

RESEARCH REPORTS

Is Utility Really Futile? A Failure to Replicate and an Extension

Kenneth P. Carson, John S. Becker, and John A. Henderson
University of Tennessee at Chattanooga

G. P. Latham and G. Whyte (1994) found that managers gave less positive evaluations to a proposal to implement a valid selection procedure when information about the utility of that procedure was also given. The authors hypothesized that the manner in which the utility information was presented might explain this result. Two empirical studies ($Ns = 145$ and 186) were conducted. Results of mean contrast analyses show that the effect found by Latham and Whyte was not replicated, that revised scenarios including utility information are more understandable to managers, and that utility information presented in a revised manner has a low-to-moderate positive effect on the acceptability ratings that managers give to these selection proposals. However, acceptability ratings remain disappointingly low no matter which scenario is presented. Future study is encouraged to examine managerial reactions to innovations suggested by psychological research.

A general question facing all professionals, including industrial/organizational (I/O) psychologists, is how best to communicate the important findings of a discipline to individuals who are in a position to use that information but who are not familiar with the intricacies of the research that led to those conclusions. Miller (1969) made this point very well for psychologists in his memorable American Psychological Association (APA) presidential address. His comments focused on the challenges and difficulties facing psychologists who were in a position to give the important findings of scientific psychology away to nonpsychologists.

Many researchers who have focused on the calculation of the utility of human resource practices have assumed

that managers want such information (e.g., Cascio, 1991b). However, some have questioned this assumption. Latham (1988) wondered whether the consumers of organizational psychology really desire an economic justification for the interventions that they consider. This thought received some empirical support when Macan and Highhouse (1994) conducted a survey that found that roughly half of the I/O psychologists and human resource practitioners surveyed did not find that their clients expressed interest in utility information.

Latham and Whyte's (1994) study on managers' reactions to utility information directly addressed this issue. They found that managers who were presented with utility data concerning a proposed selection practice were less likely to support the selection practice than if they were given the same proposal without the utility information. They summarized the inference of these findings by concluding that "industrial-organizational psychologists are advised to reconsider their assumptions regarding the information managers use when managers make human resource policy decisions" (p. 31). This finding has influenced thinking about utility. For instance, in reviewing the Latham and Whyte findings, Borman, Hanson, and Hedge (1997) observed, "these troubling results have stimulated a great deal of discussion" (p. 321). The purpose of our studies was to further investigate managerial reactions to utility information.

Latham and Whyte (1994) presented a sample of 143 experienced managers attending an executive master of business administration (EMBA) program with a written scenario that described a human resource selection problem in a hypothetical corporation. They varied the type

Kenneth P. Carson, John S. Becker, and John A. Henderson, Department of Psychology, University of Tennessee at Chattanooga. John A. Henderson is now at the Department of Management, University of Tennessee, Knoxville.

We thank Mike Biderman, Rich Metzger, Jeffrey Ryer, and Greg Stewart for helpful comments on previous versions of this article. We also thank the numerous faculty at several universities who assisted us with data collection. Study 1 is based on a master's thesis completed by John S. Becker, and Study 2 is based on a master's thesis completed by John A. Henderson, both under the direction of Kenneth P. Carson. An earlier version of this article was presented in 1997 at the 12th Annual Meeting of the Society for Industrial and Organizational Psychology, St. Louis, Missouri.

Correspondence concerning this article should be addressed to Kenneth P. Carson, Department of Psychology, University of Tennessee at Chattanooga, Chattanooga, Tennessee 37377. Electronic mail may be sent to ken-carson@utc.edu.

of information that the managers received in four ways: validity information only, a combination of validity and expectancy table information, a combination of validity and utility information (which we will refer to as validity + utility), and a combination of validity, expectancy, and utility information (see Table 1 for a summary of the scenarios used in each study). The dependent variable was an eight-item scale assessing the extent to which the participant-managers would accept the proposal including the likelihood that they would commit organizational resources to its implementation. The significant main effect for type of information showed a difference between only the validity-only and validity + utility conditions such that the proposal was rated higher when only validity information was presented. Latham and Whyte concluded that managers respond negatively to utility information when they do receive it.

Such a conclusion has important consequences. If accurate, consultants, both internal and external to the organization, would be wise to stay away from utility analyses when attempting to persuade their clients to adopt human resource innovations. Although research as to the efficacy of such innovations would remain entirely appropriate, persons in positions of giving psychology away (Miller, 1969) to managers would be advised to refrain from communicating the utility evidence that buttresses their recommendations and rely instead primarily on nondollar-based arguments. In the research arena, future investigators may rely on this conclusion as they conduct further research on managerial decision making (Borman et al., 1997; Hazer & Highhouse, 1997).

Although their conclusion is certainly tenable and is consistent with their findings, it is not the only possible explanation for their results. The Latham and Whyte (1994) explanation is that managers react negatively to something central to utility analysis. Among the possible items to which managers may negatively react are the algebraic formulas involved, the description of the manner in which standard deviation in dollars (i.e., SD_y) is calculated, or the size of the utility estimates themselves. In contrast, we hypothesized a different explanation for their

results, namely that the manner in which the utility information was presented led to their counterintuitive result.

There is a strong conceptual rationale for hypothesizing that the manner in which the proposals are presented will make a difference to the managers who evaluate them. Dutton and Ashford (1993) proposed a framework for analyzing the manner in which issues are "sold" to top management. Many researchers have argued that the process of selling an issue to top management is an important determinant of the decision to "buy" the issue. They advise that "the more an issue seller portrays an issue succinctly, the greater the level of top management's attention invested in an issue" (Dutton & Ashford, 1993, p. 416). O'Reilly (1983) made a similar argument in a paper on organizational decision making. He argued that information must be readily accessible and summarized in order for good quality decisions to be made and that "only limited amounts of information from any set are likely to be used for decision making purposes" (p. 126). Petty and Cacioppo (1986) made a similar argument in the context of a more general framework for understanding persuasive communications. One prediction of their elaboration likelihood model is that strong arguments will be more favorably received to the extent that complexity is reduced and comprehensibility increased. Similarly, flawed arguments will be more negatively evaluated when complexity is reduced and comprehensibility is increased (Petty & Cacioppo, 1986, p. 76). Taken together, these reviews suggest that the manner in which a problem or a solution to a problem is presented to individuals influences the reaction that they have to it. Specifically, material that is shorter and easier to understand will be, or at least can be, more persuasive and lead to better quality decisions (see also Chervany & Dickson, 1974; Lyles, 1987; Oskamp, 1965; Sweller, Chandler, Tierney, & Cooper, 1990).

