discussed earlier, it is often widely accepted that estimates from the basic UA formula are unrealistic (e.g., Cascio, 1993), Latham and Whyte (1994; Whyte & Latham, 1997) investigated how managers react to the results of a basic UA formula. In two separate experiments, the authors presented a group of managers with results of a UA that revealed a large financial return for implementing a valid selection procedure. Despite the UA suggesting a $601 million return for a HR program, those provided the UA were actually less likely to want to implement the selection procedure than those in other experimental groups that had less information. These results occurred when UA was explained to the managers (Latham & Whyte, 1994) and even when an expert provided additional information through a presentation (Whyte & Latham, 1997).

The UA example used by Latham and Whyte reports a 14,000% return on investment. This paper argues that this return is unrealistically high, and that it may have been the lack of credibility of these results that led to managers' avoidance of this recommendation. Latham and Whyte (1994; Whyte & Latham, 1997) used the simple UA model to calculate the utility in their studies. Both studies used the same example, which is taken largely from another published UA (Cronshaw, 1986); however, they updated the dollars to reflect 1992 prices and changed the validity from the original 0.52 to 0.40 to present "a more conservative" (Latham & Whyte, 1994: 38), and perhaps more believable, validity coefficient. In the example, an organization hires 470 of 1410 applicants, each new hire stays with the company an average of 18 years, SD_y equals $16,280, and the total cost of the selection procedure is $429,110. The utility calculation is thus as follows:[i]

$$\Delta U = 18 \times 470 \times 1.09 \times .4 \times 16290 - 429110$$

$$\Delta U = \$59{,}657{,}532$$

While there is no specific evidence to support or contradict this estimate, the result lacks face validity. In the present study, the UA adjustments are applied to this example to help illustrate their implications and to show how they yield a distribution of estimates that may be more believable.

Because the UA adjustments require additional information than was provided by Latham and Whyte (1994; Whyte & Latham, 1997) or Cronshaw (1986), the simulator generated a range of cases with the needed additional variables. For each scenario, the simulator generated values for the missing parameters. Again, for each of the missing parameters, values were determined randomly based on the distribution of values reported in Table 1. Additionally, the validity of the old selection battery was also varied, coming from a random uniform distribution from 0.05 to 0.38. Again, each variable was generated independently of the other variables. Values for the variables not generated by the simulator (i.e., years, number of people selected, applicant per hire, SD_y, cost per applicant, and device validity) were taken from the Latham and Whyte (1994) example. To illustrate the implications of these adjustments for a wide range of environmental and organizational conditions, the simulator generated 10,000 scenarios.

For each scenario, the UA adjustments are applied in the order determined by the previous usefulness analysis. Table 3 shows the distribution of utility estimates that result when applying one to all five of the UA modifications to the Latham and Whyte example. As expected, the first step had a noteworthy impact: the adjustment for economic variables reduced the mean estimate by $42,800,529. Examining the range of utility estimates for the 10,000 scenarios generated by the simulator, the Boudreau (1983a) adjustments reduced the utility estimate by over half of the initial estimate (i.e., $30,000,0001) for over 98.5% of the estimates involved.

The remaining adjustments had smaller, but still noticeable, effects on the final utility estimate. The mean estimated utility dropped $9.3 million when the existence of multiple selection devices was considered, and another $4.4 million when deviations from top-down hiring were also considered. The adjustments for including a probationary period decreased the mean expected utility by a further $0.5 million, and accounting for employee flows dropped the mean utility estimate by another $0.5 million. Overall, using all five UA adjustments reduced the original estimate of $59,657,532 to the mean utility estimates of $2,228,170: more than a 96% reduction. The median return was $1,738,861; the smallest outcome was an estimated loss of $2,577,238; and the largest predicted gain was $17,162,452; or over 71% less than the initial estimate.

Conclusions

The above results demonstrate the magnitude that adjustments to the UA formula have on utility estimates. These results show that one of the major criticisms of UA, that it yields unrealistically large estimates, may in part be a reflection of using an overly simplistic formula. Although the UA adjustments require greater complexity, such adjustments sizably change UA estimates, even upwards of millions of dollars.

The results from this study lead to two contributions for UA and HR research. First, the results reveal the relative effect that each of the published algebraic modifications to the UA formula have, and thus present some guidance for researchers and practitioners deciding upon whether certain adjustments should be included. Although the magnitude of the changes are a function of organizational and environmental factors, these results suggest that economic modifications have the largest average impact, relatively closely followed by the effect of considering multiple devices, followed by weaker effects for deviations from top-down hiring, a probationary period for new employees, and considering employment flows.

