

## The effect of attribute questions on overall liking ratings

R. Popper<sup>a,\*</sup>, W. Rosenstock<sup>b</sup>, M. Schraidt<sup>a</sup>, B.J. Kroll<sup>a</sup>

<sup>a</sup> Peryam and Kroll Research Corporation, 6323 North Avondale Avenue, Chicago, IL 60631, USA

<sup>b</sup> Kraft Foods, 801 Waukegan Road, Glenview, IL 60025, USA

Received in revised form 10 December 2003; accepted 10 December 2003

Available online 4 February 2004

### Abstract

In food acceptance tests, it is typical for respondents to rate a product on overall liking and on a series of product attributes. This study was designed to determine if the inclusion of attribute questions can alter a respondent's rating of overall liking. Five groups of respondents evaluated four variations of a dairy dessert. Each group used a different questionnaire, comprised either of (a) overall liking only; (b) overall liking plus intensity scales, which asked respondents to rate the strength of 10 sensory characteristics; (c) overall liking plus attribute liking scales, which asked respondents to rate their liking of appearance, flavor and texture of the products; (d) overall liking plus just-about-right scales, which asked respondents to indicate whether the level of 10 sensory characteristics was "too high", "too low", or "just about right"; and (e) overall liking plus attribute liking and just-about-right scales. The inclusion of just-about-right questions on the questionnaire changed the average overall liking ratings of the products, whereas intensity scales had no such effect. Attribute liking questions also tended to alter overall liking ratings, but to a lesser extent than just-about-right questions. Several hypotheses concerning the source of these context effects are discussed.

© 2003 Elsevier Ltd. All rights reserved.

**Keywords:** Questionnaire design; Context effects; Overall liking; Product attribute ratings

### 1. Introduction

The objective of most quantitative consumer research conducted in support of product development is to determine consumers' affective reaction to new or revised products. In such studies, the level of consumer acceptability is often assessed by asking consumers to rate how much they like a product overall, using, for example, the nine-point hedonic scale (Peryam & Girardot, 1952).

Product developers not only need to know the degree of overall liking, but also what about a product people like and dislike, and how these attributes might be changed to increase acceptability. For this reason, studies often include, in addition to overall liking, questions about the product attributes that are likely to determine the level of overall liking. In tests of food acceptability, these questions often concern the sensory properties of the food, such as its aroma, flavor and texture (Meilgaard, Civille, & Carr, 1991).

The attribute questions employed in food acceptance testing usually take one of three forms: (1) attribute liking questions ask consumers to rate, on a hedonic scale, how much they like the appearance, flavor or texture of a product, or like specific sensory characteristics, such as its color, sweetness, and crunchiness; (2) intensity rating questions ask respondents to rate the strength of a sensory attribute, for example, its sweetness, on a scale from "low" to "high"; and (3) just-about-right questions ask respondents to rate whether the level of a sensory attribute, for example its sweetness, is "too high", "just right" or "too low".

While attribute questions can be useful in identifying the reasons consumers like or dislike a product, there is concern that such questions can be a source of bias (Stone & Sidel, 1993), for example by making certain product attributes especially salient in the respondent's mind (Tourangeau & Rasinski, 1988). The survey design literature contains many examples of contextual effects in questionnaire design, due, for example, to question order (Gross, 1964; Schuman & Presser, 1981; Schwarz & Sudman, 1992; Sudman & Bradburn, 1982). Question order has also been found to be influential in product acceptance and preference tests (Earthy, MacFie, &

\* Corresponding author. Tel.: +1-972-769-0001; fax: +1-972-769-1172.

E-mail address: richard.popper@pk-research.com (R. Popper).

Hedderley, 1997; Ennis, 1993). For this reason, overall liking, or a similar key measure, is usually asked first, in advance of the attribute questions.

However, in acceptance tests involving multiple products, placing the overall liking question first does not eliminate the potential for bias, although it may mitigate it. Attribute questions asked about the first product still have the potential to influence ratings of the products evaluated subsequently, and thus the concern about including attribute questions persists, even when overall liking is the first question asked about each product.

Results of past studies on the effects of different types of attribute questions have been inconsistent. Mela (1989) found no effect of a saltiness intensity question on pleasantness ratings. Similarly, Vickers, Christensen, Fahrenholtz, and Gengler (1993) showed no effect of attribute liking and intensity questions on overall liking. However, Earthy et al. (1997) found that including just-about-right scales changed overall liking ratings compared to when just-about-right questions were not asked.

The present research was undertaken to compare, within the same study, the effects of including attribute liking, intensity, and just-about-right questions on overall liking.