It is therefore appropriate to examine the content of the experimental materials used by Latham and Whyte (1994) in light of these conceptual perspectives and evidence. Because the content of these materials is central to the present investigation, they are included in the Appendix. Latham and Whyte's validity-only proposal contains information about only the process of conducting a validation study but does not include any quantitative information whatsoever. The scenario comprises 365 words. Latham and Whyte's validity + utility condition appears next in the Appendix. The first four paragraphs are identical to the validity-only condition followed by the utility information. In total, this proposal is more than four times as long as the validity-only proposal (1,550 words). Furthermore, it is considerably more complex than the validity-information section. This proposal contains the actual utility formula and an extended description of the process of determining SD_y estimates using managerial judgments. Latham and Whyte report that much of the utility informa-

Table 1
Summary of the Scenarios Used in Each Study

Scenario	Latham and Whyte	Study 1	Study 2
Validity only ^a	X	X	X
Validity + expectancy ^a	X		
Validity + utility ^a	X	X	X
Validity + expectancy + utility ^a	X		
Revised validity only		X	
Revised validity + utility		X	X
SD_y -enhanced validity + utility			X

^a This scenario was authored by Latham and Whyte (1994).

tion was taken directly from Cronshaw's (1991) undergraduate textbook and from Cascio's (1991b) description of the calculation of SD_y . Their rationale for using these sources was to minimize experimenter bias. However, the effect of using these sources was to make the presentation considerably longer and much more analytically complex than the validity-only materials.

We hypothesized that the negative reaction that managers had toward the utility materials (relative to the validity-only materials) was not a function of the essential content of the utility information but was rather due to the manner of its presentation. In addition, we predicted that utility information that was shorter and easier to understand would be more favorably received. Latham and Whyte (1994) did note in the discussion of their study that subject-matter comprehension was a possible alternative explanation for their results. Because they did not include a measure of comprehension in their materials, it was not possible for them to test that possibility. They pointed to the nature of the source of the material (undergraduate texts) and the fact that none of their participants reported any comprehension problems or asked any questions of the experimenter as indicators that this alternative explanation was, in their view, unlikely. However, they also noted that issues such as this are "best resolved through further empirical investigation rather than by debate" (p. 44).

Study 1

Method

Sample. The sample consisted of 145 managers attending EMBA programs at three different universities (two public, one

private) in the southeastern United States. Table 2 summarizes the demographic information for this sample along with the similar information from the Latham and Whyte (1994) study. The samples are quite similar in terms of age and experience. Our sample had a significantly higher percentage of women, 31% versus 20%, $\chi^2(1, N = 141) = 4.49, p < .05$. Because prior academic exposure to concepts relating to human resource accounting, utility, and validity might influence the results, we had participants self-report whether these topics had ever been discussed in a course. By way of comparison, Latham and Whyte reported that no one in their sample had taken a course devoted to any of these topics.

Design. Study 1 used four scenarios to be compared with linear contrasts (Howell, 1997). The first two scenarios were a replication of the stimulus materials presented by Latham and Whyte (1994) in their validity-only and validity + utility cells. The third and fourth scenarios consisted of revised validity-only and validity + utility information. These materials are described below. Only one of the four scenarios was evaluated by each participant-manager.

We elicited the cooperation of instructors of three different EMBA programs. We randomly distributed the four scenarios to EMBA students during class or other meeting times. Because materials were distributed in a random fashion in each of the three universities, sample sizes for the four cells were slightly uneven, ranging from 34 to 40. Data for two participants were incomplete for the acceptability dependent measure. The data obtained from these two participants were included in the summary of demographic characteristics and analyses concerning the understandability dependent variable but were excluded from the analyses involving the acceptability dependent measure.

Materials. The materials used in the Latham and Whyte presentation condition were identical to Latham and Whyte's (1994) original experimental materials (provided to us by La-

Table 2
Demographic Characteristics of the Three Samples

Characteristic	Latham and Whyte	Study 1	Study 2
Total sample	143	145	186
University setting			
EMBA	143	145	92
MBA	0	0	94
Experience			
Human resources	51	61	49
Finance or accounting	44	52	79
Education			
Courses			
HRA	0	—	7
Utility	0	—	18
Validity	0	—	13
Discuss			
HRA	—	23	37
Utility	—	35	50
Validity	—	40	53
Gender			
Male	114	97	126
Female	29	44	60
Age (mean)	39.0	37.3	32.5
Work experience (mean in years)	16.0	15.7	9.9

Note. EMBA = executive master of business administration; MBA = master of business administration; HRA = human resource accounting. Dash indicates that these data were not collected for this sample.

tham) with two exceptions. First, we changed the description of the company in the scenario from "a large Canadian based multinational corporation" to "a large American based multinational corporation." Second, we removed the consultant's affiliation with the Canadian Psychological Association from the description of the consultant's experience (we retained the affiliation with the American Psychological Association, which was also present in Latham and Whyte's original materials). Both changes were made because our sample of managers were attending U.S. universities.

We altered the materials in the revised presentation condition in such a way that the material was shorter and easier to comprehend. We did not make many changes for the revised validity-only condition. Most significantly, we gave an expected validity coefficient for the proposed validity study (.40) and placed it in the context of validity coefficients generally (usually not higher than .50) and of typical validity coefficients for unstructured interviews (.20). See the Appendix for the actual wording. We chose the .40 value because that was the value used in the other three conditions of Latham and Whyte's study. The .20 value, which we chose, represents the average of the two major meta-analyses that report average validity coefficients for traditional, unstructured interviews (Hunter & Hunter, 1984; McDaniel, Whetzel, Schmidt, & Maurer, 1994). The revised validity-only proposal is 485 words in length.

