Elicitation of product attributes in an evaluation context: A comparison of three elicitation techniques

Einar Breivik *, Magne Supphellen 1

Norwegian School of Economics and Business Administration, Institute of Strategy and Management, Breiviksveien 40, 5045 Bergen, Norway

Received 24 May 2001; accepted 5 November 2001

Abstract

Research on method effects in attribute elicitation suggests that different techniques may elicit different attribute sets, which in turn are differentially associated with self-reported attitudes and purchase intentions. However, past research has failed to distinguish between elicitation of perceptual and evaluative attributes. In this paper we focus on elicitation of evaluative attributes. Since evaluation processes typically involve a set of few, but salient, attributes, substantial elicitation method effects are not expected. The present paper reports the results of an experimental comparison of three different techniques (including two variants of one of the techniques) in which the elicitation context was held constant across techniques (product evaluation context). The performance of the different elicitation techniques is compared on common output-dimensions (e.g., attribute importance and predictive ability) and procedural dimensions (task ambiguity and task congruity). The results indicate that method bias is not a serious problem in the elicitation of evaluative attributes. Theoretical and practical implications are discussed.

© 2003 Elsevier Science B.V. All rights reserved.

PsycINFO classification: 2260; 3900
JEL classification: C8
Keywords: Product attributes; Attribute elicitation; Multi-attribute models

*Corresponding author. Tel.: +47-55-95-95-81; fax: +47-55-95-94-30.
E-mail addresses: einar.breivik@nhh.no (E. Breivik), magne.supphellen@nhh.no (M. Supphellen).
1 Tel.: +47-55-95-94-21; fax: +47-55-95-94-30.

0167-4870/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
PII: S0167-4870(02)00156-3
1. Introduction

Attribute information is important for various purposes in research on choice and attitudes and is used in a wide variety of contexts, such as consumer choice, vocational choice, and the development of knowledge systems. It is used as input in, for instance, perceptual mapping, choice or attitude models, and design of product profiles. Several techniques for identifying attributes can be found in the literature (see Appendix A for a list of techniques). To the extent that different sets of attributes influence the validity of model predictions knowledge of whether different elicitation techniques may evoke different sets of attributes are crucial. Despite the importance of the topic and repeated calls for more research on attribute elicitation (e.g., Shocker & Srinivasan, 1979; Steenkamp, van Trijp, & Ten Berge, 1994; Wilkie & Pesssemier, 1973), only a few comparisons of techniques are reported in the literature.

Although previous research is scarce, recent findings in several disciplines suggest that the output of elicitation interviews is influenced by the choice of elicitation technique. For instance, studies within the vocational choice literature indicate that elicited attributes may vary in terms of meaningfulness and personal relevance depending on the choice of elicitation technique (Brown, 1987; Neimeyer, Leso, Marmarosh, Prichard, & Moore, 1992). In a conceptual review of elicitation techniques for eliciting expert knowledge Cooke (1994) concludes that significant differences between techniques should be expected. A few empirical comparisons of elicitation method have been found in the consumer behavior literature (cf. Bech-Larsen & Nielsen, 1999; Kanwar, Olson, & Sims, 1980; Lines, Breivik, & Supphellen, 1995). Differences between elicitation techniques were found for the number of elicited attributes (Bech-Larsen & Nielsen, 1999; Kanwar et al., 1980), abstractness of attributes (Kanwar et al., 1980), variability of attributes (within and across products) (Bech-Larsen & Nielsen, 1999; Lines et al., 1995) and the ability of elicited attribute sets to predict attitudes and purchase intentions (Lines et al., 1995).

However, with the exception of Kanwar et al., these studies lack an explicit control of the elicitation context. The administration of elicitation techniques requires some kind of instruction that may focus either on evaluation or perceptual description of products. Since the studies comparing elicitation techniques have allowed instructions to vary, observed method effects (or lack thereof) may stem from differences in the context imposed by elicitation instructions.

The objective of this paper is to compare method effects of different elicitation techniques in an evaluation context, since evaluative attributes are particularly important for multi-attribute modeling of attitude and choice. More specifically, we focus on consumer’s evaluation of a set of product alternatives. Two product categories, cars and restaurants, are included in the study. In contrast to previous studies, instructions are held constant across techniques with an explicit focus on evaluation. The use of two product categories allows for analyses of product specific method effects, while varying the presentation order allows for an examination of contextual influences.
The article is organized as follows: After a brief review of potential method effects in attribute elicitation, we focus on the role of the elicitation context, product category and presentation order. The subsequent sections present an outline of the experiment and the study results. Finally, findings are reviewed and discussed to explore theoretical and practical implications.