## 2. Method

### 2.1. Materials

Four dairy desserts were tested (Products 1–4). The soft dairy desserts, which are typically served in a cup, were formulated with varying levels of starch and sugar and differed in appearance, flavor, and texture.

### 2.2. Questionnaire designs

Five questionnaire designs were employed, varying in the type and number of attribute questions included (overall liking was always the first question asked):

- Overall liking only.
- Overall liking, followed by 10 intensity scales.
- Overall liking, followed by three attribute liking scales.
- Overall liking, followed by 10 just-about-right scales.
- Overall liking, followed by three attribute liking scales and 10 just-about-right scales. The combination of attribute liking and just-about-right scales on the same ballot was of particular interest because it occurs frequently in product research applications.

Table 1 lists the attributes included on the questionnaires, and Table 2 depicts the scales used for each question type. Attribute liking questions were limited to liking of appearance, flavor and consistency (texture).

Table 1

Liking, intensity, and just-about-right (JAR) attributes used to construct attribute questions

Attribute liking	Intensity or JAR attributes
• Appearance liking	• Thickness
• Flavor liking	• Smoothness
• Consistency liking	• Darkness
	• Overall flavor strength
	• Dairy flavor strength
	• Sweetness
	• Thickness in mouth
	• Smoothness in mouth
	• Creaminess in mouth
	• Mouthcoating

Table 2

Scales used for liking (overall liking and attribute liking), intensity, and just-about-right (JAR) questions

Liking (nine-point)	Intensity (nine-point)	JAR (five-point)
Like extremely	Extremely weak	Much too weak
Like very much		A little too weak
Like moderately		Just about right
Like slightly		A little too strong
Neither like nor dislike		Much too strong
Dislike slightly		
Dislike moderately		
Dislike very much		
Dislike extremely	Extremely strong	

The intensity and just-about-right questions probed 10 specific sensory attributes. The attributes used for these two question types were the same, only the scale differed. While it would have been possible to ask attribute liking questions corresponding to the 10 attributes used for the intensity and just-about-right questions, in practice attribute liking questions are often restricted to a few, general terms.

### 2.3. Procedure

Each respondent evaluated all four dairy desserts using only one of the questionnaire types. The desserts were served one at a time, with the serving order balanced across respondents. There was a timed delay of 5 min between servings. Before the first sample, and in between each of the other samples, respondents had a bite of unsalted cracker and rinsed with filtered water. Desserts were served in plastic cups identified by a three digit code number. The serving size of each dessert was 2 oz. Respondents were tested in groups, and the paper questionnaires were self-administered.

### 2.4. Subjects

A total of 832 respondents were recruited from databases in three US cities, with about 160 respondents randomly assigned to each of the five questionnaire

conditions. Respondents ranged in age between 18 and 59, and 70% of respondents were female. All respondents indicated prior to the study that they liked and consumed the type of dairy dessert tested.

2.5. Statistical analysis

The effect of questionnaire type on overall liking was assessed using a mixed model analysis of variance (SAS Institute Inc., 1999), with questionnaire type and product as the fixed factors and subject as the random factor. Product differences within a questionnaire condition were determined using analysis of variance and post hoc multiple comparisons, based on two-tailed Duncan multiple range tests.

3. Results

The average overall liking scores by product and questionnaire condition are shown in Fig. 1. There was a significant product by questionnaire interaction ( $F(12, 2481) = 2.43, p < 0.01$ ), indicating that overall liking was rated differently depending on the type of questionnaire used. Analyses of variance within a questionnaire condition resulted in significant *F*-tests for product in all questionnaire conditions ( $p < 0.001$  in each case), but the means and the pattern of statistical significance (as determined by Duncan post hoc comparisons) varied by condition.

Fig. 1 shows that in the overall liking only condition, Product 4 was liked more than any other product. The same was true when intensity attributes were asked in addition to overall liking. In contrast, when attribute

liking questions were asked, Product 4 was no longer rated higher in overall liking than Product 3, and when just-about-right scales were asked, whether alone or in combination with attribute liking questions, there was no difference in liking among Products 2–4. The biggest difference across conditions was in how respondents rated Product 4: its average overall liking was 7.6 in the overall liking only condition and dropped to 7.0 in the two conditions that included just-about-right scales.

Fig. 2 further illustrates the change in ratings of Product 4. The figure shows the percentage of respondents in each questionnaire condition rating the product a “9” (“like extremely”). Also shown is the percentage of respondents in each condition that disliked the product to some degree (rating it less than “5” on the nine-point overall liking scale).