In contrast to the relatively minor adjustments made to the revised validity-only materials, we made major changes in the revised validity + utility proposal. In terms of raw length, we reduced Latham and Whyte's 1,550-word proposal to 772 words, a reduction of greater than 50%. Perhaps as important as mere reduction in verbiage was the deletion of the technical jargon. We deleted the utility derivations and formulas as well as information concerning the specifics of the determination of SD_y . We broke the financial gain estimate down into individual gain per year, cohort gain per year, and, finally, total gain for the entire cohort over an 18-year period (the average tenure of clerical employees at this company). We altered this material in order to clarify the issues for managers not trained in the intricacies of utility analysis and to enable managers to better understand the benefits of improved selection practices. It is important to emphasize, however, that we did not change the magnitude of the quantitative content of the utility analysis (such as the validity coefficient, SD_y estimates, and dollar estimates). For example, the utility gain of \$60,208,786 over an 18-year time span that Latham and Whyte used in their original materials was also given in the revised validity + utility proposal.

Measures. This study used two primary dependent measures. The first was identical to Latham and Whyte's (1994) eight-item scale designed to measure the extent to which managers accepted the consultant's proposal. Example items include "How likely are you to implement the consultant's recommendations?" and "How will others in the company react if you decide to implement the consultant's recommendations?" Each item had a 5-point Likert-type response scale with 1 representing the negative end of the scale. Thus, scores on this acceptability scale could range from 8 to 40. We found this scale to have an internal consistency reliability of .92 using Cronbach's (1951) alpha formula (Latham & Whyte reported a reliability of .93).

The second dependent measure assesses the participant-man-

agers' perception of their understanding of the proposal. This was a two-item scale, including "How well did you understand this consultant's proposal?" and "To what extent was the proposal clearly presented?" These items also had 5-point response scales with 1 anchoring the negative dimension so that possible scale scores range from 2 to 10. The internal consistency reliability of this scale is .82.

Results and Discussion

As indicated earlier, we used linear contrasts in all analyses. The descriptive statistics for these conditions are shown in Table 3. First, we hypothesized that the Latham and Whyte (1994) effect would be replicated. To test this, we gave the two Latham and Whyte scenarios weights of 1 and -1, respectively, with the remaining two scenarios being given zero weights. The effect was not replicated. The participant-managers evaluated the validity-only and validity + utility proposals equivalently, $t(139) = 0.42$, $p = .68$, $d = 0.09$.

We further hypothesized that the revised materials would be perceived to be easier to comprehend than Latham and Whyte's (1994) original materials. To test this hypothesis, we gave both Latham and Whyte scenarios weights of -1, and we gave the two revised scenarios weights of 1. This contrast was significant, $t(141) = 2.10$, $p = .037$, $d = 0.31$, and in the expected direction. The revised scenarios were more understandable than the original Latham and Whyte scenarios. An inspection of the cell means reveals that this difference is due to the validity + utility scenario; the Latham and Whyte validity-only scenario was not significantly different from the revised scenarios, $t(141) = 0.62$, $p = .539$, $d = -0.11$.

The final hypothesis, which is the main focus of the study, predicted that participant-managers evaluating the revised scenarios would give higher acceptability ratings to the validity + utility proposal compared with the validity-only proposal. Although the mean difference was in the expected direction, this contrast was not significant, $t(139) = 1.91$, $p = .058$, $d = 0.45$.

Study 2

The results of Study 1 partially supported the hypotheses. Specifically, participant-managers understood the revised scenarios more than they did Latham and Whyte's original validity + utility scenario. This was consistent with our observation that the Latham and Whyte (1994) scenarios were more complex than the revised scenarios.

However, three features of Study 1 suggested that further investigation was warranted. The first was the unexpected nonreplication of Latham and Whyte's (1994) original finding. The reason that the Latham and Whyte effect was not replicated in our sample when using their identical materials was not clear. The demographic composition of the sample, although substantially equivalent

Table 3
Descriptive Statistics for Study 1 Dependent Variables

Condition	n	Acceptability		Understandability	
		M	SD	M	SD
Latham and Whyte validity only	40	22.05	6.50	6.45	1.85
Latham and Whyte validity + utility	35	22.69	6.12	5.63	1.94
Revised validity only	34	22.24	6.41	6.50	2.12
Revised validity + utility	34	25.26	7.08	6.79	1.67

in many respects, was different in one important way. Our sample contained a number of participant-managers who reported that they had been educationally exposed to topics relating to human resource accounting, validity, or utility. Latham and Whyte reported that no one in their sample had a course devoted to these topics; they did not ask whether the topics had ever been discussed (see Table 2). This led to the suspicion that education led the participant-managers in the current study to be less negatively predisposed to the utility material in the Latham and Whyte presentation because they were more familiar with the issues. However, one-way analyses of covariance comparing the means of the Latham and Whyte validity-only and validity + utility conditions while holding educational exposure or self-reported understanding of the material constant did not change the nature of the effect. The means remained not significantly different. The sample size used in Study 1 provided plenty of power (0.92; Cohen, 1988) to detect the large effect that Latham and Whyte found, suggesting that sampling error also was not the source of the different experimental result. So we were left knowing only what the nonreplication was not due to, not what it was due to. In such a situation, it seemed reasonable to collect more data.

The second concern was that although Study 1 did not find that utility information had a negative effect on managers, it also did not have a significant positive effect in the set of revised scenarios as we had hypothesized. However, the marginal *p* value suggests that the effect, although probably not strong, may indeed exist. This indicated that further data collection was appropriate.

The third concern was that the content of the revised validity + utility scenario was written so that the participant managers were given much less information about the utility analysis procedures. In particular, the revised validity + utility scenario reduced the description of the *SD_y* estimation procedure to one phrase. It is possible that if the participant managers were aware of the nature of these estimation procedures that the results would be different.

Therefore, we conducted Study 2. We included the two original scenarios from Latham and Whyte (1994) in this second study to see if the effect that they found could be replicated. To address the third concern, we developed a

second revision of the validity + utility scenario. This scenario was identical to the first revision except for the addition of a paragraph that explains the *SD_y* estimation procedure.