2. Method effects in attribute elicitation

Before we focus on method effects found in attribute elicitation we will address a few properties to look for in an attribute elicitation technique. The main rationale for conducting an elicitation interview is to obtain personally relevant attribute sets. One requirement for the results obtained from an attribute elicitation session would be that the elicited attribute sets predict global evaluations, such as attitudes towards products, ads, and other attitude objects. An elicitation technique would outperform another technique if, all else being equal, it produces attribute sets that possess higher predictive ability. Thus the quality of the elicited attributes with respect to predicting choice and evaluations is important. A desirable property of an elicitation technique in this respect would be its ability to elicit determinant attributes (Myers & Alpert, 1968). A determinant attribute is not only important, but also have to be useful for discriminating among alternatives. In some cases discrimination is not found on the most important attribute. For instance, safety is often ranked first in importance when consumers evaluate cars or airlines, but the same consumers do not see that cars and airlines differ widely with respect to safety. Instead other, less important, attributes will be determinant for the overall attitude or choice. Thus the ability to elicit both important attributes and attributes that obtain values that differ across alternatives are important properties to look for in elicitation techniques. The final dimension we will list here is the amount of information obtained in an elicitation interview. One indicator of the amount of information is the number of elicited attributes. An elicitation technique may be considered better if it elicits more attributes as compared to other techniques. However, the number of attributes itself does not necessarily reflect the pieces of unique information. More attributes do not add to existing information if they reflect a general factor or some form of general response tendency. Thus in addition to the number of attributes it is important to address the level of discrimination across attributes within alternatives. Higher levels of discrimination within alternatives suggest that the elicited attributes provide different types of information and do not just reflect small variations of the same dimension. The importance attached to the different dimensions listed above varies depending on the purpose for the attribute elicitation. Predictive ability may prove very important in various choice modeling tasks, while number of attributes and discrimination within alternatives may be important in the elicitation of knowledge structures such as brand associations.

The typical rationale for expecting method effects in attribute elicitation is based on differences in the structure of elicitation techniques. These differences reflect different conceptualizations of how knowledge is organized in memory. To obtain
attribute models with high predictive validity, measurement procedures should match the principles by which attributes are organized and retrieved from memory. Techniques that do not correspond to the way knowledge is organized in memory would be expected to elicit less relevant and meaningful attribute sets as compared to techniques congruent with the mental organization. The importance and role of the structure imposed by different elicitation techniques is not properly addressed in the literature. Although several causes of observed method effects have been suggested, explanations have seldom been explicitly tested.

In addition to the mental organization structure, predictive ability will also be influenced by the way the information is used. That is, predictive ability is not only determined by the amount of information and the quality offered in the attribute set, but also by how the information is combined to form an overall evaluation. There are several different combinatorial rules or decision strategies for searching and integrating information (cf. Earl, 1990). A central distinction among decision strategies is the extent to which they make tradeoffs among attributes. Strategies that make tradeoffs are referred to as compensatory strategies, while those that do not are referred to as noncompensatory strategies. There are several different compensatory strategies, such as the weighted additive rule, the equal weight heuristic, and the additive difference model, and there are several noncompensatory strategies, such as the elimination by aspect heuristic, the lexicographic heuristic and the satisfying heuristic. The different strategies may be used in isolation or in a combination of two or several different decision rules. The determination of which rule to use depends on the decision problem characteristics, such as task effects and context effects (cf. Payne, Bettman, & Johnson, 1993). When the information load is large, either due to a high number of alternatives or/and attributes, consumers tend to use noncompensatory strategies, or a phased decision strategy. Conversely, less information facilitates more use of compensatory strategies. Furthermore, the task format is found to influence the use of decision strategy (cf. Payne et al., 1993). To the extent particular elicitation techniques favor specific decision rules, the employed combinatorial rule will also affect the comparison of elicitation techniques.

Another source of potential method effects in attribute elicitation is the respondents' understanding of the elicitation procedure. For instance, method effects may stem from the level of perceived ambiguity introduced by different elicitation techniques. Research on self-disclosure has shown that task ambiguity may hamper the reporting process and result in less valid responses (Chelune, 1979). Furthermore, the perceived match between the elicitation task procedure and the employed cognitive process when considering a product may account for some of the method effects reported in previous studies. Procedures that match the subjects' style of attribute processing are likely to obtain better results than procedures with a poor match. The potential impact of procedural factors suggests that testing alternative explanations of method effects requires not only output variables, but also variables that reflect how respondents perceive the procedures of the different techniques.

Product category involvement is another variable claimed to contribute to method effects in attribute elicitation. According to Bech-Larsen and Nielsen (1999, p. 333),
larger method-effects will occur for high involvement products as compared to low involvement products because consumers have more complex cognitive structures for high involvement products. Bech-Larsen and Nielsen used this argument to explain why Lines et al. (1995) found significant method effects in terms of predictive validity in their study of cellular phones (high involvement), whereas no method-effects was found for vegetable oil (low involvement) (Bech-Larsen & Nielsen, 1999).

3. The impact of elicitation context

Several of the method effects observed in previous studies are potentially confounded due to context effects implied by the method instructions. For instance, when using the rank-ordering technique (Reynolds & Gutman, 1988), respondents are asked to rank-order a set of specific product alternatives according to their preferences. Subsequently, attributes are elicited by asking for the reasons why alternatives are ranked in that particular order. Hence, this method implies an explicit focus on evaluation. In contrast, no explicit instructions to focus on product evaluation were given for the investigated sorting techniques (Bech-Larsen & Nielsen, 1999) or free elicitation and repertory grid techniques (Lines et al., 1995). Observed method effects may therefore stem from differences in the instructions and not the elicitation technique itself. In particular, the superior predictive validity of models based on attributes from the rank ordering technique found in Lines et al.’s study could be explained by the explicit focus on evaluation implied by this technique as compared to the other elicitation techniques included.

