

A Comparison Of Evaluation Techniques For Decision Analysis Involving Large Attribute Sets

Ronald L. Thompson, (Email: ron.thompson@mba.wfu.edu), Wake Forest University
Charalambos L. Iacovou, (Email: charles.iacovou@mba.wfu.edu), Wake Forest University
Larry E. Shirland, (Email: larry.shirland@uvm.edu), University of Vermont

ABSTRACT

This study investigates the effectiveness of two multi-attribute evaluation techniques under conditions of high information load that is caused by large attribute sets. One hundred and sixty-five respondents were randomly assigned to two groups: the first one used a holistic, point allocation-based method to evaluate a list of 20 job attributes, while the second employed a triad-based technique that decomposed the evaluation task. The results suggest that the decomposed method produced more reliable results and was deemed easier to use, even though it took longer to complete the task.

INTRODUCTION

Individual decision-makers and organizational teams are frequently tasked with evaluating a set of attributes as part of their routine responsibilities. Such multi-attribute evaluations are common in organizational settings and span various contexts: steering committees prioritizing competing project proposals (Jessop, 2002), planning committees selecting among various strategic plans, administrators evaluating diverse budget requests (Shen, Lo and Wang, 1998), individual product managers deciding on which product qualities to focus, and many others.

To assist decision makers, researchers have carried out extensive investigations on multi-attribute decision making methods (Huber, 1974; Kim and Choi, 2001). A primary goal of this research stream is to identify effective evaluation methods that can be used to select among competing alternatives and to prioritize lists of attributes (Harte and Koele, 1995; Jaccard, Brinberg and Ackerman, 1986).

While a large diversity of multi-attribute evaluation techniques exists, such methods can be categorized into two groups: *holistic* and *decomposed*. Holistic methods consist of simple evaluation processes that consider all attributes in the decision set simultaneously while allowing decision-makers to explicitly state their preferences by assigning importance weights to them; decomposed methods, on the other hand, break down the evaluation task into a series of judgment-based choices that focus on a subset of the attributes at a time and then utilize statistical techniques to infer the implicit importance weights from the decision-makers' choices (Schoemaker and Waid, 1982; Vriens, Opperwal and Wedel, 1998).

The primary research question for our study is whether or not the use of a decomposed weight elicitation technique (that breaks down the evaluation task) will lead to more effective decision making under conditions of high information load that is caused by large attribute sets. To address this issue, results from the use of a holistic, point-allocation method are compared to those of a decomposed one in which attributes are presented in triads and the attribute weights are derived using goal programming. These two techniques were selected for our study because they represent the chief traits of each method type (simplicity for holistic methods and task complexity reduction for decomposed ones).

MULTI-ATTRIBUTE EVALUATION METHODS

Holistic evaluation techniques are simple processes that enable decision-makers to explicitly assign an indication of importance (using ratings, points, etc.) to each of the attributes in the set under consideration. In these methods, the decision-maker is asked to consider all attributes in the set *simultaneously* when assigning importance weights to them (Schoemaker and Waid, 1982).

Several holistic evaluation methods are being used in organizations today. The most common ones include Direct Ratings, in which the attributes are rated for their perceived importance using an interval scale, and Point Allocation (PA), in which the decision-maker is asked to distribute a fixed number of points (usually 100) among the various attributes so as to reflect their relative importance (Wittink et al., 1994).

To assess the effectiveness of holistic evaluation techniques in assessing large attribute sets, we decided to include the PA method in our study. This was done for two reasons. First, this method is among the most popular and frequently used ones (Doyle et al., 1997; Eckenrode, 1965; Jaccard et al., 1986; Wittink et al., 1994). Second, it is representative of the simplicity and time efficiency that is inherent in direct methods.

Decomposed evaluation methods break down the evaluation task into a series of comparative, judgment-based choices that are to be made by the decision-maker (Schoemaker and Waid, 1982). The attribute set is used to generate smaller comparison subsets (usually pairs) or scenarios that the decision-maker can use to select the most attractive choice among them. Statistical procedures are then employed to indirectly derive the importance weights that were implicit in the decision-maker's choices.

While decomposed techniques aim to reduce task complexity, the time requirements for comparing items in pairs can be quite significant for large attribute sets. For example, an evaluation of 25 attributes would require at least 300 pairwise comparisons, a highly time consuming task for most decision-makers. Moreover, the independent comparisons in decomposed judgments may result in conflicting choices and lack of transitivity (Flynn, Sakakibara, Schroeder, Bates, and Flynn, 1990).

To reduce the time requirements and intransitivity risk associated with pair-wise methods, a novel decomposed technique was recently introduced by Shirland et al. (2003). This technique utilizes triad-based judgments instead of pair-wise ones. To generate the importance weights, the triad-based (TB) technique employs a two stage process. First, an integer programming model is used to generate the set of triads that participants will use to rank each attribute in them. Next, the within-triad rankings are used as input to a quadratic goal programming model (using squared weighted Euclidean distances) to calculate attribute weights. (More details for the TB technique are provided in Appendix A).