Method

Sample. The sample for Study 2 consisted of 186 students from six different universities in the United States. The major difference between this sample and that of earlier samples was the inclusion of MBA as well as EMBA students. We did this to test whether MBA students respond differently than EMBA students. Demographic information concerning this sample is summarized in Table 2.

Design. Study 2 was a 4 (type of information) \times 2 (program) between-subjects design. The four levels of type of information were crossed with two levels of program (EMBA and MBA students). The dependent measures, identical to those in Study 1, were the eight-item acceptability and two-item understandability scales.

Materials. The materials used in Study 2 were identical to those used in Study 1 with two exceptions. First, we eliminated the revised validity-only condition. Second, we added an *SD_y*-enhanced revised validity + utility condition (see Table 1 for a comparison of the three studies). The enhancement consisted of the addition of a concise description of how the *SD_y* was obtained. This revision was necessary to investigate the claim that managers who are told about the way in which *SD_y* estimates are calculated will lower their acceptability ratings. In order to give a fair, balanced, and complete description of *SD_y* estimation, we took a description from a widely used textbook. This description was taken from Cascio's (1991a) undergraduate text *Applied Psychology in Personnel Management* (p. 302). This description is still considerably shorter than the one given by Cascio (1991b), which Latham and Whyte (1994) used. We used the shorter version in order to be consistent with our goal of making the proposals more understandable. All told, we expanded the material dealing with *SD_y* estimation from 1 phrase to 12 full lines. See the Appendix for the actual wording.

Results and Discussion

Means for each of the four conditions can be found in Table 4. On the basis of the results of Study 1, we conducted three planned contrasts. First, we compared the two Latham and Whyte (1994) scenarios (the two revisions were given zero weights). There was no significant

Table 4
Descriptive Statistics for Study 2 Dependent Variables

Condition	n	Acceptability		Understandability	
		M	SD	M	SD
Latham and Whyte validity only	47	22.60	5.98	6.04	1.98
Latham and Whyte validity + utility	41	22.12	6.24	5.54	1.95
Revised validity + utility	45	24.71	5.56	6.42	1.84
SD _y -enhanced validity + utility	53	24.60	5.49	6.11	1.92

difference between these two conditions, $t(182) = .38$, $p = .703$, $d = -0.08$. Therefore, the Latham and Whyte effect was not replicated once again.

Second, we compared the two revisions. We did not hypothesize a difference here although we were concerned that the deletion of almost all material relating to the estimation of SD_y may have changed the content of the utility scenario too much. This planned contrast (with the two Latham and Whyte scenarios given zero coefficients) was also not significant, $t(182) = .10$, $p = .927$, $d = 0.02$.

Next, we examined the difference in acceptability ratings between the Latham and Whyte validity + utility scenario (given a coefficient of -2) and the two validity + utility revisions (both with a coefficient of 1). This contrast was significant, $t(182) = 2.35$, $p = .02$, $d = 0.44$. Thus, we concluded that the manner of presentation does influence the way in which each scenario is evaluated. To investigate the reason for this difference, we conducted a planned contrast for the *understanding* dependent variable. With the same coefficient weights ($-2, 1, 1$), there is a significant difference in the degree to which the participant-managers reported that they understood the material, $t(182) = 2.36$, $p = .019$, $d = 0.37$.

Finally, we examined the difference in acceptability ratings between the Latham and Whyte validity-only scenario (given a coefficient of -2) and the two validity + utility revisions (both given a coefficient of 1). This comparison was significant, $t(182) = 2.00$, $p = .047$, $d = 0.35$, such that the revised validity + utility scenarios were given higher acceptability ratings compared with the validity-only version.

Summary

Taking the results of both studies along with the Latham and Whyte (1994) study, we were faced with a set of conflicting findings. The most obvious conflict was that the results of the two studies reported here did not replicate the surprising effect that Latham and Whyte found, even though identical materials were used and even though the participant-managers seemed to be taken from the same population. The second conflict was that Study 1 found that the higher ratings given to the revised validity

+ utility scenario were not statistically significant given traditional criteria, whereas Study 2 found that these higher ratings were statistically significant. The effect sizes for comparisons in each of the three studies are summarized in Table 5. Clearly, there is a discrepancy between the results obtained by Latham and Whyte and those reported in our two studies. Further, the revised presentation of validity and utility information seems to be preferred by participant-managers compared with either the validity-only or the Latham and Whyte presentation.

Conclusion

The results of these two studies lead to several conclusions. First, the effect found by Latham and Whyte (1994) has not been replicated in two subsequent studies. Power analysis suggests that it is extremely unlikely that this is the result of sampling error. Further, this failure to replicate resulted after using the same scenarios that they used and relying primarily on participants that were very close to the type of participants that they used (experienced managers enrolled in EMBA educational programs). Clearly, there is some other reason for the disparate results, which has not been identified. We are left to speculate as to the reason for this disparity. It is possible that demand characteristics may explain the result. For example, Latham and Whyte wrote that Latham has "known biases . . . regarding utility analysis" (Latham & Whyte, 1994, p. 39), and therefore Whyte was responsible for designing the stimulus materials and for data collection. However, because the data appear to have been collected from students in a program with which both Latham and Whyte were affiliated, it is possible that, despite the precautions taken to minimize demand characteristics, the study participants had been exposed to Latham's self-professed bias. In contrast, the participants in the two studies reported here came from multiple university programs, none of which any of us, the present authors, were or are affiliated with. We do not know that this is the correct explanation, only that it is one possibility. Another possibility is that there is some difference between Canadian and American managers that affects these results. We think that this is unlikely, but not impossible.

Table 5
Summary of Sample and Effect Sizes

Comparison ^a	Latham and Whyte	Study 1	Study 2	T ^b	N ^c
B-A	-.75	.09	-.08	474	158.0
C-A	—	.49	.36	331	165.5
C-B	—	.39	.44	331	165.5

Note. Dash indicates that the comparisons were not available in this study.

^a A = Latham and Whyte validity-only condition; B = Latham and Whyte validity + utility condition; C = Revised validity + utility condition. ^b Total sample size. ^c Average sample size.