Previous research on consumer decision making has found evaluation task to influence evaluation processes. Of particular interest is the distinction between conceptual and instrumental tasks (i.e. Wilton & Myers, 1986). A conceptual task requires only that the individual gain some understanding of the problem, while an instrumental task requires the individual to select a particular option from a set of alternatives. Wilton and Myers (1986) found that respondents assigned to a conceptual task used both more time and demonstrating less confidence in evaluations as compared to respondents assigned to an instrumental task. Although it is possible to ensure that instructions are kept constant across elicitation techniques, elicitation instructions tend to vary in terms of how clearly the instrumental task of the procedure is stated. Free elicitation techniques are typically not very specific in terms of instructing the respondent to select a particular option. Conversely, both rank-ordering elicitation and the repertory grid involve contrasting alternatives. However, rank-ordering is usually more geared towards preferences, while repertory grid is more oriented towards similarity judgments. To facilitate comparability between elicitation techniques the instruction should assure that the context is constant across techniques. In particular, empirical comparisons of attribute elicitation techniques should distinguish between attributes elicited for the purpose of *perceptual description* and attributes relevant for *evaluation purposes*.

In the present study we focus on attribute elicitation in a product evaluation context similar to an instrumental task as described by Wilton and Myers (1986).
According to the expectancy-value model of Fishbein and Ajzen (1975), people typically base their attitudinal judgments on a small set of salient (accessible) beliefs. Given high attribute salience, most elicitation techniques should be able to elicit the relevant attributes and thus substantial method effects induced by elicitation techniques are not likely. Furthermore, provided that the alternative set is not large we expect the respondents to employ a compensatory decision strategy based on a relatively small set of salient attributes (cf. Payne et al., 1993).

4. Product category and presentation order

Previous research has suggested that variations in product category involvement could explain performance differences between elicitation techniques (Bech-Larsen & Nielsen, 1999). Additionally, specific characteristics of the product category have the potential to influence attribute elicitation. For instance, due to the more ambiguous nature of service attributes and decision frames, services are claimed to be more susceptible to contextual influences as compared to physical products (Burton, 1990). Indeed, alleged differences between services and traditional products suggest that differences in product characteristics, more specifically product intangibility, result in evaluative consequences (cf. Murray, 1991). To the extent some elicitation techniques are better at eliciting particular types of attributes, one should expect differences in the performance of different elicitation techniques depending on the focal product category.

Elicitation of product attributes is demanding both in terms of interview time and interviewer expertise. Some elicitation procedures are relatively complex and require practice trials to familiarize the respondents with the procedure. To the extent some elicitation techniques are more demanding with respect to practice trials and interviewer time compared to others, they will likely be deemed as less useful. In the current study we examine the requirement for practice trials by examining the improvement in the performance of the elicited attribute sets from the first trial to the second trial.

Consequently, the degree of improvement from repeated trials for different elicitation techniques is an interesting source of method error for different elicitation techniques. For instance, techniques that are relatively complex and thus both time-consuming and demanding with respect to interviewer expertise, may be deemed less useful if they require practice trials.

5. Method

5.1. Design, sampling, and procedure

A field experiment was used to explore differences between elicitation techniques. Consumers in two areas of a European city were recruited by telephone. Since the data collection required personal interviews a cluster sampling procedure was used
to save interviewer travelling time and cost. A total of 160 respondents were randomly assigned to four experimental conditions corresponding to the different elicitation techniques (between subject factor). Subjects reported attributes for both restaurants and cars (within subject factor). Two questionnaires were discarded due to excessive missing data. The two product categories were selected to represent both services and traditional products. The presentation order was counterbalanced (between subject factor). Presentation order was manipulated by randomly assigning subjects to either a restaurant–car or a car–restaurant combination. Thus, the design was a $4^{(\text{elicitation technique})} \times 2^{(\text{product})} \times 2^{(\text{presentation order})}$ mixed factorial design.

Before responding to the elicitation task, all subjects were instructed to focus on evaluation of products in the given category. After the elicitation session, subjects filled in a questionnaire including questions about (a) overall attitudes to five product alternatives in each category (restaurants and cars), (b) attribute evaluations of the same products, and (c) the elicitation procedure. The attributes listed in the questionnaires were those reported by each respondent. Thus, each respondent only evaluated the products on attributes previously reported by the same respondent in the elicitation session.

5.2. Elicitation techniques

Elicitation techniques were selected on the basis of two criteria. First, to ensure a certain degree of overlap with previous studies some of the included techniques were the same as in previous studies. Second, the structure of different elicitation techniques varied as a consequence of differences in the underlying assumptions as to how knowledge is stored in memory. Thus it was deemed important to choose techniques that differ in structure and conceptual basis. The selected elicitation techniques were: direct elicitation, rank ordering elicitation and ideal description.

**Direct elicitation.** Direct elicitation is based on general theories of spreading activation (e.g., Collins & Loftus, 1975) and does not favor any particular view on consumer memory. This technique was included because it is widely used and considered as a “baseline technique for elicitation of attributes” (cf. Fishbein & Ajzen, 1975). The elicitation procedure was simple. Subjects were asked to list the attributes they would use to evaluate restaurants and cars.

**Rank ordering elicitation.** Subjects were instructed to rank order five product alternatives within each product category according to their preferences. Subsequently, respondents were asked to provide reasons for their preferred order, e.g., why car brand #1 was preferred to car brand #2, why #2 was preferred to #3 and so on. Subjects were instructed not to repeat attributes. This technique activates specific knowledge structures of specific brands and is supposed to involve a cognitive task similar to that of a real evaluation (see Desoto, 1961; Reynolds & Gutman, 1988).