To assess the effectiveness of decomposed evaluation methods in large set evaluations, we decided to include the above TB judgment method in our study. This approach decreases perceived task complexity (as it reduces the number of choice sets to be evaluated) and minimizes the number of potential intransitivities. As a result, it is likely to decrease frustration and tiredness among decision makers (Shirland et al., 2003). Moreover, for practical purposes, it is easy to execute (as it does not require the development of extensive scenarios and software exists to generate the triads from simple lists). In sum, we selected it because it is representative of the class of decomposed techniques and is easy to implement.

Extensive research has investigated the effectiveness of the various multi-attribute methods by comparing their results. The findings of this research stream are somewhat equivocal. Most studies have found significant differences across methods (Doyle et al., 1997; Jaccard et al, 1986; Schoemaker and Waid, 1982; Vriens et al., 1998), while some have not (Harte and Koele, 1995; Eckenrode, 1965). While the reconciliation of these differences is beyond the scope of our work, we offer a brief summary of the relevant issues related to the comparison of holistic and decomposed methods. For an extensive treatment of multi-attribute method differences, see Pöyhönen and Hämäläinen (2001).

As mentioned earlier, research indicates that holistic methods tend to be favored and are more frequently utilized in organizational settings than decomposed ones because of their simplicity and ease of use (Doyle et al., 1997; Eckenrode, 1965; Jaccard et al., 1986; Wittink et al., 1994). While straightforwardness is a common trait of holistic methods, past research has identified a number of issues that can cast doubt on their usefulness. First, it appears that holistic techniques do not always produce the same results. In fact, research shows that such differences across holistic methods may be systematic: Bottomley et al. (2000) found that the direct rating technique tends to produce linear weights, while point allocations result in nonlinear weights. Second, research suggests that explicit, direct evaluations (that are required by holistic methods) are likely to be biased and unreliable. Indeed, evidence suggests that direct evaluations tend to lead to systematic “overweighting” of unimportant attributes and “underweighting” of important ones (Ashton, 1974; Joyce, 1976).

Due to the above shortcomings of holistic evaluations, some researchers have concluded that asking participants to make choice-based judgments and deducing the attribute weights indirectly (using decomposed techniques) may be a better alternative (Vriens et al., 1998). In this paper, we propose that an additional reason for advocating the use of decomposed methods is the overwhelming cognitive burden that is imposed when evaluating a large number of attributes using holistic methods. Under such conditions of information overload, we believe that holistic methods may be less effective than decomposed ones.

RESEARCH HYPOTHESES

According to cognitive psychology, when humans engage in evaluative tasks, the items that are being compared must be placed in their memory (Simon, 1969). These chunks of information are stored and processed in a subset of our memory, the *working memory* (Galotti, 1999). Working Memory Capacity (WMC) researchers, however, have posited that the capacity of working memory is constrained.

An important function of working memory is to combine information that is received from the current environment with information that is retrieved from the past (such as prior experiences and related preferences that are stored in the long-term memory) so that individuals can select options and develop strategies. In the context of multi-attribute evaluations, we assert that holistic evaluations of large attribute sets are likely to exaggerate the problems associated with our limited working memory capacity by overloading decision-makers with information values and processing loads that they cannot adequately handle.

In PA evaluations, individuals are unlikely to be able to concurrently hold in their minds all attributes and their respective weights. Due to WMC limitations, those who try to do so are unlikely to be successful and are likely to produce unreliable results. As individuals get overwhelmed by the task (by recognizing that they cannot hold all of the information in their minds), they attempt to compensate by proactively decomposing the task on their own. This process, however, requires high cognitive power. According to Doyle et al. (1997) and Bottomley et al. (2000), such individuals must not only make judgments, but they must also keep a mental tally of how many points they have left to allocate to the remaining attributes. These dual demands (of concurrent assessment and dynamic point budgeting) are likely to exert a higher cognitive load on decision-makers. TB judgments, on the other hand, do not experience such tremendous cognitive demands because the task is decomposed for the evaluator into smaller “bite size” chunks. Thus the information processing requirements and strategies are likely to be different between the two methods.

Given the above differences between TB and PA in terms of complexity and cognitive loads, we expect that these two methods will not yield similar results in terms of attribute weights. This is consistent with prior studies that have shown systematic differences in evaluations across multi-attribute methods (Doyle et al., 1997; Jaccard et al., 1986; Pöyhönen and Hämmäläinen, 2001). As a result, we posit that:

H1: In large attribute set evaluations, the TB method will not produce the same results as the PA method.

Evidence suggests that when working with large attribute sets, the reliability of multi-attribute evaluations decreases (Pöyhönen and Hämmäläinen, 2001; Srivastava, Connolly and Beach, 1995; von Winterfeldt and Edwards,

1986). We suspect, however, that the degradation of performance (due to attribute set size) is likely to affect the two methods differently.