Note that while this usefulness analysis provides insights into general steps UA researchers or users should follow (i.e., including the first two modifications would likely have a large practical impact, and the other modifications may also have large impacts depending on the circumstance), these results do not suggest a single set of modifications that should be blindly implemented. Rather, users of UA should consider their specific needs and adapt UA to meet those needs. These needs may be a function of the how complex the UA can be, the nature of the HR intervention, the business environment, and/or the audience for the UA results. Thus, UA should not follow one method blindly. These results show that all UA modifications have noteworthy effects, with some effects noticeably larger than others; it is up to the user of the UA, though, to use the modifications that are appropriate.

Morrow and colleagues (Morrow et al., 1997) provide a superb example of this. When using a UA to help evaluate the effectiveness of training programs, they admittedly used an approximation that included error. Their point was that “[human resource] decisions need to be based upon good information; perfectly accurate information is not critical” (Morrow et al., 1997: 113), nor is it even possible as “any field application of UA will rely upon an effect size calculated with an imperfect quasiexperimental design” (Morrow et al., 1997: 114). They thus conclude that managerial acceptance of a UA model is particularly important. The results of the present study can inform future work like Morrow et al.’s by informing researchers and practitioners of the potential effect of UA model adjustments, and thus helping them agree upon a utility model to implement. The present study’s results suggest that if the decision context does involve variables relevant to these adjustments, then such adjustments may make sizable changes to the UA estimate. This study’s results do not suggest that all adjustments need to be invariably applied; however, they do reveal that such adjustment should not be dismissed due to a flawed assumption that such adjustments would have minor effects on subsequent results.

Table 3. Frequency Distributions of Utility Analysis Estimates for Various Adjustments to Latham and Whyte’s (1994) Utility Analysis

UA estimates in millions of dollars	*Economic Variables*	*and multiple selection devices*	*and deviations from top-down hiring*	*and probationary period*	*and employee flows*
−3.0 to −2.0					2
−2.0 to −1.0					28
−1.0 to 0.0					617
0.0 to 1.0		71	1143	1428	1924
1.0 to 2.0		784	1952	2154	1988
2.0 to 3.0		800	1760	1857	1694
3.0 to 4.0		754	1440	1479	1274
4.0 to 5.0		717	1120	1062	897
5.0 to 6.0		770	831	730	624
6.0 to 7.0	1	763	582	495	382
7.0 to 8.0	45	778	428	305	243
8.0 to 9.0	173	746	268	196	126
9.0 to 10.0	302	667	183	107	84
10.0 to 11.0	472	606	119	83	51
11.0 to 12.0	624	529	73	47	35
12.0 to 13.0	744	439	41	24	13
13.0 to 14.0	782	383	21	17	7
14.0 to 15.0	840	299	15	6	7
15.0 to 16.0	836	258	13	3	2
16.0 to 17.0	776	163	6	5	1
17.0 to 18.0	692	129	2	1	1
18.0 to 19.0	668	98	2	1	
19.0 to 20.0	583	76	1		
20.0 to 21.0	490	53			
21.0 to 22.0	403	45			
22.0 to 23.0	335	33			
23.0 to 24.0	289	30			
24.0 to 25.0	213	10			
25.0 to 26.0	187	13			
26.0 to 27.0	152	12			
27.0 to 28.0	106	7			
28.0 to 29.0	90	2			
29.0 to 30.0	62	1			
30.0 to 31.0	45				
31.0 to 32.0	36				
32.0 to 33.0	22				
33.0 to 34.0	17				
34.0 to 35.0	9				
35.0 to 36.0	6				
36.0 to 37.0					
Mean	$ 16.9	$ 7.6	$ 3.2	$ 2.7	$ 2.2

Note: All values expressed in millions. Unadjusted UA yields $59.7 million estimate.

The second contribution of this paper is to reveal that the estimated impact of human resource interventions is dependent on a number of organizational-specific and environmental variables. The magnitude of these changes to the UA estimates suggests that the trend of past published UA demonstrations to invariably suggest large returns for HR investments is inappropriate. This supports the view that the value of HR interventions for organizational performance are contingent on organizational and environmental characteristics, as opposed to the notion that there is a set of “best practices” which invariably contribute to organizational performance. Thus, evidence of validity or effect sizes from meta-analyses, or list of “best practices” from certain organizations, should not be interpreted as evidence of utility. Rather, this paper suggests that the value of HR interventions may at least in part be situationally specific.

