J PROD INNOV MANAG 2006;23:359–370

THE JOURNAL OF
PRODUCT
INNOVATION
MANAGEMENT

An Illustration of Potential Sources of Concept-Test Error*

Richard R. Klink and Gerard A. Athaide

A critical step in prelaunch market analysis needing improvement is concept testing. This article reviews the literature on the three basic design decisions inherent to concept testing: (1) stimuli design; (2) respondent selection; and (3) response measurement. By incorporating findings from diffusion theory, the current review identifies a number of potential sources of concept-test error (e.g., failing to account for adoption orientation could unintentionally mask the response of earlier adopters). Through an exploratory study that replicates in many ways a typical concept test, the present study illustrates how results of conventional concept testing can be sensitive to respondents' adoption orientation and the response measure used. This study offers implications for NPD practice that include accounting for the adoption orientation of respondents, using appropriate response measures such as affective questions for later adopters, and incorporating more product-related information and repeat exposure for later adopters.

Introduction

Although important for corporate success (Schmidt and Calantone, 2002), new product development (NPD) is risky because of high failure rates (Griffin, 1997). These high failure rates are often attributed to a lack of adequate prelaunch market analysis, aptly known as the *fuzzy front end* (Dahl and Moreau, 2002; Zien and Buckler, 1997). Indeed, companies have indicated that the area of NPD needing the most improvement was prelaunch market analysis (Cooper, 1993; Montoya-Weiss and O'Driscoll, 2000; Reid and de Brentani, 2004). A critical stage in prelaunch market analysis needing further improvement is *concept testing* (Dahl and Moreau, 2002; Page, 1993), which refers to estimating "customer reactions to a product idea before committing substantial funds to it" (Moore, 1982, p. 279).

Administered via a questionnaire, a typical concept test contains a statement describing the novel features of the innovation relative to existing alternatives in the marketplace (Crawford and Di Benedetto, 2003). After viewing the concept statement, respondents answer questions pertaining to market assessment, as well as other possible questions that can address the importance of various product features, price, and suggestions for improving the concept (Krishnan and Ulrich, 2001).

The appeal of using concept testing is that it provides a quick, low-cost approach to market assessment prior to committing significant investments (Acito and Hustad, 1981; Page and Rosenbaum, 1992). However, the ability of concept-test results to estimate the potential success of a new product in the marketplace is a function of the soundness of its design. Consequently, the literature on concept testing has focused attention on three basic design decisions inherent to concept testing: (1) stimuli design (i.e., What do we show respondents?); (2) respondent selection (i.e., To which potential customers do we talk?); and (3) response measurement (i.e., How do we measure their responses?).

The present research builds on extant literature by incorporating insights from the diffusion literature to identify how conventional methods of conducting and

Address correspondence to: Richard R. Klink, Sellinger School of Business and Management, Loyola College in Maryland, 4501 North Charles Street, Baltimore, MD 21210-2699. Tel.: (410) 617-5546. Fax: (410) 617-2117. E-mail: rklink@loyola.edu.
* Both authors contributed equally to this article.

interpreting concept tests can lead to error in assessing the perceived marketability of a candidate new product. Further, an exploratory study illustrates the potential impact of these sources of error on concept-test results and offers practical recommendations to reduce this error that include measuring the adoption orientation of the sample, disaggregating the response of earlier from later adopters, and using appropriate response measures for different adopter categories.

The remainder of this article is organized as follows. Initially, a review of existing concept-test literature with a focus on the aforementioned design decisions is reported. Potential sources of concept test error are then identified. The exploratory study empirically demonstrates how some of these sources can lead to error. Finally, the article concludes with implications for NPD practice and theory testing.

Background on Concept Testing

Crawford and Di Benedetto (2003) offered that concept testing can serve three important purposes in NPD: (1) identify poor concepts that can be eliminated; (2) provide an initial estimate of the sales or trial rate of the new product; and (3) help develop the idea. Cooper (1993, p. 155) noted that concept testing attempts to "assess market acceptance and expected sales revenues." Although Lees and Wright (2004) cautioned that concept tests may not yield accurate forecasts of trial rates, there is evidence that the purchase intent question in a typical concept test is often used to obtain an initial estimate of a new product's sales potential. For example, Crawford and Di Benedetto (2003) observed that managers initially combine the number of people who indicate "definitely would buy or "probably would buy" on the purchase intent question into an aggregate score called the top-two-boxes score. Marketers then use past experience or an industry rule of thumb (Bell, 1994) to convert this score into a prediction of actual purchase or sales.