With large attribute sets, individuals that are considering all attributes simultaneously (in holistic evaluations) are more likely to be overwhelmed than individuals who consider each attribute in subsets (such as pairs or triads). For example, using the PA method to evaluate a set of 20 attributes a decision-maker would have to think about an inordinately high number of comparisons to simultaneously compare each of the attributes vis-à-vis the other 19 ones. Since this is a physically untenable task, individuals evaluating large sets using PA will tend to engage in complex cognitive processes in their working memory. These processes will include rapid mental comparisons (to contrast attributes against each other, resulting in large combination sets), the temporary storing of attributes and their respective tentative weights, and the dynamic computation of unallocated points at any given instance. This is likely to be too mentally taxing and is therefore likely to result in reduced task performance, especially when a task involves over 15 attributes (Malhotra, 1982).

By breaking up the task into triads -using Shirland et al.'s (2003) technique-, the evaluation task could be transformed into 69 sequential considerations of only 6 combinations (3!). Given the reduced demands on WMC by the TB method, we expect that individuals will be able to make more accurate, consistent assessments. Because of this, we anticipate that TB evaluations will outperform PA ones in terms of reliability. Thus, we hypothesize that:

H2: In large attribute set evaluations, the TB method will produce results that are more reliable than the ones produced by the PA method.

Due to the high and complex cognitive demands that are imposed on decision-makers by the PA method, we assert that PA evaluations are likely to be perceived as less favorable than TB ones. Holistic evaluations of large sets are likely to cause frustration and may be perceived as an arduous task because of the increased demands for the participants' cognitive processing. This is likely to be interpreted as a lack of consideration for the participants and can be resented (Bottomley, et al., 2000). As decision-makers are more likely to find triad-based judgments less cognitively taxing, we anticipate that:

H3: In large attribute set evaluations, the TB method will be perceived as easier to complete than the PA method.

In terms of the time needed to carry out an evaluation, it appears that PA may be a more efficient method. This is consistent with the results of several method comparison studies carried out by Eckenrode (1965). His findings suggest that holistic methods are more efficient than decomposed, judgment-based ones. Moreover, he found out that as the number of attributes increases so does the time efficiency advantage of holistic methods. According to Eckenrode (1965, p. 188), the time taken to complete an evaluation task increases linearly when utilizing holistic methods, while it does so exponentially when using decomposed techniques. Given this, we anticipate that it will take longer to complete the TB comparisons than the PA task when evaluating large sets. This conjecture is consistent with prior arguments about the simplicity of direct methods (Doyle et al., 1997; Jaccard et al., 1986). Thus, we expect that:

H4: In large attribute set evaluations, the TB method will take longer to complete than the PA method.

To assess the above four hypotheses, we carried out an empirical comparison of the two selected multi-attribute evaluation methods. The methodology that we employed to conduct our study is described next.

RESEARCH METHODS

The research instrument consisted of a paper-based evaluation task involving twenty job attributes. We selected an attribute set size of 20, as this was the number that previous researchers have indicated represents a mentally taxing load (Keller and Staelin, 1987; Malhotra, 1982; von Winterfeld and Edwards, 1986, pg. 312). The list of attributes and accompanying descriptions were adapted from Iacovou et al. (2004), and are presented in Appendix B.

Two types of the instrument were generated. The first one contained the PA evaluation task, while the other contained the TB one. The PA instrument asked respondents to allocate a total of 100 points among the 20 attributes. Appendix C presents sample instructions for the PA method. To generate the triads for the TB instrument, we followed the approach described in Shirland et al. (2003). Appendix D provides an example of the instructions provided to respondents, as well as some sample triads. For conciseness, the complete TB instrument has not been reproduced here; it is available from the authors upon request.

Empirical evidence suggests that the order of items under comparison is likely to affect judgments (Wanke, Schwarz, and Noelle-Neuman, 1995). Given this, we posit that reliable methods are more likely to produce similar results than unreliable ones independent of the order of the attributes (see H2). To assess the reliability of the evaluations provided by each method we generated two versions of the instrument (for each evaluation method) by manipulating the presentation order of the items to be assessed. The first TB instrument listed the attributes to be weighted in a specific order (which was randomly determined) and the second one presented them in reversed order. For the PA approach, one set of questionnaires listed the attributes in alphabetical order, and the second set listed them in reversed alphabetical order. Given the above setup, four versions of the questionnaire were prepared, differing in terms of the evaluation method and the presentation order of the attributes (see Figure 1).

Figure 1: Study Design

<u>Scale</u> 1 – TB	1A	1B
2 – PA	2A	2B

Order
A – Forward B - Reverse

To measure the perceived ease of use of each method, the participants were asked to respond to the statement “please indicate the extent you agree or disagree with the following statement: this questionnaire was easy to understand and complete” using a 7-point scale (1 being “strongly disagree” and 7 being “strongly agree”). To measure the time taken to complete the evaluation task, the participants were asked to record (1) the time as they were about to begin working on the task (after reading the instructions) and (2) the time as they completed the task (before rating the perceived ease of the method). The difference between these two time points was recorded as the time needed to complete the task.

Participants were undergraduate business students at a public university. We believe that this is an appropriate sample for our study as job attribute evaluations are perceived to be both important and relevant by students (Bundy and Norris, 1992; Carpenter and Strawser, 1970).