A second conclusion is that presenting information to managers in a more user-friendly manner results in some improvement in the rated acceptability of the selection proposals. The size of this effect is in the low-medium range in Cohen's (1988) scheme (*ds* of 0.45 and 0.37 in the two studies). In addition, adding more description about the *SD*_y estimation procedures did not lead to a decline in acceptability of the proposal at all, although this particular conclusion is based only on the data in Study 2.

Thus, the results obtained here suggest that utility information, presented in a user-friendly manner, may add to the efficacy of a proposal to improve a selection system. In particular, managers were not put off by the size of the utility estimate alone, because the same estimate appears in both versions. This should be welcome news to persons at all points on the scientist-practitioner continuum. Researchers who study the adequacy of utility models should be heartened that the work in which they have been engaged is not seen as entirely useless (or worse) by practicing managers. Human resource practitioners and consultants should be encouraged to look for ways to include utility data in their presentations to organizations considering making changes in their selection systems. However, because the managers who participated in these studies were not making real-world decisions, caution should be used in generalizing these results to actual decision-making contexts.

This upbeat perspective is tempered by a view of the data that should not be overlooked. The means for all of the scenarios are in the 22-26 range on an 8-40 scale, the midpoint of which is 24. This means that when combining all the conditions together, managers on average were not particularly impressed with the proposed new selection system. Even the highest mean of the two studies reported here, which occurred in the revised validity + utility condition for Study 1, had a magnitude of only 25.26. These values are quite similar to those obtained in the Latham and Whyte study. In fact, the highest value from the Latham and Whyte study (26.4) is not significantly different from the highest value in the present investigation, 25.26, $t(68) = 0.723$, $p = .30$. So although we find that utility information is not necessarily negatively viewed, there remains considerable opportunity for

the acceptance ratings of selection proposals to be improved. There are a number of possible factors that may lead the proposals to be more positively viewed. For instance, Hazer and Highhouse (1997) recently found that managers appear to prefer utility estimates based on the 40% rule compared with those based on *SD*_y estimates, such as those used in the studies reported here. In addition, Macan and Highhouse's (1994) survey results suggest that some managers may want utility information whereas others may not. This suggests that individual difference variables may be a fruitful avenue of further study. Further conceptual analysis and empirical study are required in order to identify what these factors might be.

The present results are consistent with the theoretical propositions reviewed in the introduction (Dutton & Ashford, 1993; O'Reilly, 1983; Petty & Cacioppo, 1986). In a complex decision environment, decision makers attend to and use information more when it is presented in a manner that is shorter in length and higher in comprehensibility. Of particular interest is Petty and Cacioppo's elaboration likelihood model, which suggests that strong arguments will be more positively evaluated and weak arguments will be more negatively evaluated when they are better understood. The fact that understandability and acceptability ratings both increased suggests that utility estimates are strong evidence. Of course, none of these factors has an absolute value. That is, a proposal is not perfectly comprehensible—it is only more or less comprehensible than some alternative. Future research should examine ways in which proposals dealing with human resource issues can be presented in an optimal fashion. As noted above, there remains considerable room for improvement in managers' evaluation of these materials. In addition, future research might attempt to more directly measure a manager's comprehension of the utility information (along the lines of Hazer and Highhouse's, 1997, study) or, alternatively, it might be appropriate to use a more qualitative data-gathering technique to assess what managers are actually thinking about when presented with these proposals.

In conclusion, we return to Miller's (1969) APA presidential address. He argued that society would benefit if psychologists would learn to appropriately communicate the findings of scientific psychology to laypersons who

are in a position to use them. We close with a quote from this seminal address:

Psychological principles and techniques can usefully be applied by everyone. If our suggestions actually work, people should be eager to learn more. If they do not work, we should improve them. But we should not try to give people something whose value they cannot recognize, then complain when they do not return for a second meeting. (Miller, 1969, p. 1073)

References

- Borman, W., Hanson, M., & Hedge, J. (1997). Personnel selection. *Annual Review of Psychology*, 48, 299–337.
- Cascio, W. F. (1991a). *Applied psychology in personnel management* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Cascio, W. F. (1991b). *Costing human resources: The financial impact of behavior in organizations* (3rd ed.). Boston: PWS-Kent Publishing Co.
- Chervany, N. L., & Dickson, G. W. (1974). An experimental evaluation of information overload in a production environment. *Management Science*, 20, 1335–1344.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Cronshaw, S. F. (1991). *Industrial psychology in Canada*. Waterloo, Ontario, Canada: North Waterloo Academic Press.
- Dutton, J. E., & Ashford, S. J. (1993). Selling issues to top management. *Academy of Management Review*, 18, 397–428.
- Hazer, J. T., & Highhouse, S. (1997). Factors influencing managers' reactions to utility analysis: Effects of SD, method, information frame, and focal intervention. *Journal of Applied Psychology*, 82, 104–112.
- Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury Press.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–98.
- Latham, G. P. (1988). Human resource training and development. *Annual Review of Psychology*, 39, 545–582.
- Latham, G. P., & Whyte, G. (1994). The futility of utility analysis. *Personnel Psychology*, 47, 31–46.
- Lyles, M. A. (1987). Defining strategic problems: Subjective criteria of executives. *Organization Studies*, 8, 263–279.
- Macan, T. H., & Highhouse, S. (1994). Communicating the utility of human resource activities: A survey of I/O and HR professionals. *Journal of Business and Psychology*, 8, 425–436.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive meta-analysis. *Journal of Applied Psychology*, 79, 599–616.
- Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist*, 24, 1063–1075.
- O'Reilly, C. A. (1983). The use of information in organizational decision making: A model and some propositions. *Research in Organizational Behavior*, 5, 103–139.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29, 261–265.
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.
- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, 119, 176–192.

(Appendix follows)

Appendix

Stimulus Materials

Latham and Whyte's Validity-Only Condition and their Validity + Utility Condition are printed with permission from the authors.

Latham and Whyte's Validity-Only Condition

Imagine that you are a vice president of a large American based multi-national corporation. The company has over 10,000 employees, and last year's sales exceeded \$4 billion dollars. The company is highly regarded for the quality of its products. Lately, however, some people within the organization have raised concerns about the quality of the clerical/administrative personnel that the firm has been hiring.