Similarly, the Booz Allen Sales Estimating System (BASES, 2005) group uses concept-test data (1) to evaluate a concept's sales potential and (2) to compare its volume potential relative to other concepts. To generate these sales estimates, BASES integrates responses to the purchase intent question with responses to other measures like value perception, transaction size, and purchase frequency. The resulting sales forecast can be adjusted further by incorporating other factors like the expected type and level of supporting marketing efforts and idiosyncratic regional characteristics. Although sales forecasts drawn from concept testing may not be accurate, the forecasts are adequate for determining whether the candidate product should advance to the next stage of the NPD process (Lees and Wright, 2004).

Mistakes in this go–no-go decision are considered concept-test error, which comprises two error types (Crawford and Di Benedetto, 2003; Schmidt and Calantone, 1997). The first type of error occurs when concept-test results suggest going forward with the candidate product when in fact a no-go decision is appropriate. The second type of error occurs when concept-test results lead to a no-go decision when in fact a go decision is warranted. As noted, the ability of concept-test results to estimate the potential success of a new product in the marketplace is a function of the soundness of its design. Therefore, this article reviews the current state of knowledge regarding the three basic designs fundamental to concept tests: stimuli design, respondent selection, and response measurement.

Stimuli Design

Extant literature on stimuli design advocates that a concept statement should be clear and realistic and should not oversell the concept (Crawford and Di Benedetto, 2003). Although the statement can be worded in a commercial or noncommercial format, the concept's difference from existing alternatives in the marketplace should be unambiguous and credible. To emphasize the concept's benefits, marketers often provide additional props like rough sketches, three-dimensional models, virtual prototypes, or even working prototypes. Because the concept's ideational content may interact with the particular execution of the concept, Haley and Gatty (1971) and Iuso (1975) recommended avoiding such props. However, Cooper

BIOGRAPHICAL SKETCHES

Dr. Richard R. Klink received his Ph.D. from the University of Pittsburgh and is associate professor of marketing in the Joseph A. Sellinger School of Business and Management at Loyola College in Maryland. His research interests include new product development and management with a focus on branding issues. His work has appeared in the *Journal of Marketing Research, Marketing Letters, Journal of Marketing Theory and Practice, Journal of Marketing Education*, and other outlets.

Dr. Gerard A. Athaide received his Ph.D. from Syracuse University and is associate professor of marketing in the Joseph A. Sellinger School of Business and Management at Loyola College in Maryland. His research interests focus on new product development, innovation management, and international marketing.

(1993) and Crawford and Di Benedetto (2003) suggested that going beyond the concept description and getting as close as possible to the final product by including artist drawings and crude working models increases the validity of concept tests. It should be noted that when the concept's prime benefit is experiential (e.g., the aroma of a new perfume, the visual appeal of artwork), the stimulus may not represent the concept well. In such cases, concept testing may not be useful (Crawford and Di Benedetto, 2003).

Respondent Selection

Though respondent selection has received little academic research attention in concept testing, it has been advocated to talk to individuals highly knowledgeable in the relevant product category. Schoormans, Ortt, and de Bont (1995) argued that product expertise allows individuals to understand product information faster, to fill in missing information, to learn more easily, to discriminate between important and unimportant aspects of a product, and to better infer benefits from a product's physical attributes. For major innovations, these authors found that similarity between the evaluations of a product concept and an actual product are greater for product experts than novices. For minor innovations, experts were found (1) to give more consistent evaluations within a given concept test and (2) to generate more stable evaluations over time than novices. They concluded, "We advise the use of consumer expertise to select respondents for concept tests of both minor and major innovations" (p. 161).

However, Moreau, Lehmann, and Markman (2001) found that the relationship between expertise and concept evaluation was not uniform for all types of innovations. Specifically, though the information processing benefits afforded by expertise—in the primary base domain—were found to increase comprehension and perceived net benefits for continuous innovations, expertise reduced both comprehension and perceived net benefits for discontinuous innovations. The authors concluded "that experts are, ceteris paribus, not more prone than novices to adopt discontinuous new products" (p. 28).

Response Measurement

As mentioned, concept testing can provide an initial estimate of a candidate product's sales (Crawford and Di Benedetto, 2003). This estimate is typically derived from the key measure of purchase intent (Dubas, Dubas, and Atwong, 1999) and also from measures of liking (Cooper, 1993; Iuso, 1975). Often, at least with academic studies, liking and purchase intent scales are combined to arrive at an aggregate measure of customer response (see, e.g., Keller and Aaker, 1992; Klink and Smith, 2001; Moreau, Lehmann, and Markman, 2001). Furthermore, concept tests also may include other questions to address various marketing mix execution decisions associated with a new product launch, such as what price to charge and which benefits to promote.

Potential Sources of Concept-Test Error

The present review raises several concerns about conventional concept-test design that could lead to concept-test error. Regarding respondent selection, Cooper (1993) recommended that concept tests use individuals that are representative of the target market. Often, earlier adopters represent the main target market of a new product introduction (Mahajan and Muller, 1998). Even when earlier adopters are not the primary target, understanding their response is also important because they often influence the purchase behavior of later adopters (Rogers, 1995). The concern in the present study is that concept tests currently do not account for adoption orientation and that failing to do so may unintentionally mask the response of earlier adopters. Specifically, given that earlier adopters represent a small subset of the population to adopt—approximately 2.5 percent for innovators and 13.5 percent for early adopters according to Rogers (1995)—random samples from the general population will primarily reflect the response tendencies of later adopters.