One hundred sixty five students participated in the study by completing one of the four questionnaires, which was randomly assigned to them. Sixty one (37%) were female, and one hundred four were male. The vast majority were traditional students, ranging in age from 19 to 23. The questionnaires were distributed in classes, and were completed over a two-week period. Respondents were given the option of not participating, and no incentives were offered for participation. Of those who were asked to participate, all agreed to do so.

RESULTS

Multiple statistical tests were carried out to assess the hypotheses. In addition to the constructs discussed above, two additional measures were incorporated in our statistical tests: the respondent’s level of interest in job evaluations and gender. These were assessed using self-reported measures. The interest level of the student was measured by their level of agreement with the statement: “job characteristics are important to me at this point in my life.”

Past research shows that these two factors are likely to influence one’s evaluations and were selected as control variables. In terms of interest, individuals who find the task more relevant and appealing are more likely to spend more time and engage in more thoughtful evaluations than those that do not (Bundy and Norris, 1992; Carpenter and Strawser, 1970). With respect to gender, it appears that one’s sex may play a dual role in evaluations. Research shows that men and women tend to engage in dissimilar information processing strategies in cognitive, problem solving tasks (Goddard, Dritschel and Burton, 1998). Moreover, evidence suggests that males and females are likely to have different preferences in terms of job attributes, especially in younger populations such as students (Browne, 1997; Konrad and Langton, 1991; Tolbert and Moen, 1998). To control for such differences, both variables (interest and gender) were incorporated in our tests.

To assess the first two hypotheses, we carried out a MANCOVA multivariate analysis. The dependent variables were the weights for the twenty job attributes. For the PA method, the points allocated by the respondents were used; for the TB method, the weights were calculated by the goal programming model. Four factors were used as predictors of the dependent variables: METHOD (representing whether the respondent used the PA or TB evaluation technique), ORDER (representing whether the attributes were presented in a forward or reverse fashion in the instrument), GENDER (as a control variable) and INTEREST (as a control variable/covariate). The results of the MANCOVA analysis are shown in Table 1. The METHOD variable was statistically significant ($p < .01$), and hence H1 was supported.

In terms of the second hypothesis (on the reliability of the methods), while the results suggest that both methods were impacted by the presentation order of the attributes (ORDER), it appears that they were affected differently by it (see significant ORDER by METHOD interaction effect). This is consistent with H2. To further explore this effect, we computed the average mean rating change (again based on the attribute weights) across ORDER. More specifically, we computed the mean weight for each attribute for respondents in cell 1A from Table 1 (TB, Forward) as well as for cell 1B (TB, Reverse), and computed the difference between them. We then repeated this process to compute the differences between means for the PA method (comparing cell 2A and 2B from Table 1). The sum of the average mean rating change (across ORDER, using absolute values) for TB was 17.2, while the sum of the average mean rating changes for PA was 20.9. This suggests that there was greater variability in responses for the PA method when the presentation order was reversed. Taken together with the MANCOVA findings, these results provide support for H2, suggesting that the TB method provides more consistent results than the PA one.

Table 1: Mancova Results For Test Of Hypotheses 1 & 2

Effect	Pillai’s Trace	F Value	Pr. > F
Interest	.163	1.33	.170
Order	.221	1.94	.014
Method	.419	4.95	.000
Gender	.201	1.72	.036
Order*Method	.228	2.02	.010
Order*Gender	.121	0.94	.537
Method*Gender	.150	1.21	.255
Order*Method*Gender	.147	1.19	.277

Note that Gender did have a statistically significant influence on the attribute ratings. This result is not surprising, as other researchers have noted differences between males and females when evaluating job attributes (Browne, 1997; Konrad and Langton, 1991; Tolbert and Moen, 1998).

H3 (on the perceived ease of completing the task) was tested by using a univariate ANOVA test. The dependent variable was the response to the perceived ease of use (EASE) indicator. To tease out the precise effects of the METHOD variable on EASE, several control variables were included in the test. These included: INTEREST, GENDER, ORDER and TIME (which was the time taken to complete the task). The results are shown in Table 2.

Table 2: Anova Results For Test Of Hypothesis 3

Effect	Mean Square	F Value	Pr. > F
Interest	3.21	1.04	.311
Time	1.00	.323	.571
Order	1.01	.325	.569
Method	9.81	3.16	.077
Gender	0.01	.003	.956
Order*Method	1.51	.486	.487
Order*Gender	0.76	.246	.620
Method*Gender	2.29	.740	.391
Order*Method*Gender	0.51	.163	.687

As the results show, none of these control factors seem to have influenced the participant’s perceptions. The only variable that was statistically significant was METHOD (at $p < .10$). The mean EASE response for the TB method was 4.6 (on a scale of 1 to 7, with 7 being most positive), with a standard deviation of 1.8; for the PA technique, the mean and standard deviation were 3.9 and 1.9, respectively. These results provide modest support to H3.

To assess H4, we performed a t-test comparing the time taken to complete the evaluation task (TIME) across evaluations methods (METHOD). The results suggest that the PA method was more efficient than the TB one ($t=20.69, p < .01$). To complete the TB task the students took on average 14.9 minutes (with a standard deviation of 4.1 minutes), while it took participants 3.5 minutes (with a standard deviation of 1.8 minutes) to carry out the PA evaluation. This finding provides support to H4.