Previous research has demonstrated the superior predictive validity of attribute sets based on this technique (Lines et al., 1995). However, the calculation of predictive ability was based on evaluations of the same set of alternatives used to elicit
product attributes. Consequently, high predictive ability is likely due to this somewhat tautological procedure. In the present study, two different sets of product alternatives were used for the elicitation of product attributes and the evaluation of alternatives. Furthermore, two different variants of set construction were used in the study. In the first variant, attributes were derived from a set of product alternatives with lower product quality as compared to the alternatives subsequently evaluated in the questionnaire. In the other variant we employed alternative sets that included product alternatives of similar quality in both the elicitation and the evaluation set. Levels of quality were determined in a pretest. 2

**Ideal description.** Whereas direct elicitation and rank ordering elicitation have been tested in previous research, ideal description has not. Still, ideal description was included here because it is tailored to a specific theory pertinent to an evaluation context. With this technique, respondents were instructed to describe an ideal product (in each product category) that she/he would evaluate most favorably. This approach is consistent with the notion of category ideals (Barsalou, 1985). In this perspective on categorization, the best example of a product category may be the ideal product, rather than an average prototypical product. According to Barsalou, ideals are particularly pertinent to goal-oriented categories, which are categories established for a specific purpose, for instance the evaluation of a product.

The respondents were distributed as follows: 43 individuals responded to the direct elicitation condition (DE), 39 to the description of an ideal condition (DI), 38 to rank ordering elicitation with similar product sets (ROSIM) and 38 to the rank ordering elicitation with dissimilar sets (RODIS).

5.3. Dependent variables

**Output variables.** Five output-variables were included: number of attributes, attribute importance, attribute variability, within-alternative variance, and predictive validity. All variables have been included in previous comparisons of elicitation techniques. Number of attributes was measured by counting the attributes elicited from each individual. Perceived importance was measured on a seven-point scale including “unimportant” and “very important” as end points. Attribute variability and within alternative variance were measured as suggested by Lines et al. (1995) and Bech-Larsen and Nielsen (1999). The attribute variability measure is based on the variation in evaluations of each attribute across products, measured as the mean standard deviations across alternatives for all attributes. In contrast, within-alternative variance is measured as the average variability in evaluations of attributes within each product.

---

2 A random sample (N = 47) rated a large number of cars and restaurants on perceived quality. Based on these ratings three groups of five products were selected from each category: two groups with similar average quality and one group with significantly lower average quality as compared to the other two groups (p < 0.01 in both categories).
The measurement of predictive ability warrants some discussion. Previous studies (cf. Bech-Larsen & Nielsen, 1999; Lines et al., 1995) employed the following procedure for measuring predictive ability. Predictive ability was calculated on an individual level in terms of individual correlations between estimated and self-reported attitudes. Estimated attitudes were calculated by means of a standard linear multi-attribute model. This procedure implies summing the product terms, perceived importance multiplied with the evaluation of the alternative, of the elicited attributes. Self-reported attitudes were measured on three 7-point scales similar to the procedure employed in this study (anchors: “not at all favorable”–“very favorable”, “good–bad”, and “like very much–dislike very much”). Product attributes were also evaluated on 7-point scales, with “very poor” and “very good” as end points. However, this procedure is problematic due to the fundamental indeterminancy of correlating constructs composed of product terms, such as the estimated attitude, with other constructs, such as the self-reported attitude (cf. Bagozzi, 1984). This implies that the obtained correlation is scale dependent. Previous investigations have found that correlation coefficients between constructs composed of product terms and other constructs are sensitive to scale changes. The influence of scale variations varies from small to dramatic, and in some cases even the sign of the correlation changed for varying scale levels (cf. Bagozzi, 1984; Evans, 1991). The proposed procedure to overcome the fundamental indeterminancy is to use a hierarchical regression (cf. Bagozzi, 1984). However, since the individual correlations in this study are based on only five observations, corresponding to the number of alternatives, it is impossible to employ a hierarchical regression for each individual. Furthermore, since each individual employs his/her own elicited attributes, it is impossible to aggregate evaluative responses and perceived importance across individuals. Thus a hierarchical regression within method is not feasible. Consequently, we investigated the sensitivity of the obtained correlations with respect to scale variations. 3 We also assessed the use of two alternative correlation coefficients, the Spearman rank and Pearson product–moment correlation. In our investigation the Spearman rank correlation was found to be relatively insensitive to the imposed scale transformations (the correlations were identical in 152 out of 157 cases), while the Pearson product–moment correlation appeared to be highly sensitive. Hence, we employed the Spearman rank correlation in this study. Next we contrasted this measure of predictive ability with two alternative procedures for measuring predictive ability. The two alternative measures did not involve product terms and thus are not susceptible to the fundamental indeterminancy problem addressed above. The first measure contained an estimated overall evaluation based on the summation of the evaluative judgments for each attribute without using the measures of perceived importance. This procedure corresponds to the equal weight heuristic. The estimated overall evaluation was correlated with the self-reported overall evaluations. The second measure was a

---

3 To assess the sensitivity of the obtained correlations to scale transformations we introduced two different scale transformations. First we added a large positive constant to the evaluative beliefs (+15) and then we subtracted a constant (4) to the evaluative beliefs and the importance weights.
correlation of the evaluation on the first attribute with the self-reported overall evaluation for each individual, which reflects a noncompensatory decision strategy (lexicographic heuristic). A t-test revealed that the latter procedure performed worse (lower predictive ability) compared to the two other approaches. There was no difference between the two first procedures (the multi-attribute model and the equal weight model), suggesting that the inclusion of perceived importance weights did not add to the predictive ability. Hence, we employed the correlation between the self-reported and the summated evaluative judgments as the measure of predictive ability.