Importantly, it should be noted that the common practice of using expertise in respondent selection is a poor surrogate for adoption orientation. Indeed, Klink and Smith (2001) found only a modest correlation between expertise and consumer innovativeness. Further, as noted earlier, Moreau, Lehmann, and Markman (2001) found that in the case of discontinuous innovations, expertise reduced both comprehension and perceived net benefits. They concluded that for really new products, experts may be more accurately characterized as laggards than innovators.

Although it has been argued here that random samples from the general population may primarily reflect the responses of later adopters, their responses may not even be representative of their behavior in the marketplace because of stimuli design concerns. More specifically, in the marketplace, new product launches are often accompanied by significant promotional efforts that can provide extensive information and repeat exposure for many in the target audience. On the other hand, conventional concept tests typically expose respondents to relatively limited information only once and then immediately assess their reaction. This setting may be particularly unnatural for later adopters, as they are not inclined to adopt an innovation on initial exposure. Rather, they often wait for additional information to become available. Also, relative to earlier adopters, later adopters often are more influenced by information from nonmarketing sources, such as word of mouth and observation of the product in use (Bass, 1969; Gatignon and Robertson, 1985; Mahajan and Muller, 1998). However, conventional concept statements only contain information from marketing sources, irrespective of their design, such as commercial versus noncommercial format. In short, the divergence between the marketplace and the concept-test setting regarding the amount, type, and frequency of information available is another potential source of concept-test error.

Concerning response measurement, the measure chosen to assess the response of later adopters also could contribute to concept-test error. Specifically, in light of the previous discussion, the common use of purchase intent in concept testing could further place later adopters in an unrealistic position—that is, to consider adopting a product on initial exposure. Furthermore, if managers follow academic lead by aggregating purchase intent and liking scale items, the resulting measure of response is also a potential source of concept-test error. More specifically, the tripartite view of attitude offers that attitude is comprised of three components—affect, cognition, and conation, or purchase intent—which possess convergent and discriminant validity (Bagozzi et al., 1979). Relative to affect or cognition, conation has been found to be more highly correlated with actual behavior (Ostrom, 1969). To the extent that affective and conative scales are pooled together during interpretation of concept-test results, the ability of the resulting measure to predict actual behaviors related to the potential new product may be compromised.

Lastly, a compounding effect may be observed from pooling these potential sources of concept-test error. To understand why, consider that a salient trait distinguishing earlier from later adopters is risk sensitivity—that is, later adopters are more risk averse than earlier adopters (Rogers, 1995). When a sample comprised primarily of risk-averse individuals, or later adopters, is placed in an unnatural or risk-filled condition—that is, one that asks respondents their purchase intentions at time $T_{\emptyset}$ without relevant information (e.g., information from nonmarketing sources)—one might expect that the potential for concept-test error may be further elevated.

Having identified potential sources of concept-test error, the present study's goal is now to illustrate how these sources can lead to go–no-go decision errors. It should be noted that accounting for these sources will not eliminate all error. Because product adoption is a dynamic process that can occur over extensive periods of time, concept-testing methods in the lab cannot be expected to mirror actual marketplace conditions and hence to estimate market potential without error. Furthermore, an inherent trade-off exists between type I and type II errors—that is, a reduction in one type of error can increase occurrences of the other (Hair et al., 1998). Finally, because concept-test errors are errors in decision making, what is decision error for one firm may not be for another. Simply put, the correctness of a go–no-go decision is contingent on a number of firm-specific factors, such as available resources, required rates or return, and risk tolerance.

Although concerns have been identified for each of the three basic design decisions, the present exploratory study focuses on illustrating respondent selection and response measurement as potential sources of concept-test error for two reasons. First, prior research distinguishes a concept's ideation from its execution (e.g., prototype versus no prototype) and goes so far as to argue that "concept testing is supposed to provide a market evaluation of ideational content itself rather than any particular execution(s) of that content" (Iuso, 1975, p. 228). The threat posed to concept-test results by respondent selection and response measurement concerns relate to ideation and would thus underlie concerns posed by execution or stimuli design.

Second, from a more practical standpoint, managers may be unable or unwilling to reduce the discrepancy between the marketplace and concept-test settings through stimuli design, as it may not be feasible (e.g., incorporating information from personal sources) or costly in terms of time and money (e.g., incorporating repeat exposure), thereby compromising the appeal of concept testing.