Note that although 3.5 minutes may not appear on the surface to be very long to complete the task, it is consistent with results from previous experiments of this type (Eckenrode, 1965). Indeed, the fact that participants using the TB method spent an average of almost 15 minutes completing the task, and still evaluated it as being easier to complete than those using the PA approach, provides compelling evidence that the TB approach provided a much better opportunity for participants to become actively engaged in the decision making process.

DISCUSSION AND IMPLICATIONS

As anticipated, our results suggest that the PA and TB techniques are likely to yield different evaluations. This finding is consistent with prior work that suggests that different evaluation methods will result in dissimilar outputs, and that different information processing strategies are likely to be pursued by users of various techniques. While this is not a novel finding, it provides additional evidence to a long stream of research on method comparisons.

In terms of reliability, our results suggest that TB-based evaluations tend to be more robust (in terms of order effects). This is a significant discovery as most prior work on method reliability did not consider the issue of attribute set size. In fact, most prior studies suggest that the impact of attribute set size on reliability is universal across all methods. Our findings, however, indicate that method selection and attribute set size seem to have an interaction effect on performance and reliability. When assessing large sets, our work suggests that the high levels of task complexity and information load that are imposed by holistic methods are likely to lead to less reliable results

compared to those of decomposed methods. Given this, it appears that the TB method (and perhaps other decomposed techniques as well) may be a more appropriate choice when dealing with a large set of attributes. This difference on robustness across methods introduces an additional reason for the apparent dissimilarities in evaluations that are produced by different techniques.

With respect to ease of use, our results show that the TB method was more favorably perceived (than the PA one) by the participants. This is an important finding as simplicity and ease of use have been thought of as the key traits of holistic methods. Our findings, however, are inconsistent with prevailing thought. Once again, we believe that a moderating effect may be at play here. Many previous findings are based on evaluations of ten or fewer attributes, while ours are based on twenty attributes. Consequently, one could conclude that although holistic methods may be preferred over decomposed ones in small set assessments, such preferences may be reversed when dealing with longer lists of attributes. This suggests that the decomposing nature of TB and other indirect methods has a suppression effect on task complexity and information load that is indeed recognized and appreciated by decision-makers. Lastly, our work confirms the time efficiency of holistic methods. This is consistent with prior findings, and seems to hold in evaluations of larger attribute sets as well.

Before discussing the implications of our work, we would like to review its limitations. In this regard, our ability to decidedly declare decomposed techniques better than holistic ones in large set evaluations is constrained by several factors. Firstly, we are unsure about whether our findings will extend to other techniques. Although we selected the TB and PA methods because they represent the core of each method type (decomposed and holistic, respectively), additional empirical work must be undertaken (to compare other techniques) before we generalize our results and conclude that one method type is consistently better than the other one in the context of large set assessments.

Secondly, our work was somewhat limited in assessing the quality of the evaluations produced by each method. While our findings suggest that the two methods yield different results (in terms of attribute weights, reliability and perceived ease of use), we don't know whether one technique was truly "better" than the other, in terms of identifying weights that are likely to be more "accurate" in an *objective* sense. This issue is inherent in several multi-attribute method comparison studies because identifying a true, objective solution to a multi-attribute evaluation task is a challenging undertaking. Lastly, we recognize that our work is limited by our selection of a single-item measure for the perceived ease of use construct. This fact may partially explain the moderate support that H3 received (at $p < .10$). Given the face validity of the item, we believe that the findings related to this construct will generalize to future studies utilizing more complex, multi-item scales.

In terms of research contributions, we believe that our work has identified a relatively neglected issue in the research on multi-attribute evaluations: that "attribute set size" is a factor that is likely to play an important role in the performance of evaluation methods. Past work has ignored this in explaining the seemingly inconsistent findings in method comparisons. Indeed, past work focused on the "direct" effects of method selection on evaluation outcomes, without much consideration on the length of the attribute lists. Our findings, however, suggest that an interaction effect between these two variables may be at play, and thus can be partly responsible for the inconsistent findings across method comparison studies.

With respect to practical significance, we believe that our work is of relevance to team leaders and individual decision-makers who are faced with complex multi-attribute assessments. As prior work shows, there is a tendency to select holistic methods for such assessments (because time and other resources are frequently scarce in organizational settings). Our findings, however, suggest that an over-reliance by decision-makers on time considerations in selecting an evaluation method may be inappropriate under certain conditions. Specifically, decision-makers may be ill served by using simple, holistic methods when faced with large set evaluations as their use may cognitively overwhelm the evaluators and thus may result in inaccurate, unreliable decisions. Consequently, the use of sophisticated decomposed techniques may be a more appropriate and effective choice under such circumstances.

In terms of a technique recommendation (among the various decomposed methods), because different techniques are likely to yield different results, we recommend that decision-makers utilize multiple elicitation

techniques to appraise large multi-attribute sets. If multiple methods are used, discussion of the various method outputs can be a good starting point for reaching a consensus-based prioritization of the attributes. As we recognize that this suggestion may be infeasible in many organizations (because of resource limitations), we recommend the use of a single sophisticated method (such as the triad-based one that was used in our study) to conduct large set appraisals. Our results suggest that such methods seem to effectively break down the evaluation into a set of manageable choice-based tasks.