Procedural variables. A factor analysis (maximum likelihood, promax rotation) of the procedural variables revealed three factors. The items can be found in Appendix B. Similar patterns were found for both product categories and three factors accounted for 70% of the variance for both product categories. The first factor included three items and was labelled task ambiguity. The second and third factors both reflected the extent to which the structure of the elicitation task was perceived to be in accordance with the way the respondents make their evaluations. The factors were labelled general task congruity and specific task congruity. General task congruity contained four items reflecting the correspondence between the task and the respondents perceived evaluation process. The items were general, referring to the respondent’s way of thinking. Specific task congruity included two items specifically referring to the correspondence between the elicitation task and the respondent’s perceived evaluation of product alternatives. The factor correlation matrix revealed that task ambiguity and general task congruity were positively correlated for both product categories (0.42 for cars and 0.46 for restaurants). Furthermore, general task congruity and specific task congruity were negatively correlated for both the car and restaurant category (−0.37 and −0.49, respectively). The correlation between task ambiguity and specific task congruity was close to zero for the car product category, while it was negative for the restaurant category (−0.3). Task ambiguity was found to be reasonably reliable for both product categories (alpha = 0.75 and 0.86 for cars and restaurants respectively). Both of the other dimensions were found to be highly reliable (alpha > 0.88 for all conditions). In the subsequent analysis we use the mean of the items in each scale for the three different dimensions.

6. Results

The subsequent presentation of the results from the comparison of the included elicitation techniques is divided into two sections. The first section includes the results for the output dimensions while the second section contains the results for the procedural dimensions.

Output dimensions. Table 1 presents the findings for four dimensions: number of attributes, attribute importance, attribute variability and within-alternative variance. The main effect of elicitation technique for number of attributes was significant ($F_{\text{technique}} = 4.361, p < 0.01, \text{df} = 3149$). Pairwise comparisons reveal that this effect was due to lower numbers of attributes in the two conditions with rank ordering
techniques (RODIS and ROSIM), in particular the condition with rank ordering with dissimilar sets (RODIS). However, the difference between DI and ROSIM failed to reach significance. RODIS and ROSIM were not found to be significantly different. Also, a marginal effect of product category on number of attributes was observed ($F_{\text{category}} = 3.256, p < 0.1, df = 1149$).

No statistically significant effects were found for attribute importance. For attribute variability, there was a significant main effect of elicitation technique ($F_{\text{technique}} = 3.097, p < 0.05, df = 3147$). This effect was due to higher levels of attribute variability for the two rank ordering techniques. The results also showed a significant effect of product category on attribute variability ($F_{\text{category}} = 43.167, p < 0.01, df = 1147$) with higher attribute variability for the restaurant category compared to the car category. The interaction between product category and presentation order was also significant ($F_{\text{category} \times \text{order}} = 5.556, p < 0.05, df = 1147$). Specifically, attribute variability was found to be higher for cars when this category was presented first.

No effect of elicitation technique was observed on within-alternative variance. However, both a product category effect and an interaction effect of presentation order and product category were found for within-alternative variance ($F_{\text{category}} = 4.257, p < 0.05$ and $F_{\text{category} \times \text{order}} = 7.955, p < 0.01, df = 1144$). Specifically, higher levels of within-alternative variance were observed for the restaurant category. This difference was notable when restaurants were presented first, while it is nonexistent when cars were presented first.

The results for predictive ability are presented in Table 2. No effect of elicitation technique was found for this variable. However, a significant effect of product category was found for predictive ability ($F_{\text{category}} = 6.219, p < 0.05, df = 1139$), indicating higher predictive ability for restaurants than for cars.

### Table 1
Comparison of elicitation techniques (means and standard deviations)

<table>
<thead>
<tr>
<th>Elicitation Technique</th>
<th>Car</th>
<th>Restaurant</th>
<th>Car</th>
<th>Restaurant</th>
<th>Car</th>
<th>Restaurant</th>
<th>Car</th>
<th>Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of attributes</td>
<td>4.49 (1.40)</td>
<td>4.95 (1.60)</td>
<td>5.82 (0.83)</td>
<td>5.59 (0.88)</td>
<td>1.12 (0.52)</td>
<td>1.40 (0.54)</td>
<td>0.91 (0.42)</td>
<td>0.96 (0.40)</td>
</tr>
<tr>
<td>Mean attribute importance</td>
<td>43 (43)</td>
<td>43 (43)</td>
<td>43 (43)</td>
<td>43 (43)</td>
<td>41 (41)</td>
<td>41 (41)</td>
<td>41 (41)</td>
<td>41 (41)</td>
</tr>
<tr>
<td>Attribute variability</td>
<td>4.33 (2.00)</td>
<td>4.85 (1.83)</td>
<td>5.77 (0.81)</td>
<td>5.56 (0.69)</td>
<td>1.13 (0.53)</td>
<td>1.50 (0.50)</td>
<td>0.78 (0.52)</td>
<td>0.97 (0.46)</td>
</tr>
<tr>
<td>Rank ordering and elicitation (similar sets)</td>
<td>4.32 (1.67)</td>
<td>4.08 (1.44)</td>
<td>5.38 (1.09)</td>
<td>5.54 (0.67)</td>
<td>1.30 (0.49)</td>
<td>1.72 (0.48)</td>
<td>0.94 (0.55)</td>
<td>1.00 (0.50)</td>
</tr>
<tr>
<td>Rank ordering and elicitation (non-similar sets)</td>
<td>3.61 (1.41)</td>
<td>3.92 (1.24)</td>
<td>5.25 (1.03)</td>
<td>5.21 (0.75)</td>
<td>1.40 (0.67)</td>
<td>1.60 (0.47)</td>
<td>0.89 (0.56)</td>
<td>0.97 (0.48)</td>
</tr>
</tbody>
</table>