Modifications of the UA model, like those described and applied in the present study, are important for potentially improving the accuracy of the measurement technique; however, the ultimate goal of UA applications is to influence decision making. To date, UA has not had its intended effect on decision making (Boudreau, 1996; Cascio, 1996). If UA is to become a useful decision aid, research is needed in a number of areas to help overcome this general trend. First, laboratory studies like Latham and Whyte’s can contribute to this goal; however, such research needs to move beyond simply looking at end-results of artificial decision environments and help understand the cognitive processes involved when people react to decision aids, particularly in human resources. Although such research would lose the contextual and organizational factors which would affect real-world decision making, the controlled laboratory environment could help yield a better understanding of how individuals process UA information, and therefore how the technique or training can be adjusted to improve human resource decision making.

Second, research needs to examine UA implementations for actual decisions (Boudreau, 1996). Although less controlled than a laboratory study, such research can provide needed information on how actual decision-makers use UA to justify or make decisions. Such research should pay attention to the effects that decision contexts and organizational characteristics have when judging managerial reactions to UA results.

To this effect, research is also needed to help specify the types of environments in which HR interventions are being implemented. This study’s analyses used simulation to represent a range of realistic situations, but when applied to actual decision making, better approximations of the additional required variables would lead to greater accuracy. Hence, descriptive research is needed to provide a clearer picture of the level of these crucial variables in organizations. Questions such as “how many applicants apply for jobs?" and "what is the relationship between predicted performance and the probability of accepting a job?" are important issues not only for UA research, but for staffing research in general. Furthermore, this study assumed zero correlation between the simulation parameters. Research is also needed on the relationships between contextual variables for human resource decision making.[ii] If such intercorrelations significantly deviate from 0, then utility estimates may be notably different. Future research should therefore look into the nature of these relationships and the implications of such relationships on UA estimates.

There are still a number of limitations thwarting the widespread usefulness of UA. While this study shows how the UA adjustments can provide substantially different estimates, we are still a long way from UA being a useful tool for managers. Notably, there are two fundamental problems: (a) the complexity involved in UA computation, and (b) the lack of training of HR managers. These problems are a reflection of technology and training, and not inherently a flaw of UA. UA, even in its basic form, is often viewed as complex. Some have criticized the inclusion of just the economic variables (Boudreau 1983a) as "so long and complex as to be daunting . . . and as such is very difficult for personnel psychologists and HR managers to understand" (Hunter et al., 1988: 527). Incorporating the adjustments of UA discussed in this paper essentially eliminates the possibility of performing UA by hand. Some adjustments, such as those discussed by De Corte (1994) for estimating the effect of a probationary period, require computations beyond the capability of most spreadsheet programs. It should be noted, though, that just as it was not surprising that techniques, such as structural equations modelling, were not popular before computer programs were available to solve the necessary equations, researchers should not be surprised that sophisticated utility models are not employed when paper and pencils are often the only tools available for UA computation. Technology must be developed that facilitates UA, including the adjustments discussed in the research.

But perhaps more important than technology, managers need to understand UA and be trained in the use of the technology. In the Latham and Whyte (1994) study, "none of the managers . . . had taken courses on utility analysis, HR accounting, or validation procedures" (p. 36), and in the Whyte and Latham (1997) study, "none of [the managers] had been exposed to utility analysis prior to the study" (p. 603). Although it is likely that most managers are not educated in UA procedures, this is a fault of HR education (or perhaps staffing and training in HR positions) rather than a fault of UA. It is incumbent on researchers of UA to explore how utility estimates affect decision making so as to inform UA research of ways to make the device more of a decision making aid (Boudreau, 1996).

For a complex decision making tool to be useful, the users of the decision aid must desire the information it provides and be trained in its use. We should not be surprised that an individual untrained with a use of a decision aid fails to adhere to the results of the aid. For example, we would not be surprised to find that managers with no experience in economics, accounting, and finance would not trust the recommendations of the Merton and Scholes options pricing formula (e.g., Black & Scholes, 1973). This formula is as complex, if not more complex, than UA but is extensively used and has had an impact that merited it winning the Nobel prize in economics (e.g., *Economist*, 1997).

As HR must increasingly justify its use of resources and demonstrate its value to other functions in the organization, HR managers must become able to communicate the merit of HR interventions. As educators and researchers of HR, we must advance the level of knowledge in the field and keep managers (and future managers) abreast of these advances to provide them with the tools they need for the future.

References

Ashe, R. L., Jr. 1990, April. The legal defensibility of assessment centers and in-basket exercises. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Miami Beach, FL.

Barrett, G. V., Caldwell, M. S., & Alexander, R. A. 1989. The predictive stability of ability requirements for task performance: A critical reanalysis. Human Performance, 2: 167–181.

Becker, B. E. 1989. The influence of labor markets on human resource utility estimates. Personnel Psychology, 42: 531–546.