CONCLUSION

Despite many years of research and the development of numerous alternative techniques, we still have little understanding of how best to improve decision making effectiveness when decision makers are faced with the task of prioritizing a large number of attributes. In this study we have shown that decomposed techniques may hold the promise of improving the reliability of large set evaluations, alleviating one of the major concerns expressed for other methods. Moreover, the decomposed approach used in our study was deemed to be easier to use than the holistic method. This should improve participants' satisfaction with the prioritization process. For those involved in decision environments requiring large attribute set evaluations, we believe that decomposed methods offer a promising alternative to simplistic holistic techniques.

SUGGESTIONS FOR FUTURE RESEARCH

This work provides a foundation for future research in the area of multi-attribute decision making. Additional work in this area may include varying the number of attributes to obtain a sense of the exact point at which "information overload" becomes an issue and decomposed techniques lead to more consistent results than holistic evaluation methods. Also, it may involve the comparison of other holistic and decomposed elicitation techniques. Finally, future research may focus on evaluating the effectiveness of multi-attribute decisions for large set evaluations using both objective (whenever feasible) and subjective measures of decision quality (such as perceived satisfaction with the evaluation outcomes).

REFERENCES

1. Ashton, R. H. (1974). Cue Utilization and Expert Judgments: A Comparison of Independent Auditors With Other Judges. *Journal of Applied Psychology*, 59, 437-444.
2. Bottomley, P. A., J. R. Doyle, and R. H. Green. (2000). Testing the Reliability of Weight Elicitation Methods: Direct Ratings Versus Point Allocation. *Journal of Marketing Research*, 37(4), 508-513.
3. Browne, B. A. (1997). Gender and Preferences for Job Attributes: A Cross Cultural Comparison. *Sex Roles*, 37(1/2), 61-71.
4. Bundy, P. and D. Norris. (1992). What Accounting Students Consider Important in the Job Selection Process. *Journal of Applied Business Research*, 8(2), 1-6.
5. Carpenter, C. G. and R. H. Strawser. (1970). Job Selection Preferences of Accounting Students. *Journal of Accountancy*, 84-86.
6. Doyle, J. R., R. H. Green, and P. A. Bottomley. (1997). Judging Relative Importance: Direct Rating and Point Allocation Are not Equivalent. *Organizational Behavior and Human Decision Processes*, 70(1), 65-72.
7. Eckenrode, R. T. (1965). Weighting Multiple Criteria. *Management Science*, 12(3), 180-192.
8. Flynn, B. B., S. Sakakibara, R. G. Schroeder, K. A. Bates, and E. J. Flynn. (1990). Empirical Research Methods in Operations Management. *Journal of Operations Management*, 9(2), 250-284.
9. Galotti, K. M. (1999). *Cognitive Psychology: In and Out of the Laboratory*, Belmont, CA: Brooks/Cole Wadsworth.
10. Goddard, L., B. Dritschel, and A. Burton. (1998). Gender Differences in the Dual-Task Effects on Autobiographical Memory Retrieval During Social Problem Solving. *British Journal of Psychology*, 89(4), 611-628.
11. Harte, J. M. and P. Koele. (1995). A Comparison of Different Methods for the Elicitation of Attribute Weights: Structural Modeling, Process Tracing and Self-Reports. *Organizational Behavior and Human Decision Processes*, 64(1), 49-64.