*Note: Standard deviations in parantheses. Cells sizes in italics.*
Procedural dimensions. Table 3 presents the results for the procedural dimensions. For task ambiguity there was a significant two-way interaction effect of product category and presentation order ($F_{\text{category} \times \text{order}} = 11.836, p < 0.01, \text{df} = 1146$) and a three-way interaction effect between product category, presentation order and elicitation technique ($F_{\text{category} \times \text{order} \times \text{technique}} = 2.684, p < 0.05, \text{df} = 3146$). The two-way interaction reveals that the task ambiguity for both product categories decreased when presented after the other. The three-way interaction effect reflects that the “trial task” was less useful for the rank ordering techniques.

Finally, a significant main effect of elicitation technique was found for general task congruency ($F_{\text{technique}} = 2.988, p < 0.05, \text{df} = 3146$). Again, the rank ordering techniques were found to perform worse than direct elicitation and ideal description, being less congruent to the respondent’s stated way of thinking. The final dimension, specific task congruency, revealed a significant elicitation technique effect ($F = 7.418, p < 0.01, \text{df} = 3146$) and a significant interaction between product category and presentation order ($F = 8.563, p < 0.01, \text{df} = 1146$). Rank ordering techniques were found to perform worse than the other elicitation techniques. In accordance with the effects for task ambiguity, the two-way interaction shows that especially for restaurants, perceived specific task congruity increased the second time an elicitation technique was used.

7. Discussion

Contrary to findings in previous research, this study revealed only marginal elicitation method effects. A summary of the identified differences between elicitation techniques is presented in Table 4. Since most of the identified differences were found
Table 3
Perceptions of procedural variables (means and standard deviations)

<table>
<thead>
<tr>
<th></th>
<th>Task ambiguity</th>
<th>General task congruency</th>
<th>Specific task congruency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car</td>
<td>Restaurant</td>
<td>Car</td>
</tr>
<tr>
<td></td>
<td>Presented 1st</td>
<td>Presented 2nd</td>
<td>Presented 1st</td>
</tr>
<tr>
<td>Free elicitation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.09</td>
<td>(1.48)</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td>Description of an</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ideal</td>
<td>2.62</td>
<td>(1.39)</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Rank ordering and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elicitation (similar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sets)</td>
<td>2.32</td>
<td>(1.41)</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>Rank ordering and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>elicitation (non-similar sets)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.13</td>
<td>(1.15)</td>
<td>2.25</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>16</td>
<td>21</td>
</tr>
</tbody>
</table>

Note: Standard deviations in parentheses. Cell sizes in italics.
with respect to the rank ordering techniques in comparison to the other techniques, we have simplified the presentation by merging the rank ordering techniques with each other and listing direct elicitation and ideal description together.

No significant differences were observed between elicitation techniques on major output dimensions such as attribute importance and predictive validity. Most method effects were observed for the procedural dimensions: task ambiguity and the task congruity dimensions. These variables have not been tested in prior comparisons of attribute elicitation techniques. Contrary to previous suggestions (Desoto, 1961; Lines et al., 1995), the rank ordering techniques were perceived as less congruent with the respondents’ natural way of evaluating products as compared to direct elicitation and ideal description. Also, the task was found to be more ambiguous for the rank ordering techniques compared to direct elicitation and ideal description.

The findings do suggest that the performance of attribute elicitation varies for different product categories. All output dimensions, with the exception of attribute importance, revealed performance differences between restaurants and cars. The elicited attribute sets for restaurants contained more attributes, attributes with higher attribute variability, higher levels of within-alternative variance and higher predictive ability as compared to the car category. However, the results did not differ across elicitation techniques, suggesting that none of the included elicitation techniques performed better for one product category as opposed to the other in eliciting attributes. No significant product category effects were identified for the procedural dimensions.

8. Implications

The lack of method effects for the major output variables in this study lends support to the contention that, irrespective of the choice of technique, the most salient attributes relevant for product evaluation will tend to surface. Notably, this result was obtained with three (four) principally different techniques and for two medium/high involvement products. Perhaps most important, no specific method
was found to elicit attributes with superior ability to predict attitudes or purchase intentions. According to the concept of determinant attributes (Myers & Alpert, 1968), attribute importance is not sufficient for influencing choice. Attributes have to be able to discriminate between relevant alternatives to influence the choice process. Our findings show that the rank-ordering techniques elicited attributes with higher levels of attribute variability (across products), whereas no differences were observed across methods for attribute importance. However, the rank-ordering techniques did not perform better compared to the other techniques with respect to predictive ability.  