That respondent selection and response measurement concerns are sources of concept-test error would be validated by a significant correlation between adoption orientation and response measures on a concept test. To illustrate how these concerns can affect the interpretation of concept-test results, the present study replicates in many ways a conventional concept test. It gauges the market response to several product concepts and thereby produces results similar to what a decision maker, or manager, would use to select one or more alternatives for further development (Dubas, Dubas, and Atwong, 1999; Page and Rosenbaum, 1992).

Study

The subjects of the study were 143 working professionals—with an average age of 29 years—enrolled in an evening graduate program. Before receiving a survey booklet, a greeter told subjects that a research firm was interested in gathering their reactions to some potential new products. The first page of the survey booklet oriented subjects to the task. On the following page appeared a concept statement for a new product idea. See Appendix A for the concept statements. After reading the statement, subjects turned the page and answered questions related to the potential new product and their product adoption behavior. Each subject received one of four possible concept statements to evaluate. The procedure took approximately 10–12 minutes.

Measure Development and Validation

For measure development, the recommendations of Churchill (1979) and Gerbing and Anderson (1988) were followed. For each construct, the literature helped identify the domain of the constructs and generate the following multi-item scales.

Purchase intent. Purchase intent was measured with three items, anchored by 1 = very low to 7 = very high: "The likelihood of purchasing the XXXX is," "The probability that you would buy the XXXX is," and "Your willingness to buy the XXXX is." These items were averaged into one score.

Product liking. Product liking was measured with the following five seven-point adjective scales: 1 = dislike to 7 = like; 1 = undesirable to 7 = desirable; 1 = unfavorable to 7 = favorable; 1 = unpleasing to 7 = pleasing; and 1 = awful to 7 = nice (Ajzen and Fishbein, 1980). These items were averaged into one score.

Consumer innovativeness. Innovativeness has been conceptualized as the desire or willingness to try new and different experiences (Hirschman, 1980). The extent to which this translates into product adoption behavior for a particular person, however, tends to be product-category specific (Gatignon and Robertson, 1985; Hirschman, 1980). In accordance with this view, the current study uses the same innovativeness measure used in prior research (e.g., Klink and Smith, 2001), which borrows closely from the domain specific measure developed and validated by Goldsmith and Hofacker (1991).

To be more specific, subjects in the study were asked to refer to their experiences with the general product category of interest. For instance, for smoke detectors the category was home safety and security products. To assist subjects with their understanding of the product category, sample products of the category were noted. For example, in the case of the home safety and security product category, reference products included such items as smoke detectors, fire extinguishers, carbon monoxide detectors, security alarms. Then, using three seven-point items (1 = strongly disagree to 7 = strongly agree), subjects were asked to indicate the extent to which they agreed or disagreed with a series of statements related to the timing of their purchases of new products in the category. Specifically, the scale score comprised the average of the following items: (1) "Overall, I often buy the latest home safety and security products"; (2) "When I see a new brand of home safety and security product in the store, I often buy it because it is new"; and (3) "I tend to purchase the latest home safety and security products before others do." Scores on these three items were averaged to derive the scale score.

The multi-item scales were subjected to confirmatory factor analysis to establish unidimensionality (Gerbing and Anderson, 1988). The measurement model was estimated by maximum likelihood using covariance analysis of linear structural equations (CALIS, or SAS procedure). For this model, each item was restricted to load only on its own construct, and the constructs were allowed to intercorrelate. The results of the measurement model indicate a good fit to the data (Bagozzi and Yi, 1991): $\chi^2 = 89.59$;

df = 41; $p < .0001$; Bentler and Bonett's (1980) non-normed index (NNFI) = .95; Bentler and Bonett's (1980) NFI = .94; Bentler's (1989) comparative fit index (CFI) = .96; RMSEA = .09.

Tests were conducted to assess convergent and discriminant validity. Convergent validity was assessed by reviewing the t-tests for the standardized factor loadings. Each of the factor loadings was significant at the .001 level, providing support for the convergent validity of the constructs. To establish discriminant validity, a series of new measurement models are estimated in which the correlation between the constructs of interest were constrained to unity. In each case, the chi-square difference test was significant, which suggests that the constructs were distinct. For example, fixing the covariance between purchase intent and consumer innovativeness to 1 resulted in a difference in χ^2 (1) = 246.83, which is significant at the .001 level. Finally, to measure the internal consistency of the items measuring a given construct, the composite reliability index (Fornell and Larcker, 1981) is computed for each construct in the model. The indices for the purchase intent, product liking, and consumer innovativeness scales were .92, .95, and .84, respectively, which indicate that all constructs have good reliabilities.

Stimuli

The product concepts included a new lawnmower, a new smoke detector, a new dry cleaning appliance, and a new laptop computer. Following Iuso's (1975) recommendation that the concepts should be of interest to potential customers, these products were selected through the use of pretests. As recommended by Cooper (1993), each product idea was described in a concept statement of approximately 100 words. Because most new products tend to be moderately innovative (Cooper, 1993) and because concept testing is best suited for new product ideas that are not radically different from existing products (Moore, 1982), each statement was written to communicate a moderately novel feature. For example, much like a home security alarm system, the smoke detector was said to be able to alert local authorities.