The percentage of respondents liking the product extremely was highest in the overall liking only condition and declined when any type of attribute question was added (even intensity questions). The percentage of dislikers did not change in the presence of intensity questions, but increased noticeably in the presence of just-about-right scales.

Table 3 summarizes the just-about-right ratings after collapsing the original five-point scale to three points. Since the just-about-right ratings were similar in the two questionnaire conditions that employed these scales, the data from the two conditions were combined. Table 3 shows that respondents were most critical of Product 1 (which scored the lowest in overall liking), with only four attributes rated “just about right” by 50% or more of the respondents. Respondents were more positive with respect to the characteristics of the other three products, although even Product 4 was criticized for being “not smooth enough” in appearance (33%) and “too weak” in overall flavor (41%).

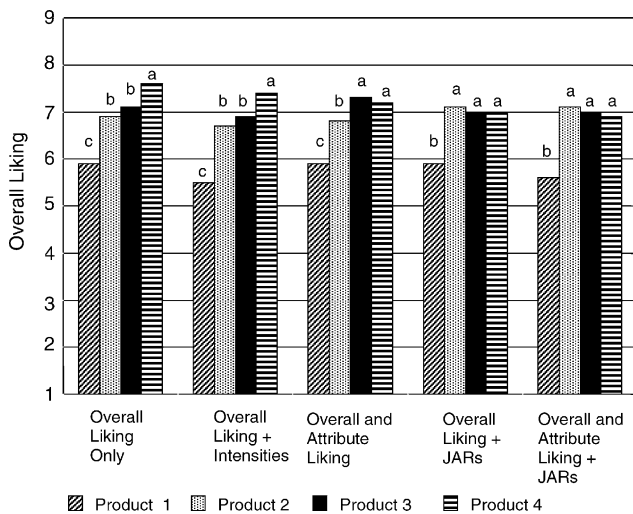


Fig. 1. Mean overall liking ratings for the products by questionnaire condition (JAR = just-about-right). Within a questionnaire condition, the overall liking means that share a common letter do not differ significantly from one another ( $p < 0.05$ ).

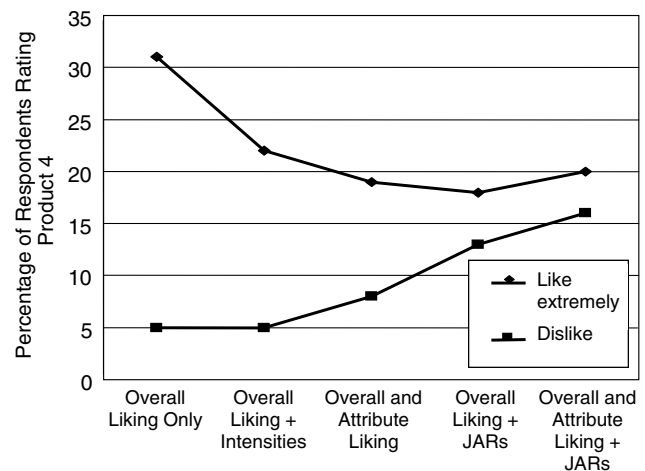


Fig. 2. Percentage of respondents by questionnaire condition that rated Product 4 a “9” (“like extremely”) or rated it less than “5” (i.e. disliked it).

Table 3  
Response percentages for the just-about-right (JAR) questions

Attribute	Product 1	Product 2	Product 3	Product 4
Appearance too thick/JAR/too thin	18/52/31	16/69/16	19/74/8	29/62/9
Appearance too smooth/JAR/not smooth enough	12/60/28	8/70/24	6/71/24	5/63/33
Too dark/JAR/too light	61/36/3	17/70/14	9/66/26	2/30/69
Overall flavor too strong/JAR/too weak	38/28/34	16/52/33	13/51/37	5/55/41
Milk flavor too strong/JAR/too weak	15/36/49	11/59/30	12/65/24	20/67/14
Too sweet/JAR/not sweet enough	32/38/29	20/59/22	18/58/24	14/66/20
Too thick/JAR/too thin	20/48/33	16/61/23	18/71/13	27/67/7
Too smooth in mouth/JAR/not smooth enough	12/63/26	8/74/19	6/77/18	6/73/22
Too creamy/JAR/not creamy enough	13/52/36	9/67/24	10/71/20	16/67/18
Too much mouthcoating/JAR/not enough mouthcoating	29/48/24	17/67/17	17/70/14	22/65/14

The original five-point scales were collapsed to three points and averaged over the two questionnaire conditions in which these scales were used.