The results for the rank ordering technique are consistent with the suggestion of Schibrowsky and Peltier (1995, p. 71), that ranking processes tend to be bottom-up or top-down oriented based on only one key attribute. In line with this conjecture we suggest that the initial ranking of products was based on only one determinant attribute. However, when explanations were asked for at every level of the rank order, respondents had to come up with more attributes. In doing this, they would likely activate salient attributes distinguishing between the two specific products contrasted at each level of the rank order. These processes resulted in a higher level of attribute variability for this method, but not higher levels of attribute importance.

Whereas only minor effects of elicitation technique were observed for the output dimensions, the rank ordering techniques tended to perform worse than direct elicitation and ideal description on the procedural variables, task ambiguity and general and specific task congruity as perceived by the respondents. In particular, the lower congruity between rank ordering tasks and the subjects’ natural way of evaluating products ran counter to previous conceptions of ranking as a natural consumer activity (Desoto, 1961; Schibrowsky & Peltier, 1995). It should be noted that task congruity in this study was measured both with respect to personal styles of evaluation and explicit evaluation of product alternatives.

Based on the results in this study, direct elicitation and ideal description are recommended rather than rank-ordering techniques for eliciting attributes pertinent to product evaluation, due to the time saved and the easier administration associated with direct elicitation and ideal description. Our finding that ideal description and direct elicitation performed better the second time these techniques were administered to the same subjects in terms of a reduction in task ambiguity, indicates that warm-up tasks should be used for these methods. Even though warm-ups are used,

---

4 A regression analysis where the mean perceived importance of the elicited attributes, attribute variability, within alternative variance and number of alternatives were regressed on predictive ability showed that only within alternative variance had a significant negative effect on predictive ability in both product categories. Attribute variability had a positive effect on predictive ability in the restaurant category, but failed to reach significance in the car category.

5 In our discussion of predictive ability we investigated the performance of using only the first attribute for predicting the self-reported overall evaluation. The results suggested that this model performed worse than the one using all the attribute information. Furthermore, a hierarchical regression analysis within each elicitation technique including the four first attributes revealed that in most cases the addition of new variables produced a significant change in F-value.
ideal description and direct elicitation are still less cumbersome and time-consuming than rank ordering techniques.

9. Limitations and directions for further research

Although a comparison of three techniques represents an improvement over most comparable studies in the area (Neimeyer et al., 1992), there are several other techniques that might have yielded other results if included in the study (see Appendix A). However, if instructions are held constant and focus explicitly on evaluation or choice, we believe that only minor method effects will occur because such attributes are very salient in memory and easily evoked by all methods. Substantial method effects may result, however, if instructions focus on perceptual description and not evaluation or choice, since perceptual attributes are more likely to vary in terms salience and accessability.

In this study we used a linear compensatory combination model to calculate overall evaluation and predictive ability, more specifically we employed the equal weight heuristic (Payne et al., 1993). Other kinds of compensatory models may produce different results. For example, the ideal-point model (Lehmann, 1971) would probably favor the ideal description technique. Although compensatory models are relatively robust (Lutz & James, 1977), noncompensatory models might also favor specific techniques (see Fader & McAlister, 1990). The present research context employed few alternatives and thus the information load on each respondent was not overwhelming. The results also revealed that the respondents utilized most of the elicited information. For instance, a hierarchical regression including the four first elicited attributes revealed that the addition of new variables produced a significant increase in the $F$-value in most cases for all the different elicitation techniques. In a situation containing more information (e.g. more alternatives) respondents might employ different decision strategies or combinations of strategies that would utilize less of the available information. Thus a noncompensatory model might prove more useful. To the extent that these models favor a particular elicitation technique, one would expect to find differences between elicitation techniques.

Although we examined method effects on common output dimensions and procedural variables in this study, the analysis is limited with regard to comparative dimensions. Several other interesting dimensions, such as level of attribute abstraction and attribute type (i.e. relative proportions of characteristic-, beneficial- and image-attributes across techniques) have not yet been explored (see Lefkoff-Hagius & Mason, 1993). Investigations of these dimensions may introduce a more theoretical foundation for comparing elicitation techniques, and suggest practical guidelines for selection of elicitation techniques for different research objectives.

Different effects were observed for the two product categories included in this study. This finding implies that elicitation method effects can be category-dependent. In the study we included two categories representing a service and a traditional product. Although we did not find any effects of elicitation technique, other product categories and elicitation techniques may produce different results. Also, other category
effects such as product category involvement have been proposed to have an influence on attribute elicitation. For instance, since the elicitation of attributes for products with socially sensitive attributes (such as status, making a good impression on other people, etc.) is particularly difficult because of social desirability bias (Fisher, 1993), different techniques may produce different results.

The high performance of the new technique, the ideal description technique, suggests that the theory of category ideals (Barsalou, 1985) is highly relevant to attribute elicitation. Future research may delve deeper into this theory in order to improve the ideal description technique and increase our understanding of how product attributes should be elicited from consumer memory.