Results

To validate this study's contention that accounting for adoption orientation is important when concept testing, two separate regressions were run to gain insight on the relationship between adoption orientation and response measurement. The first found a significant effect of consumer innovativeness on purchase intent ($B = .339$; $p < .001$), whereas the second found only a modestly significant relationship between consumer innovativeness and liking ($B = .108$; $p = .10$). A t-test comparing the unstandardized slope coefficients for consumer innovativeness found that consumer innovativeness had a significantly different effect on purchase intent than liking ($t = 2.625$; $p < .005$). These findings support the contention that adoption orientation matters when gauging new product response, particularly with purchase intent.

To illustrate how concept-test error can result from the aforementioned adoption orientation and response measurement concerns, Table 1 reports the results of each concept on the measures of interest. Because the determination of whether a concept should go forward or not is firm specific, results of each concept are interpreted relative to others. See, for instance, Dubas, Dubas, and Atwong (1999) and Trebbi and Flesch (1993) for a similar approach.

The sample was split into separate groups based on the consumer innovativeness measure. Because innovators and early adopters share, to some extent, their risk-taking propensity and their ability to influence later adopters (Rogers, 1995), studies of diffusion and adoption dynamics often combine these adopter categories and compare them with remaining adopter categories or mass market (Mahajan and Muller, 1998; Moore, 1991). The resulting size of this earlier adopter group is approximately 16 percent of the population to adopt—recall that innovators and early adopters comprise approximately 2.5 percent and 13.5 percent of the population to adopt, respectively—which is often computed at one standard deviation away from the mean on adoption orientation scores. As such, for each of the product categories, this study's subsamples were split at the mean less one standard deviation on the consumer innovativeness measure. The resulting groups are labeled earlier adopters and mass market.

As can be seen in Table 1, the rankings of the product concepts provided by earlier adopters differ from those provided by the mass market. On the key measure of purchase intent, concept 1 received the highest score by earlier adopters, whereas concept 4 received the highest score by the mass market. As one might expect, when looking at the sample as a whole (i.e., earlier adopters and mass market combined), the rankings of the concepts mirror those of the mass

Table 1. Mean Values by Respondent Categories

	Product Concept 1 (Lawnmower)	Product Concept 2 (Smoke Detector)	Product Concept 3 (Dry Cleaning Unit)	Product Concept 4 (Laptop Computer)
Earlier Adopters and Mass Market				
Liking	4.97 (3)[a]	4.59 (4)	5.57 (2)	5.62 (1)
Purchase Intent (PI)	3.50 (3)	3.01 (4)	3.71 (2)	4.23 (1)
Averaged Liking and PI	4.24 (3)	3.80 (4)	4.64 (2)	4.93 (1)
	$n = 35$	$n = 36$	$n = 34$	$n = 38$
Earlier Adopters				
Liking	5.40 (3)	4.80 (4)	5.55 (2)	5.94 (1)
Purchase Intent (PI)	5.40 (1)	3.52 (4)	4.67 (2)	4.48 (3)
Averaged Liking and PI	5.40 (1)	4.16 (4)	5.11 (3)	5.21 (2)
	$n = 5$	$n = 7$	$n = 4$	$n = 7$
Mass Market				
Liking	4.89 (3)	4.54 (4)	5.58 (1)	5.55 (2)
Purchase Intent (PI)	3.15 (3)	2.89 (4)	3.58 (2)	4.17 (1)
Averaged Liking and PI	4.02 (3)	3.72 (4)	4.58 (2)	4.86 (1)
	$n = 30$	$n = 29$	$n = 30$	$n = 31$

[a] Numbers in parentheses denote the relative ranking of concepts for the given measure and respondent group.

market and hence mask those of the critical earlier adopter group. Interestingly, though ranked first on purchase intent by the mass market, concept 4 was only ranked third on this measure for earlier adopters. This divergence in response further underscores the importance of accounting for adoption orientation of the sample. Specifically, even though the mass market might favor concept 4, because earlier adopters responded less favorably to this concept relative to other alternatives, it may not last in the marketplace long enough to reach later adopters. Furthermore, the concept's success in the marketplace could be placed in further jeopardy should earlier adopters pass unfavorable information about the product to later adopters. In short, failing to account for adoption orientation could lead to erroneous go–no-go decisions.

With respect to this study's response measurement concerns, Table 1 reveals that the ranking of concepts based on purchase intent differs from those based on liking and an aggregate score combining liking and purchase intent. Thus, aggregation of measures could be a potential source of concept-test error. Although the mass market's rankings based on purchase intent do not differ from their aggregate rankings, it is noted again that the validity of their purchase intent responses may be compromised by the aforementioned unnatural condition in which they are placed.