Potential new hires are interviewed before a decision is made to extend an offer, but at present no systematic procedures exist by which clerical/administrative employees are selected. Although you are skeptical that the methods by which your firm selects such personnel can be improved upon, an organizational consultant has been retained to investigate the issue.

This consultant specializes in the development and validation of selection practices that attempt to enable companies to select high performers in a legally defensible manner. She is also a member of the American Psychological Association, and graduated ten years ago with a Ph.D. from a prestigious university.

This particular consultant was contacted because she has done research showing that employee performance can be predicted through the use of selection testing. After a discussion of the issues, the consultant has recommended that the following steps be taken:

1. Develop a test to measure employee performance that is tailor-made to your firm;
2. Administer this test across all relevant categories of clerical/administrative personnel;
3. Correlate the results of the test with performance on the job to see how well the test predicts performance;
4. Use the results of this test to alter existing selection practices so as to improve employee performance in the clerical/administrative category.

The consultant has stated that the cost of designing and validating the selection test will be \$6,100. Because you expect to hire about 470 new clerical/administrative employees this year from a large pool of applicants it is estimated that the total implementation cost of the selection program will be \$423,000. This amount reflects recruiting costs, test administrators' salary, computer test

scoring, and outlays for test booklets. Total costs of the consultant's recommendations, if followed, are therefore approximately \$429,100.

As vice president of the company, it is up to you to decide whether to implement the consultant's recommendations.

Latham and Whyte's Validity + Utility Condition

Imagine that you are a vice president of a large American based multi-national corporation. The company has over 10,000 employees, and last year's sales exceeded \$4 billion dollars. The company is highly regarded for the quality of its products. Lately, however, some people within the organization have raised concerns about the quality of the clerical/administrative personnel that the firm has been hiring.

Potential new hires are interviewed before a decision is made to extend an offer, but at present no systematic procedures exist by which clerical/administrative employees are selected. Although you are skeptical that the methods by which your firm selects such personnel can be improved upon, an organizational consultant has been retained to investigate the issue.

This consultant specializes in the development and validation of selection practices that attempt to enable companies to select high performers in a legally defensible manner. She is also a member of the American Psychological Association, and graduated ten years ago with a Ph.D. from a prestigious university.

This particular consultant was contacted because she has done research showing that employee performance can be predicted through the use of selection testing. After a discussion of the issues, the consultant has recommended that the following steps be taken:

1. Develop a test to measure employee performance that is tailor-made to your firm;
2. Administer this test across all relevant categories of clerical/administrative personnel;
3. Correlate the results of the test with performance on the job to see how well the test predicts performance;
4. Use the results of this test to alter existing selection practices so as to improve employee performance in the clerical/administrative category.

The consultant has stated that the cost of designing and validating the selection test will be \$6,100. Because you expect to hire about 470 new clerical/administrative employees this year from a large pool of applicants it is

estimated that the total implementation cost of the selection program will be \$423,000. This amount reflects recruiting costs, test administrators' salary, computer test scoring, and outlays for test booklets. Total costs of the consultant's recommendations, if followed, are therefore approximately \$429,100.

The consultant estimates that the approximate financial gain achievable if her advice is followed is \$60,208,786. This amount represents her estimate of the dollar gain from the use of new selection procedures for one year, as compared to continuing to use existing methods.

The algebraically derived model the consultant used to calculate this amount is relied on by many industrial psychologists today and was developed and refined by researchers over many years. The model is as follows:

$$DU_R = t(N_s SD_y r_{xy} l / f) - N_s C_{is} - N_s C_{os}$$

where U_R is the dollar gain from the use of the selection test for one year as compared to selecting people using existing procedures; t is the average length of time the selectees stay with the organization; N_s is the number of people to be selected this year, SD_y is the standard deviation of job performance in dollars; r_{xy} is the correlation between test scores and job performance as determined by previous research; l is the height of the normal curve at the cutting point on the predictor (a statistic which is obtained from the f or selection ratio); f is the selection ratio (the proportion of people selected as compared to the number of people who applied for the job); C_{is} is the implementation cost of the selection program per person selected; and C_{os} is the original cost of designing and validating the selection instrument per person selected.

To compute the dollar gain, input estimates for the terms in the model were obtained from research studies and your firm's personnel records, and then entered into the above equation. The input estimates were as follows:

1. The average length of time new hires in the clerical/administrative category stay with the firm, or t is 18 years;
2. The number to be selected this year (N_s) is 470;
3. SD_y Refers to the standard deviation of job performance in dollars. The larger the difference in how well people perform on the job as measured by SD_y , the higher the potential gain from using selection practices that will distinguish between poor and good performers. While most of the input estimates for the model were obtained from the firm records, SD_y cannot be obtained in this way but must be estimated. The SD_y term is difficult to estimate, but it must be estimated in order to express in dollars the potential gain to be achieved from following the recommendations. SD_y was calculated using supervisors of clerical/administrative personnel. In estimating SD_y , the instructions to the supervisors were as follows:

The dollar utility estimates we are asking you to make are critical in estimating the relative dollar value to the firm of different selection methods. In answering these questions, you will have to make some very difficult judgments. We realize they are difficult and that they are judgments or estimates. You will have to ponder for some time before giving each estimate, and there is probably no way you can be absolutely certain your estimate is accurate when you do reach a decision. But keep in mind three things:

1. The alternative to estimates of this kind are cost accounting procedures to the evaluation of job performance. Such applications are usually prohibitively expensive. And, in the end, they produce only imperfect estimates, like this estimation procedure.

2. Your estimates will be averaged in with those of other supervisors of clerical/administrative employees. Thus errors produced by too high and too low estimates will tend to be averaged out, providing more accurate final estimates.

3. The decisions that must be made about selection methods do not require that all estimates be accurate down to the last dollar. Substantially accurate estimates will lead to the same decisions as perfectly accurate estimates.

Based on your experience with clerical/administrative employees, we would like to estimate the yearly value to the firm of the products and services provided by the average clerical/administrative employee. Consider the quality and quantity of output typical of the average employee and the value of this output, it may help to consider what the cost would be of having an outside firm provide these products and services.