Acknowledgements

The authors thank the Research Council of Norway for financial support and reviewers and editor for valuable suggestions and comments.
### Appendix A

#### Attribute elicitation techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Procedure</th>
<th>Theoretical basis</th>
<th>Typical kind of attribute searched for</th>
<th>Selected references</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Q-sort techniques</td>
<td>A group of techniques with the common feature that respondents are asked to divide a set of brands into piles and then describe why brands are placed together. Different sorting principles give rise to different variants.</td>
<td>Measurement of values (Thurstone, 1927)</td>
<td>Perceptual attributes</td>
<td>Bech-Larsen and Nielsen (1999), Steenkamp et al. (1994)</td>
</tr>
<tr>
<td>2. Repertory grid</td>
<td>In the most common variant, subjects are given three brands at a time and asked to indicate which two are most alike and on which attributes they differ from the third.</td>
<td>Personal construct theory (Kelly, 1955)</td>
<td>Perceptual attributes</td>
<td>Fransella and Bannister (1977), Sampson (1972)</td>
</tr>
<tr>
<td>3. Direct elicitation</td>
<td>Subjects are asked directly to list specific types of product attributes (e.g., those relevant for evaluation)</td>
<td>Semantic network theory (e.g., Collins &amp; Loftus, 1975)</td>
<td>Evaluative attributes</td>
<td>Olson and Muderrisoglu (1979), Steenkamp et al. (1994)</td>
</tr>
<tr>
<td>4. Free one-word association</td>
<td>Subjects are instructed to report all the words that come to mind when exposed to different brand names in a product category. To avoid chain-formats typical of free association, the focal brand is repeated for each association.</td>
<td>Theories of perceptual meaning (e.g., Szalay &amp; Deese, 1978)</td>
<td>Perceptual attributes</td>
<td>Friedmann (1986), Green, Wind, and Jain (1972)</td>
</tr>
<tr>
<td>5. Ideal descriptions</td>
<td>Subjects are asked to describe the characteristics of an ideal brand within a given product category</td>
<td>Theories of category ideals (Barsalou, 1985)</td>
<td>Evaluative attributes</td>
<td>No studies found</td>
</tr>
<tr>
<td>Step</td>
<td>Methodology</td>
<td>Description</td>
<td>Theories/Techniques</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>6. Rank ordering elicitiation</td>
<td>Subjects first rank order a set of brands and then describe the criteria for the reported order (why #1 was preferred to #2, etc.)</td>
<td>Decision theory (Desoto, 1961)</td>
<td>Bech-Larsen and Nielsen (1999), Reynolds and Gutman (1988)</td>
<td></td>
</tr>
<tr>
<td>7. Selection from a predefined list</td>
<td>First a large number of potentially relevant attributes are listed by experts (e.g. managers). Next, subjects pick personally relevant attributes from the list</td>
<td>No specific</td>
<td>Brown (1987), Lines et al. (1995), Neimeyer et al. (1992)</td>
<td></td>
</tr>
<tr>
<td>8. Visual techniques</td>
<td>Subject are instructed to use some kind of visual stimulus to evoke brand perceptions. One variant asks subjects to look through magazines and select pictures which they in some way associate with the product. Descriptions of why the pictures were included are content-analyzed in search of attribute info</td>
<td>Theories of mental imagery (e.g., Kosslyn, 1994)</td>
<td>Zaltman (1997)</td>
<td></td>
</tr>
<tr>
<td>9. Indirect questioning</td>
<td>Direct elicitation, one-word association or other techniques conducted in the third person (e.g., “what attributes do you believe most people would consider?”)</td>
<td>Theories of projection (Sherwood, 1981)</td>
<td>Socially sensitive attributes Haire (1950), Supphellen (2000)</td>
<td></td>
</tr>
<tr>
<td>10. Group interviews</td>
<td>Groups of subjects discuss which attributes are relevant until consensus is reached on a limited set of attributes</td>
<td>No specific</td>
<td>Both perceptual and evaluative attributes Armacost and Hosseini (1994), Ratneshwar and Shocker (1991)</td>
<td></td>
</tr>
</tbody>
</table>

*The definition of product attribute used here comprises all kinds of perceptions linked to a product in memory (Finn, 1985; Lefkoff-Hagius & Mason, 1993). The ten techniques presented are to be considered categories or classes of techniques rather than specific procedures.

This column indicates the kind of attributes typically searched for when using the various techniques. However, it should be noted that it is usually possible to use the techniques for many different purposes depending on the instructions given.
Appendix B

<table>
<thead>
<tr>
<th>Constructs and items</th>
<th>Note</th>
<th>Alpha level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task ambiguity:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When I received this task I was not sure what to do</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was very unsure about what I was supposed to do</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The task instruction was ambiguous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was totally clear about what to do</td>
<td>Reversed scale item deleted</td>
<td></td>
</tr>
<tr>
<td>when I received this task</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>General task congruity:</strong></td>
<td>All items reversed</td>
<td>0.91 0.89</td>
</tr>
<tr>
<td>This was an unusual way for me to think</td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is not natural for me to think like this</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I did not feel comfortable with this way of thinking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This is an artificial way of thinking</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Specific task congruity:</strong></td>
<td>Item deleted</td>
<td>0.92 0.96</td>
</tr>
<tr>
<td>I think like this when I evaluate cars/restaurants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is natural for me to think like this when I evaluate cars/restaurants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think in a completely different manner when I evaluate whether a car/restaurant is good or bad</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