It should be noted that a mean split on a single-item, product-category knowledge measure used in Klink and Smith (2001) revealed that experts' ranking of concepts based on purchase intent was the same as the mass market's rankings, thus differing from earlier adopters. This provides further support that experts can be laggards. Given that the concepts tested here are not discontinuous innovations, this finding is interesting because Moreau, Lehmann, and Markman (2001) predicted that experts are more likely to be laggards when the innovation is discontinuous. Furthermore, the correlation between product category knowledge and purchase intent was insignificant ($p > .88$).

Discussion and Managerial Implications

The present study found that adoption orientation matters when concept testing. The study also illustrates how the ranking of concepts can be contingent on the adoption orientation of respondents and the response measure used. A number of implications arise from these findings and illustration.

Managerial Implications

At a minimum, this study suggests that NPD managers need to be cognizant of how the setting in which respondents evaluate new product ideas in a concept test differs from the marketplace condition in which consumers evaluate new products. Moreover, practical steps in concept-test design are suggested to reduce this discrepancy and hence the potential for go–no-go decision error. Importantly, several of these steps require very little additional NPD costs and therefore do not jeopardize the appeal of concept testing.

Sample. Conventional NPD practice and theory propose that concept tests account for the expertise of respondents. However, a closer look at the extant literature reveals several pragmatic difficulties in following such prescriptions. First, although research consistently advocates and supports talking to experts with continuous innovations, results of prior research are opposing in whether experts or novices should be used with discontinuous innovations. Second, the ability to utilize the proper respondent pool is limited by managers' ability to properly discern, a priori, continuous from discontinuous innovations. This is not a trivial problem because it is not unusual for a manager to view an innovation as continuous while the marketplace views it as discontinuous, and vice versa—Post-it notes are an example (Mohr, 2001; Rangan and Bartus, 1995). Third, the literature lacks directives on dealing with innovations that lie more toward the middle of this continuum. Given that most concept tests involve moderately continuous innovations (Crawford and Di Benedetto, 2003; Moore, 1982), this too is not a trivial problem. Fourth, as Moreau, Lehmann, and Markman (2001) noted, innovations can tap into knowledge structures from different domains (e.g., primary versus secondary). A product that draws on innovation from several domains could make classifying products—as innovative or not—and respondents—as experts or not—even more problematic for managers.

Because concept testing can help provide an initial estimate of a new product's sales, it would seem critical for concept tests to select respondents based on traits that have been empirically linked to adoption. In this study's review of the diffusion and adoption literature, no research was found that directly linked expertise with actual adoption. On the other hand, consumer innovativeness has been empirically linked with actual adoption (e.g., Goldsmith and Hofacker, 1991). Importantly, similar to prior work, this study's results did not find a significant relationship between consumer innovativeness and expertise.

The need to consider adoption orientation in concept testing was illustrated by this study. Understanding how earlier adopters will respond to a new product offering seems particularly important in instances when these individuals have relatively greater influence on later adopters. For example, research suggests that they may exert greater influence for products from which pleasure is derived from usage, such as movies, or when association with the product provides a form of self-expression, such as automobiles (Bloch, 1986; Feick and Price, 1987).

This is not to say that estimating the response of later adopters is unimportant. Indeed, Mahajan and Muller (1998) cited instances when targeting the majority—early majority and late majority—is more likely to be preferred, such as when market acceptance of the new product is slow or when profit margins decline slowly over time. The point here is that concept tests need to accurately gauge the response of their target market, whether earlier adopters or not, and that failing to account for differences in adoption orientation may lead to erroneous go–no-go decisions.

Response measurement. This study suggests how the decision to go forward or not with the candidate product may be sensitive to the primary response measure used to inform the decision. Because concept testing is used to assess potential market acceptance, purchase intent will likely remain a key measure to inform the go–no-go decision. Although purchase intent appears to be an appropriate measure for earlier adopters, it may not be for later adopters, as they often wait for more information, including information from more personal sources, before considering adoption. As such, relying solely on purchase intent measures to gauge the marketplace acceptance for later adopters could lead to concept-test error. Given the unidimensionalist view that attitudes are solely affective (Allen, Machleit, and Kleine, 1992; Lutz, 1991) and that attitudes endure over time (Allen, Machleit, and Kleine, 1992), measures of liking taken during a concept test (i.e., at time $T_{\O}$) may more accurately reflect later adopters' actual response in the marketplace (i.e., at a later time). While being aware of the potential for error from pooling measures, managers may use other measures, such as perceptions of affordability or availability, to modify purchase intent to gauge market acceptance of a new product (Jamieson and Bass, 1989). Lastly, managers might consider creating new questions for concept tests that probe the adoption decision for later adopters, such as, "Would you purchase this product if a friend recommended it?"