Since different respondents participated in the different questionnaire conditions, an additional analysis was conducted to check for the possibility that respondent differences might have contributed to the differences among the questionnaires. Mean overall liking ratings were computed for each product when it was served first, and these first position liking ratings were compared across conditions. Since subjects rated overall liking before answering the attribute questions, the overall liking rating of the first product was unaffected by the later questions and any differences among conditions would be due to factors other than the attribute questions.

Fig. 3 shows the mean first position overall liking ratings for each questionnaire type (number of respondents per questionnaire condition  $\approx 40$ ). A product by questionnaire analysis of variance yielded a significant main effect for product ( $F(3, 812) = 6.09, p < 0.001$ ), although product differences were smaller in first position (in the absence of a product context) than when

based on all serving positions, a familiar finding in product research (Vickers et al., 1993). More importantly, however, the analysis of variance failed to find any significant product by questionnaire interaction ( $F(12, 812) = 0.88, p = 0.56$ ), indicating no systematic respondent differences across the questionnaire conditions.

#### 4. Discussion

The results show that different attribute questions have a differential effect on overall liking ratings. Intensity questions affected the overall liking ratings the least, just-about-right scales the most. When no attribute questions were asked or when intensity questions were asked in addition to overall liking, Product 4 was liked the most. When just-about-right scales were included, Products 2–4 were at parity for overall liking. Attribute liking questions by themselves had a smaller effect than just-about-right questions, although their presence on the questionnaire did change some of the conclusions regarding product differences. Since the number of attribute liking scales was limited in this study, confirmation that attribute liking questions have a lesser effect than just-about-right questions requires research in which an equal number of attribute liking questions and just-about-right questions are asked.

The finding that intensity scales and just-about-right scales differ in their effects on overall liking helps explain past apparent inconsistencies with regard to the effect of attribute questions. Mela (1989) and Vickers et al. (1993) used intensity scales and observed no effect of including attribute questions on overall liking, whereas Earthy et al. (1997) used just-about-right scales and did observe an effect.

Research on survey design has demonstrated that questions early in a survey can effect respondents' answers to later, more general questions, such as their brand preference or their overall position on a social issue (e.g. Bickart, 1993). These effects are typically

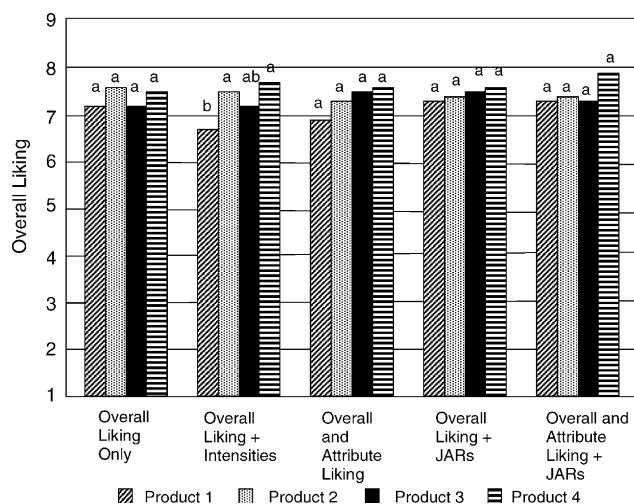


Fig. 3. Mean overall liking ratings in first position by questionnaire condition (JAR = just-about-right). Within a questionnaire condition, the overall liking means that share a common letter do not differ significantly from one another ( $p < 0.05$ ).

attributed to the fact that questions asked early in the survey focus the respondent's attention on certain aspects and change the frame of reference for answering later, more general questions (Schuman & Presser, 1981; Tourangeau & Rasinski, 1988).

The attribute questions included in this study may have made certain product dimensions (for example, the products' appearance) more salient in the consumer's mind than when only overall liking was asked. However, such attentional effects cannot account for the difference in the effect of intensity and just-about-right ratings, since respondents were led to consider the same set of product characteristics in both cases. In fact, intensity questions had only minimal effects on respondents' overall liking ratings, suggesting that the mention of these product attributes did little, by itself, to alter their response.

Intensity scales differ from just-about-right questions in that the latter have an affective component. Just-about-right questions imply that there is an ideal product, one that is "just right". Such questions may make a respondent more critical and aware of the several ways a product falls short of perfection. The just-about-right responses in this study (see Table 3) show that respondents found that there was room for improvement in the case of every product. In comparison to just-about-right scales, intensity scales are affect-neutral. Attribute liking questions also ask for affective judgments, but without the implication of an "ideal" product. This may be the reason, not the difference in the number of questions, that attribute liking questions had a smaller effect on overall liking than just-about-right questions.