Based on my experience, I estimate the value to the firm of the average clerical/administrative employee at _____ dollars per year.

We would now like for you to consider the "superior" clerical/administrative employee. Let us define a superior performer as some one who is at the 85th percentile. That is, his or her performance is better than 85% of his or her fellow employees, and only 15% turn in better performances. Consider the quality and quantity of the output typical of the superior employee. Then estimate the value of these products and services. In placing an overall dollar amount on this output, it may again help to consider what the cost would be of having an outside firm provide these products and services.

Based on my experience, I estimate the value to the firm of the superior clerical/administrative employee at _____ dollars per year.

Finally, we would like you to consider the "low-performing" clerical/administrative employee. Let us define a low-performing employee as one who is at the 15th percentile. That is, 85% of all clerical/administrative em-

employees turn in performances better than the low-performing employee, and only 15% turn in worse performances. Consider the quality and quantity of the output typical of the low-performing employee. Then estimate the value of these products and services. In placing an overall dollar amount on this output, it may again help to consider what the cost would be of having an outside firm provide these products and services.

Based on my experience, I estimate the value to the firm of the low-performing clerical/administrative employee at _____ dollars per year.

SD_y was calculated as the average of the estimated difference in dollar value of yearly job performance between superior and average clerical/administrative employees, and between average and low-performing clerical/administrative employees. SD_y , the average of these two estimations was \$16,290. Therefore according to this approach, $SD_y = \$16,290$.

4. The correlation between performance on the test and job performance was estimated from previous research to be $r_{xy} = .40$.

5. The height of the normal curve at the cutting point on the predictor is .3621. The selection ratio F (the number of people selected expressed as a proportion of the number of people who applied for the job) was conservatively estimated at .33 (one selectee for every three applicants), therefore $1/F = 1.10$;

6. The implementation cost of the selection program per person selected, or C_{is} , was estimated at \$900;

7. The original cost of designing and validating the selection instrument per person selected or C_{os} , was estimated as \$13.

When these estimates are entered into the equation given above, the resulting dollar gain was calculated as approximately \$60,208,786. This figure represents the estimated net returns from using new selection procedures for one year when the returns are aggregated over the average length of time that the new selectees are predicted to stay with the firm.

$$DU_R = 18 (470 \times 16,290 \times .40 \times 1.10) \\ - 470 (900) - 470 (13)$$

$$DU_R = \$60,208,786.$$

As vice president of the company, it is up to you to decide whether to implement the consultant's recommendations.

Revised Validity-Only Condition

Imagine you are the vice president of human resources in a large American based multi-national company. This corporation employs over 10,000 people and had sales of over \$4 billion dollars last year. The corporation has a

reputation for high quality products. Recently concerns have been raised about the quality of clerical/administrative personnel.

Potential new hires are interviewed before a decision is made to extend an offer, but at present no systematic procedures exist by which clerical/administrative employees are selected. Although you are skeptical that the methods by which your firm selects such personnel can be improved upon, an organizational consultant has been retained to investigate the issue.

This consultant specializes in the development and validation of selection practices that attempt to enable companies to select high performers in a legally defensible manner. She is also a member of the American Psychological Association, and graduated ten years ago with a Ph.D. from a prestigious university.

This particular consultant was contacted because she has done research showing that employee performance can be predicted through the use of selection testing. After a discussion of the issues, the consultant submits a preliminary proposal. This is a summary of her proposal:

Interviews like the ones your company uses have traditionally had lower levels of validity than test based selection systems. Validity is simply an expression of the relationship between something used to predict job performance and actual job performance. A validity coefficient of zero means there is no relationship; a coefficient of 1.00 means there is a perfect relationship. It is very rare for validity coefficients to exceed .50 in a selection context. The higher the validity coefficient, the more able we are to accurately predict who will be successful on the job. The average validity for the type of interview your company is using is approximately .20, whereas a test designed for another client yielded a coefficient of .40, which is consistent with research evidence for these types of tests.

For your company, a validity study consisting of the following steps will be performed:

1. Develop a test to measure employee performance that is tailor-made to your firm;
2. Administer this test across all relevant categories of clerical/administrative personnel;
3. Correlate the results of the test with performance on the job to see how well the test predicts performance;
4. Use the results of this test to alter existing selection practices so as to improve employee performance in the clerical/administrative category.

The cost of designing and validating the selection test will be \$6,100. Because you expect to hire about 470 new clerical/administrative employees this year from a large pool of applicants we estimated that the total implementation cost of the selection program will be \$423,000. This amount reflects recruiting costs, test administrators' sal-

ary, computer test scoring, and outlays for test booklets. Total costs are approximately \$429,100.

As vice president of the company, it is up to you to decide whether to implement the consultant's recommendations.

Revised Validity + Utility Condition

Imagine you are the vice president of human resources in a large American based multi-national company. This corporation employs over 10,000 people and had sales of over \$4 billion dollars last year. The corporation has a reputation for high quality products. Recently concerns have been raised about the quality of clerical/administrative personnel.

Potential new hires are interviewed before a decision is made to extend an offer, but at present no systematic procedures exist by which clerical/administrative employees are selected. Although you are skeptical that the methods by which your firm selects such personnel can be improved upon, an organizational consultant has been retained to investigate the issue.

This consultant specializes in the development and validation of selection practices that attempt to enable companies to select high performers in a legally defensible manner. She is also a member of the American Psychological Association, and graduated ten years ago with a Ph.D. from a prestigious university.

This particular consultant was contacted because she has done research showing that employee performance can be predicted through the use of selection testing. After a discussion of the issues, the consultant submits a preliminary proposal. This is a summary of her proposal:

Interviews like the ones your company uses have traditionally had lower levels of validity than test based selection systems. Validity is simply an expression of the relationship between something used to predict job performance and actual job performance. A validity coefficient of zero means there is no relationship; a coefficient of 1.00 means there is a perfect relationship. It is very rare for validity coefficients to exceed .50 in a selection context. The higher the validity coefficient, the more able we are to accurately predict who will be successful on the job. The average validity for the type of interview your company is using is approximately .20, whereas a test designed for another client yielded a coefficient of .40, which is consistent with research evidence for these types of tests.