Stimuli design. Even though this study did not directly examine concept execution, implications about stimuli design arise from its findings and concerns regarding adoption orientation and measurement. As noted, this study illustrates how earlier adopters could respond differently than later adopters to stimuli presented in a concept test. It is also known that, relative to earlier adopters, later adopters often rely on additional information to evaluate new products. Yet conventional concept tests provide uniform stimuli to respondents, regardless of adoption orientation. Because risk and level of information are widely held to be inversely related, the limited information contained in a concept test creates a relatively risky environment to evaluate alternatives. Particularly for later adopters who are a risk-sensitive group, managers who wish to go beyond pure ideational concept testing should provide as much information as would be available in the marketplace when possible. Thus, in these instances, using additional props such as prototypes, advertisements, and brochures in concept tests is recommended, especially when gauging later adopters' response is critical.

Regarding the type of information, recall that later adopters are more inclined to adopt after receiving positive information from a personal source, such as a friend. Although it may be difficult to instill the trust that comes with a friend's advice in marketing stimuli, concept tests might include information that attempts to replicate it, such as, "This product has received rave reviews from others including your friends," to see whether concept test results are sensitive to such information.

Another discrepancy involving stimuli between the marketplace and the concept-test setting involves repeat exposure. Managers may wish to incorporate repeat exposure particularly when concept testing complex product information, as it can reduce risk or uncertainty (Cox and Cox, 2002). However, as mentioned, incorporating repeat exposure may infringe on the very appeal of using concept tests—namely, repeat exposure requires additional time and money.

Research Implications

The concerns presented in this study about possible sources of concept-test error suggest the need for a program of future research across the three concept-test design areas. First, even though the purpose of this study was illustrative of how concept-test error could arise, the data must be interpreted with care. The small size of the early adopter category in the study makes further analysis of differences between adopter groups difficult. Before investigating differences among adopter groups, future research might first examine how to determine these groups. Similar to prior work, the present study applied the rule of thumb of one standard deviation away to bisect earlier adopters from the mass market uniformly across products in the study. Yet prior research on consumer innovativeness suggests the need to look at adoption behavior at the product-category level. Specifically, Gatignon and Robertson (1985) and Hirschman (1980) questioned the existence of a general innovator across product categories and instead favored a view of the innovator that is product-category specific. As such, future research could build on the earlier work of Mahajan, Muller, and Srivastava (1990), which looked at adopter groups for 11 consumer durable products to identify how the size of respective adopter groups vary by product category.

Second, gauging the response of concept-test participants is also an avenue for future inquiry. Future research could examine how measures of affect, purchase intent, and also beliefs differ by participant's adoption orientation. Furthermore, these differences could be investigated over time to better understand how measures from a concept test can reflect the future response of the mass market.

Third, considerable opportunity exists for future research to investigate stimuli design issues in concept testing, particularly when these issues are investigated concurrently with respondent selection and measurement issues. Specifically, future research can investigate how concept test execution—that is, the amount, type, and frequency of information presented—interacts with adoption orientation of participants and the response measure used.

Finally, the design decisions addressed here—sample selection, response measurement, and stimuli design—are fundamental to a wide range of theory testing, including NPD studies. Particular care should be given to these design decisions when testing theories involving new product acceptance. Much like concept testing, NPD studies that involve gauging new product acceptance are often not longitudinal. However, the results of such studies often support or refute theories intended for a dynamic process—product adoption or diffusion. Studies that overlook the differences between the setting in which individuals evaluate new products in the lab versus the marketplace may lead to theories that do not generalize to the marketplace.

References

Acito, Franklin and Hustad, Thomas P. (1981). Industrial Product Concept-Testing. *Industrial Marketing Management* 10(3):157–64.

Ajzen, Icek and Fishbein, Martin (1980). *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Allen, Chris T., Machleit, Karen and Kleine, Susan S. (1992). A Comparison of Attitudes and Emotions as Predictors of Behavior at Diverse Levels of Behavioral Experience. *Journal of Consumer Research* 18:493–504 (March).

Bagozzi, Richard P., Tybout, Alice M., Craig, C. Samuel and Sternthal, Brian (1979). The Construct Validity of the Tripartite Classification of Attitudes. *Journal of Marketing Research* 16:88–95 (February).

Bagozzi, Richard P. and Yi, Y. (1998). On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science* 16(1):74–94.

Bass, Frank M. (1969). A New Product Growth Model for Consumer Durables. *Management Science* 15:215–27 (January).

Bell, Marie (1994). Nestle Refrigerated Foods: Contadina Pasta & Pizza (A). In: *Harvard Business School Cases*. Boston: Harvard Business School Press.

Bentler, P.M. (1989). *EQS Structural Equations Program*. Los Angeles: BMDP Statistical Software.