12. Huber, G. (1974). Multiattribute Utility Models: A Review of Field and Field-Like Studies. *Management Science*, 20, 1393-1402.
13. Iacovou, C., L. Shirland, and R. Thompson. (2004) Job Selection Preferences of Business Students. *Journal of Applied Business Research*, 20(1), 87-98.
14. Jaccard, J., D. Brinberg, and L. J. Ackerman. (1986). Assessing Attribute Importance: A Comparison of Six Methods. *Journal of Consumer Research*, 12(4), 463-468.
15. Jessop, A. (2002). Prioritisation of an IT Budget Within a Local Authority. *Journal of the Operational Research Society*, 53, 33-46.
16. Joyce, E. J. (1976). Expert Judgment in Audit Program Planning. *Journal of Accounting Research* (Supplement), 14, 29-60.
17. Keller, K. and R. Staelin. (1987). Effects of Quality and Quantity of Information on Decision Effectiveness. *Journal of Consumer Research*, 14, 200-213.
18. Kim, J. K. and S. H. Choi. (2001). A Utility Range-based Interactive Group Support System for Multiattribute Decision Making. *Computers & Operations Research*, 28(5), 485-503.
19. Konrad, A. M. and N. Langton. (1991). Sex Differences in Job Preferences, Workplace Segregation, and Compensating Earning Differentials: The Case of Stanford MBA's. *Academy of Management Proceedings*, 368-373.
20. Malhotra, N. (1982). Information Load and Consumer Decision Making. *Journal of Consumer Research*, 8, 419-429.
21. Pöyhönen, M. and R. P. Hämmäläinen. (2001). On the Convergence of Multiattribute Weighting Methods. *European Journal of Operations Research* 129(3), 569-585.
22. Schoemaker, P. J. H. and C. C. Waid. (1982). An Experimental Comparison of Different Approaches to Determining Weights in Additive Utility Models. *Management Science*, 28(2), 182-196.
23. Shen, Q., K. K. Lo, and Q. Wang. (1998). Priority Setting in Maintenance Management: A Modified Multi-attribute Approach Using Analytic Hierarchy Process, *Construction Management and Economics*, 16, 693-702.
24. Shirland, L., R. Jesse, R. Thompson, and C. Iacovou. (2003). Determining Attribute Weights Using Mathematical Programming. *Omega*, Vol. 31(6), December, 423-437.
25. Simon, H. (1969). *The Sciences of the Artificial*, Cambridge, MA: MIT Press.
26. Srivastava, J., T. Connolly, and L. R. Beach. (1995). Do Ranks Suffice? A Comparison of Alternative Weighting Approaches in Value Elicitation. *Organizational Behavior and Human Decision Processes* 63(1), 112-117.
27. Tolbert, P. and P. Moen. (1998). Men's and Women's Definitions of "Good" Jobs: Similarities and Differences by Age and Across Time. *Work and Occupations*, 25(2), May, 168-194.
28. von Winterfeld, D. and W. Edwards. (1986). *Decision Analysis and Behavioral Research*. Cambridge, UK: Cambridge University Press.
29. Vriens, M., H. Opperwal, and M. Wedel. (1998). Ratings-based Versus Choice-based Latent Class Conjoint Models: An Empirical Comparison. *Market Research Society* 40(3), 237-248.
30. Wanke, M., N. Schwarz, and E. Noelle-Neuman. (1995). Asking Comparative Questions: The Impact of the Direction of Comparison. *Public Opinion Quarterly*, 59, 347-372.
31. Wittink, D. R., M. Viens, and W. Burhenne. (1994). Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections. *International Journal of Research in Marketing*, 11, 41-52.

**APPENDIX A
An Overview of the Decomposed, Triad-based Method**

As described in more detail by Shirland et al. (2003), this technique utilizes triad-based judgments instead of pair-wise ones. To generate the importance weights, the triad-based (TB) technique employs a two stage process. First, an integer programming model is used to generate the set of triads that participants will use to rank each attribute in them. The model is formulated as follows:

$$\text{minimize } \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n t_{ijk}, \tag{1}$$

where t_{ijk} is a 0-1 variable denoting the triad containing attributes $i, j,$ and $k,$ respectively; and n denotes the number of attributes under consideration.

The first set of constraints ensures that each (i,j) pair is evaluated at least once. A constraint of the following type is necessary in order to ensure that each pair appears at least once.

$$\sum_{k=j+1}^n t_{ijk} + \sum_{\substack{k=i+1 \\ j-i > 1}}^{j-1} t_{ijk} + \sum_{\substack{k=1 \\ i > 1}}^{i-1} t_{ijk} \geq 1, \tag{2}$$

for all (i, j) pairs with $i = 1, 2, 3, \dots, (n-1)$ and $j = i+1, i+2, i+3, \dots, n.$

The following set of constraints ensures that a particular (i,j) pair does not appear more than twice in the optimal solution.

$$\sum_{k=j+1}^n t_{ijk} + \sum_{\substack{k=i+1 \\ j-i > 1}}^{j-1} t_{ijk} + \sum_{\substack{k=1 \\ i > 1}}^{i-1} t_{ijk} \leq 2, \tag{3}$$

for all (i, j) pairs with $i = 1, 2, 3, \dots, (n-1)$ and $j = i+1, i+2, i+3, \dots, n.$

Next, the within-triad rankings are used as input to a quadratic goal programming model (using squared weighted Euclidean distances) to calculate attribute weights. In each triad, respondents are asked to choose the most important attribute and give it a rank of 1. The next attribute is given an equal rank if it is equally important or one unit higher if it is less important. The third attribute is ranked accordingly. Possible rankings for a triad include (1,1,1), (1,2,3), (1,2,2), (1,1,2). The formulation is as follows:

$$\text{minimize } \sum_{\substack{h=1, \\ j,k \in S_h}}^2 a_{jkh} e^{-2} + \sum_{j,k \in S_0} a_{jk0} \left(e^{-2} + e^{+2} \right), \tag{4}$$

where:

a_{jkh} = the number of responses where attribute j was rated more important than attribute k and $h = 1$ or $2,$

a_{jk0} = the number of responses where attributes j and k were rated equally important.

S_h = the set of all pairs (j,k) of the n attributes where attribute j is rated more important than attribute $k.$

S_0 = the set of all pairs (i,k) of the n attributes where attribute j is rated equally important to attribute $k.$

e_{jkh}^{-2} = deviation variable indicating the amount by which the difference in weights between attribute j and k deviate from the value of h.

$e_{jk0}^{\pm 2}$ = deviation variables for equally weighted attributes.