If just-about-right questions were to make respondents more critical, one might expect the presence of these questions to reduce the level of overall liking. That was the case in this study for Product 4. However, Earthy et al. (1997) have shown that the presence of just-about-right scales can raise, not just lower overall liking. Hence, just-about-right scales may be altering respondents' evaluation criteria.

One way in which just-about-right questions could affect the evaluation criteria is by leading respondents to think about why they like or dislike a product. Such introspection may lead respondents to attempt to base their liking response on the attributes suggested. Wilson and his colleagues (Wilson & Schooler, 1991; Wilson, LaFleur, & Anderson, 1996) have conducted a series of experimental studies in social psychology that have demonstrated the potentially disruptive effects of "why questions" on preferences. The studies show that first asking subjects to introspect regarding their reasons for liking or disliking a product can change their answer to a subsequent preference question. The studies further demonstrate that the preferences expressed after answering such "why questions" tend to be less pre-

dictive of later preferences and actual behavior than the preferences expressed without forcing introspection.

Wilson's interpretation of his results is that people often lack conscious, well-articulated reasons for their attitudes and preferences. When asked to reflect on such reasons, they focus on the most salient attributes, and when then asked for their preferences, base their answers on the attributes that seem the most plausible reasons for their preference, even though, in reality, these attributes may have little bearing on what they prefer.

In the present study, the just-about-right questions followed the overall liking rating of any one product. Even so, however, the presence of the just-about-right questions may have initiated a process of introspection and rationalization that affected the evaluation process. Intensity questions, on the other hand, apparently did not. Wilson's work suggests that even open-ended questions, such as "what did you like or dislike about this product" can prove to be biasing in product testing, although Vickers et al. (1993) observed no such effect.

One way to avoid the potentially biasing effects of attribute questions is to ask overall liking for each product first, before asking any attribute questions. Having evaluated each product on overall liking, a respondent then would be presented with the products a second time, this time rating them on the attribute scales. While perhaps viable in some situations, there are a number of practical issues in implementing this test protocol. Requiring two separate evaluations of each product would increase the length of the interview and the amount of product required. Both factors would result in increased study costs. In addition, in studies involving a large number of products, doubling the number of samples to be evaluated may be very fatiguing for the respondents.

The practical implications of this research are that research managers need to consider carefully whether they wish to include attribute questions on their questionnaires, in particular just-about-right questions. While just-about-right questions appear to provide actionable data to product developers, including them may pose a risk of invalidating the most critical piece of information, namely how much consumers like the product overall.

## References

- Bickart, B. A. (1993). Question-order effects and brand evaluations: The moderating role of consumer knowledge. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research*. New York: Springer-Verlag.
- Earthy, P. J., MacFie, H. J. H., & Hedderley, D. (1997). Effect of question order on sensory perception and preference in central location trials. *Journal of Sensory Studies*, 12, 215–237.
- Ennis, D. M. (1993). *Is preferential choice affected by attribute scoring in blind consumer product testing?* Presented at the American

- Society for Testing and Materials (ASTM) Meeting, November, 1993, Deerfield Beach, FL.
- Gross, E. J. (1964). The effect of question sequence on measures of buying interest. *Journal of Advertising Research*, 5, 40–41.
- Meilgaard, M., Civille, G. V., & Carr, B. T. (1991). *Sensory evaluation techniques*. Boca Raton, FL: CRC Press.
- Mela, D. J. (1989). A comparison of single and concurrent evaluations of sensory and hedonic attributes. *Journal of Food Science*, 54(4), 1098–1100.
- Peryam, D. R., & Girardot, N. F. (1952). Advanced taste test method. *Food Engineering*, 24(7), 58–61, 194.
- SAS Institute Inc. (1999). *SAS/STAT user's guide version 8*. Cary, NC: SAS Institute Inc.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments in question form, wording, and context*. New York: Academic Press.
- Schwarz, N., & Sudman, S. (1992). *Context effects in social and psychological research*. New York: Springer-Verlag.
- Stone, H., & Sidel, J. (1993). *Sensory evaluation practices*. New York: Academic Press.
- Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299–314.
- Vickers, Z. M., Christensen, C. M., Fahrenholtz, S. K., & Gengler, I. M. (1993). Effect of questionnaire design and the number of samples tasted on hedonic ratings. *Journal of Sensory Studies*, 8, 189–200.
- Wilson, T. D., LaFleur, S. J., & Anderson, D. E. (1996). The validity and consequences of verbal reports about attitudes. In N. Schwarz & S. Sudman (Eds.), *Answering questions* (pp. 91–114). San Francisco: Jossey-Bass Publishers.
- Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60(2), 181–192.