Black, F., & Scholes, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81: 637–654.

Boudreau, J. W. 1983a. Economic considerations in estimating the utility of human resource productivity improvements. *Personnel Psychology*, 36: 551–557.

Boudreau, J. W. 1983b. Effects of employee flows on utility analysis of human resource productivity improvement programs. *Personnel Psychology*, 68: 396–407.

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, Volume 1*: 621–745. Palo Alto, CA: Consulting Psychologist Press.

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.

Brogden, H. E. 1949. When testing pays off. *Personnel Psychology*, 2: 171–183.

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

Cascio, W. F. 1993. Assessing the utility of selection decisions: Theoretical and practical considerations. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations*: 310–340. San Francisco, CA: Jossey-Bass.

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* (14th ed.). Boston, MA: Kent.

Cronbach, L. J., & Glesser, G. C. 1965. *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.

Cronshaw, S. F. 1986. The utility of employment testing for clerical/administrative trades in the Canadian Military. *Canadian Journal of Administrative Sciences*, 3: 376–385.

Cronshaw, S. F. 1997. Lo! The stimulus speaks: The insider's view on Whyte and Latham's "The Futility of Utility Analysis." *Personnel Psychology*, 50: 611–616.

Darlington, R. B. 1968. Multiple regression in psychological research and practice. *Psychological Bulletin*, 69: 161–182.

De Corte, W. 1994. Utility analysis for the one-cohort selection-retention decision with a probationary period. *Journal of Applied Psychology*, 79: 402–411.

Economist. 1997. The right option: The Nobel Prize for economists. October 18: 75.

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

Gatewood, R. D., & Feild, H. S. 1994. *Human resource selection* (2nd ed.). Chicago, IL: Dryden Press.

Henry, R. A., & Hulin, C. L. 1987. Stability of skilled performance across time: Some generalizations and limitations on utilities. *Journal of Applied Psychology*, 72: 457–462.

Hogarth, R. M., & Einhorn, H. J. 1976. Optimal strategies for personnel selection when candidates can reject job offers. *Journal of Business*, 49: 479–495.

Hulin, C. L., Henry, R. A., & Noon, S. L. 1990. Adding a dimension: Time as a factor in the generalizability of predictive behaviors. *Psychological Bulletin*, 107: 328–340.

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

Huselid, M. A. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, 38: 635–672.

Judiesch, M. K., Schmidt, F. L., & Hunter, J. E. 1993. Has the problem of judgment in utility analysis been solved? *Journal of Applied Psychology*, 78: 903–911.

Latham, G. P., & Whyte, G. 1994. The futility of utility analysis. *Personnel Psychology*, 47: 31–46.

Morrow, C. C., Jarrett, M. Q., & Rupinski, M. T. (1997). An investigation of the effect and economic utility of corporate-wide training. *Personnel Psychology*, 50: 91–119.

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.

Pfeffer, J. 1994. *Competitive advantage through people: Unleashing the power of the work force*. Boston, MA: Harvard Business School Press.

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

Raju, N. S., Burke, M. J., & Maurer, T. J. 1995. A note on direct range restriction adjustments in utility analysis. *Personnel Psychology*, 48: 143–149.

Raju, N. S., Burke, M. J., & Normand, J. 1990. A new approach for utility analysis. *Journal of Applied Psychology*, 77: 234–250.

Raju, N. S., Burke, M. J., Normand, J., & Lezotte, D. V. 1993. What would be if what is wasn't? Rejoinder to Judiesch, Schmidt, and Hunter (1993). *Journal of Applied Psychology*, 78: 912–916.

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

Terpstra, D. E., & Rozell, E. J. 1993. The relationship of staffing practices to organizational level measures of performance. *Personnel Psychology*, 46: 27–48.

Whyte, G., & Latham, G. P. 1997. The futility of utility analysis revisited: When even an expert fails. *Personnel Psychology*, 50: 601–610.

Williams, C. R., & Livingstone, L. P. 1994. Another look at the relationship between performance and voluntary turnover. *Academy of Management Journal*, 37: 269–298.

Wintermantel, R. E., & Mattimore, K. L. 1997. In the changing world of human resources: Matching measures to mission. *Human Resource Management*, 36: 337–342.

---

[i] In the original Cronshaw (1986) example, and in Latham and Whyte's examples, Zx was set equal to 1.10. For a selection ratio of 1/3, however, the value is actually closer to 1.09. We used the value 1.09 to reflect the more accurate value, and thus our results deviate slightly from those of Latham and Whyte.

[ii] This idea was proposed by one of the anonymous reviewers.