Subject to:

$$w_k - w_j + e_{jkh}^- \geq h \quad \text{for all } (j,k) \in S_h; \quad h = 1, 2 \quad (5)$$

where:

w_k and w_j indicate the weights of attribute k and j respectively.

e_{jkh}^- is the amount by which $(w_k - w_j)$ is less than the minimum target value of h . For those attributes that are rated equal in importance there will be both a negative and positive deviation variable in Equation 5 and h will be 0.

Two additional constraints are necessary in order to view the attribute weights as rankings.

$$\sum_{j=1}^n w_j = W_{tot}, \quad (6)$$

$$w_j \geq 1 \quad \text{and} \quad w_j \leq n \quad \text{for } j = 1, 2, \dots, n, \quad (7)$$

where W_{tot} denotes the sum of the integers from 1 to n .

APPENDIX B List Of Job Attributes In Instrument

1. **Benefits** – health, dental, retirement, vacation, sick leave, tuition remission, relocation reimbursement, maternity/paternity leave, child care
2. **City size** – rural, suburban, urban
3. **Company recognition** – knowledge of company/brand name
4. **Company size** – number of employees, number of locations, annual revenue, market capitalization
5. **Ease of commute** – round-trip time required and potential stress involved
6. **Environmental reputation** – policies and track record in support of environmental issues
7. **Geographic region** – country, area within country such as U.S. Northeast
8. **Growth potential** – opportunities for promotions, opportunities for personal growth
9. **Job responsibility** – ability to make a difference and/or influence change
10. **Job variety/complexity** – multiple tasks and responsibilities that change
11. **Organization type** – not-for-profit, government agency, manufacturing, financial services, public service, peace corp., etc.
12. **Professional development programs** – required, formal, multi-month/year program
13. **Social responsibility** – community involvement, supportive policies on alternative lifestyles,
14. **Stock options** – opportunity to purchase ownership in the company at a given price
15. **Training classes** – short duration, specific topic training sessions, 1-day spreadsheet software training
16. **Travel opportunities** - options exist for off-site assignments, either domestic or international
17. **Travel requirements** - frequent travel is an expected part of the job
18. **Work culture** – dress code, employee demographics, formal vs. informal communications
19. **Work flexibility** – flexible work hours, opportunity to work from home, etc.
20. **Work style** – team versus individual, pace of work, level of human interaction

**APPENDIX C
PA Instrument Instructions**

To indicate the relative importance of all twenty job attributes, please allocate 100 points among them. Please allocate more points to those attributes that are important to you. For example, if you believe that attribute X is twice as important as attribute Y, please allocate twice as many points to attribute X compared to the points you allocate to attribute Y. Although you may allocate any number of points (including zero) to a specific attribute, the total points allocated should equal 100.

<u>Job Attribute</u>	<u>Point Allocation</u>
• Benefits	_____
• City size	_____
• Company recognition	_____
• Company size	_____
• Ease of commute	_____
• Environmental reputation	_____
• Geographic region	_____
• Growth potential	_____
• Job responsibility	_____
• Job variety/complexity	_____
• Organization type	_____
• Professional development programs	_____
• Social responsibility	_____
• Stock options	_____
• Training classes	_____
• Travel opportunities	_____
• Travel requirements	_____
• Work culture	_____
• Work flexibility	_____
• Work style	_____
Total:	100

**APPENDIX D
TB Instrument Instructions And Sample Triads**

EMPLOYMENT SURVEY

We are interested in your opinions concerning the importance of various job attributes. Your individual answers will be kept completely confidential. Thank you in advance for your participation.

INSTRUCTIONS

Procedure – Rank each three-choice set according to the importance they should be given in choosing a company to work for after graduation. Refer to the Item List Definitions for clarification.

Work across the page. Judge only three choices at a time:

1. Decide which of the three you feel is “most important” among just those three. Place a “1” in the space to the left of that choice. There should be at least one score of “1” for every set of three on the questionnaire.

2. Now, decide which is next most important. Make its score 2 if you judge it less important than the one you scored first. Make it the same score if you judge it to be equal in importance. Place the score in the space to the left of that choice
3. Judge your third choice. Compare it to the one that was your second choice. Make the score of your third choice 1 point higher than your second choice if you judge it less important than your second choice. If it is equally important to your second choice make it the same score as your second choice.

Scoring Examples

1	Growth Potential	2	Benefits	3	Ease of Commute
---	------------------	---	----------	---	-----------------

This response says “Growth Potential” is judged most important among the three choices. It is more important than “Benefits”. “Benefits” is more important than “Ease of Commute”.

2	Growth Potential	2	Benefits	1	Ease of Commute
---	------------------	---	----------	---	-----------------

This response says “Ease of Commute” is judged most important among the three choices. “Benefits” and “Growth Potential” are judged to be equally important.

1	Growth Potential	1	Benefits	1	Ease of Commute
---	------------------	---	----------	---	-----------------

This response says that all three attributes are of equal importance.

SAMPLE TRIADS

- | | | |
|---|---|--|
| <input type="checkbox"/> Stock options | <input type="checkbox"/> Work flexibility | <input type="checkbox"/> Social responsibility |
| <input type="checkbox"/> Growth potential | <input type="checkbox"/> Ease of commute | <input type="checkbox"/> City size |