For your company, a validity study consisting of the following steps will be performed:

1. Develop a test to measure employee performance that is tailor-made to your firm;
2. Administer this test across all relevant categories of clerical/administrative personnel;
3. Correlate the results of the test with performance on the job to see how well the test predicts performance;

4. Use the results of this test to alter existing selection practices so as to improve employee performance in the clerical/administrative category.

The cost of designing and validating the selection test will be \$6,100. Because you expect to hire about 470 new clerical/administrative employees this year from a large pool of applicants we estimated that the total implementation cost of the selection program will be \$423,000. This amount reflects recruiting costs, test administrators' salary, computer test scoring, and outlays for test booklets. Total costs are approximately \$429,100.

These costs should be evaluated in the context of the return the company can expect to receive. If the validity study which we are proposing does in fact indicate that the new test has a validity coefficient of about .40 (which is expected based on past experience and research), this would mean that more productive employees are being selected than is currently the case. More productive employees are obviously beneficial since they produce more for the company for the same labor costs as their less productive counterparts. Utility analysis is the process of estimating the benefit to the company of selecting better, more productive, employees.

A utility analysis was conducted based on information gained from supervisors in your company about the value of more productive employees in this particular job category, the expected validity of the new test, the number of applicants for the jobs, and on the cost estimates of administering the new selection procedure. This analysis indicates that the average clerical/administrative employee will be "worth" \$7,117 more to the organization than the average employee selected under the current system because the better employees hired based on the new procedure will learn their jobs more quickly and continue to improve as new policies and technology are put in place. Since the company expects to hire 470 employees in this job category over the course of the year, this savings equals \$3,344,933 ($470 \times 7,117$) per year to the company. Further, since employees in this job category stay with the company for an average of 18 years, the total benefit to the company over the life span of the employees selected this year is estimated to be \$60,208,786 ($18 \times 3,344,933$). In sum, the utility analysis shows that improved selection has an impressive economic benefit to the company.

As vice president of the company, it is up to you to decide whether to implement the consultant's recommendations.

SD_y-Enhanced Revised Validity + Utility Condition

Imagine that you are a vice president of a large American based multi-national corporation. The company has over 10,000 employees, and last year's sales exceeded \$4

billion dollars. The company is highly regarded for the quality of its products. Lately, however, some people within the organization have raised concerns about the quality of the clerical/administrative personnel that the firm has been hiring.

Potential new hires are interviewed before a decision is made to extend an offer, but at present no systematic procedures exist by which clerical/administrative employees are selected. Although you are skeptical that the methods by which your firm selects such personnel can be improved upon, an organizational consultant has been retained to investigate the issue.

This consultant specializes in the development and validation of selection practices that attempt to enable companies to select high performers in a legally defensible manner. She is also a member of the American Psychological Association, and graduated ten years ago with a Ph.D. from a prestigious university.

This particular consultant was contacted because she has done research showing that employee performance can be predicted through the use of selection testing. After a discussion of the issues, the consultant submits a preliminary proposal. This is a summary of her proposal:

Interviews like the ones your company uses have traditionally had lower levels of validity than test based selection systems. Validity is simply an expression of the relationship between something used to predict job performance and actual job performance. A validity coefficient of zero means there is no relationship; a coefficient of 1.00 means there is a perfect relationship. It is very rare for validity coefficients to exceed .50 in a selection context. The higher the validity coefficient, the more able we are to accurately predict who will be successful on the job. The average validity for the type of interview your company is using is approximately .20, whereas a test designed for another client yielded a coefficient of .40, which is consistent with research evidence for these types of tests.

For your company, a validity study consisting of the following steps will be performed:

1. Develop a test to measure employee performance that is tailor-made to your firm;
2. Administer this test across all relevant categories of clerical/administrative personnel;
3. Correlate the results of the test with performance on the job to see how well the test predicts performance;
4. Use the results of this test to alter existing selection practices so as to improve employee performance in the clerical/administrative category.

The cost of designing and validating the selection test will be \$6,100. Because you expect to hire about 470 new clerical/administrative employees this year from a large pool of applicants we estimated that the total implementation cost of the selection program will be \$423,000. This

amount reflects recruiting costs, test administrators' salary, computer test scoring, and outlays for test booklets. Total costs are approximately \$429,100.

These costs should be evaluated in the context of the return the company can expect to receive. If the validity study which we are proposing does in fact indicate that the new test has a validity coefficient of about .40 (which is expected based on past experience and research), this would mean that more productive employees are being selected than is currently the case. More productive employees are obviously beneficial since they produce more for the company for the same labor costs as their less productive counterparts. Utility analysis is the process of estimating the benefit to the company of selecting better, more productive, employees.

A utility analysis was conducted based on the expected validity of the new test, the number of applicants for the jobs, the cost estimates of administering the new selection procedure, and the information gained from supervisors in your company about the value of more productive employees in this particular job category. The value of more productive employees was obtained utilizing the Schmidt-Hunter global estimation procedure. This method is based on the following reasoning: If job performance in dollar terms is distributed normally, then the difference between the value to the organization of the products and services produced by the average (50th percentile) employee and those produced by an employee at the 85th percentile in performance is equal to SD_y . To facilitate these judgments, raters are told to imagine how much the goods and services would cost if provided by an outside consulting firm. The magnitude of SD_y then is inferred from the difference between these two estimates.

This analysis indicated that the average clerical/administrative employee will be "worth" \$7,117 more to the organization than the average employee selected under the current system because the better employees hired based on the new procedure will learn their jobs more quickly and continue to improve as new policies and technology are put in place. Since the company expects to hire 470 employees in this job category over the course of the year, this savings equals \$3,344,933 (470×7117) per year to the company. Further, since employees in this job category stay with the company for an average of 18 years, the total benefit to the company over the life span of the employees selected this year is estimated to be \$60,208,786 ($18 \times 3,344,933$). In sum, the utility analysis shows that improved selection has an impressive economic benefit to the company.

As vice president of the company, it is up to you to decide whether to implement the consultant's recommendations.

Received March 7, 1997

Revision received July 9, 1997

Accepted July 9, 1997 ■