Bentler, P.M. and Bonett, D.G. (1980). Significance Tests and Goodness-of-Fit in the Analysis of Covariance Structures. *Psychological Bulletin* 88(3):588–606.

Bloch, Peter H. (1986). The Product Enthusiast: Implications for Marketing Strategy. *Journal of Consumer Marketing* 3:51–61 (Summer).

Booz Allen Sales Estimating System (BASES) (2005). PreBASES: Volumetric Early Concept Screening that Works. http://www.bases.com/services/prebases.html.

Churchill Jr., Gilbert A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research* 21:39–53 (February).

Cooper, Robert G. (1993). *Winning at New Products: Accelerating the Process from Idea to Launch*. Reading, MA: Addison-Wesley Publishing Company.

Cox, Dena and Cox, Anthony D. (2002). Beyond First Impressions: The Effects of Repeated Exposure on Consumer Liking of Visually Complex and Simple Product Designs. *Journal of the Academy of Marketing Science* 30:119–30 (Spring).

Crawford, C. Merle and Di Benedetto, C. Anthony (2003). *New Products Management*, 7th ed., New York: McGraw-Hill Irwin.

Dahl, Darren W. and Moreau, Page (2002). The Influence and Value of Analogical Thinking during New Product Ideation. *Journal of Marketing Research* 39:47–60 (February).

Dubas, Khalid M., Dubas, Saeed M. and Atwong, Catherine (1999). Some Difficulties in Predicting New Product Trial Using Concept Test Scores. *Journal of Product and Brand Management* 8(1): 50–60.

Feick, Lawrence F. and Price, Linda L. (1987). The Market Maven: A Diffuser of Marketplace Information. *Journal of Marketing* 51: 83–97 (January).

Fornell, C. and Larcker, D.F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* 18(3):39–50.

Gatignon, Hubert and Robertson, Thomas S. (1985). A Propositional Inventory for New Diffusion Research. *Journal of Consumer Research* 11:849–67 (March).

Gerbing, David W. and Anderson, James C. (1988). An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment. *Journal of Marketing Research* 25:186–92 (May).

Goldsmith, Ronald E. and Hofacker, Charles F. (1991). Measuring Consumer Innovativeness. *Journal of the Academy of Marketing Science* 19(3):209–21.

Griffin, Abbie (1997). PDMA Research on New Product Development Practices: Updating Trends and Benchmarking Best Practices. *Journal of Product Innovation Management* 14:429–58 (November).

Hair Jr., Joseph F., Anderson, Rolph E., Tatham, Ronald L. and Black, William C. (1998). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Haley, Russell I. and Gatty, Ronald (1971). The Trouble with Concept Testing. *Journal of Marketing Research* 8:230–2 (May).

Hirschman, Elizabeth C. (1980). Innovativeness, Novelty Seeking, and Consumer Creativity. *Journal of Consumer Research* 7:283–95 (December).

Iuso, Bill (1975). Concept Testing: An Appropriate Approach. *Journal of Marketing Research* 12(2):228–31.

Jamieson, Linda F. and Bass, Frank M. (1989). Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods. *Journal of Marketing Research* 26:336–45 (August).

Keller, Kevin L. and Aaker, David A. (1992). The Effects of Sequential Introduction of Brand Extensions. *Journal of Marketing Research* 29:35–50 (February).

Klink, Richard R. and Smith, Daniel C. (2001). Threats to the External Validity of Brand Extension Research. *Journal of Marketing Research* 38:326–35 (August).

Krishnan, V. and Ulrich, Karl T. (2001). Product Development Decisions: A Review of the Literature. *Management Science* 47:1–21 (January).

Lees, Gavin and Wright, Malcolm (2004). The Effect of Concept Formulation on Concept Test Scores. *Journal of Product Innovation Management* 21:389–400 (November).

Lutz, Richard J. (1991). The Role of Attitude Theory in Marketing. In: *Perspectives in Consumer Behavior*. Harold H. Kassarjian and Thomas S. Robertson (eds.). Englewood Cliffs, NJ: Prentice-Hall, Inc., 317–39.

Mahajan, Vijay and Muller, Eitan (1998). When Is It Worthwhile Targeting the Majority Instead of the Innovators in a New Product Launch? *Journal of Marketing Research* 35:488–95 (November).

Mahajan, Vijay, Muller, Eitan and Srivastava, Rajendra K. (1990). Determination of Adopter Categories Using Innovation Diffusion Models. *Journal of Marketing Research* 27:37–50 (February).

Mohr, Jakki (2001). *Marketing of High-Technology Products and Innovations*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Montoya-Weiss, Mitzi M. and O'Driscoll, Tony M. (2000). Applying Performance Support Technology in the Fuzzy Front End. *Journal of Product Innovation Management* 17:143–61 (January).

Moore, Geoffrey A. (1991). *Crossing the Chasm*. New York: Harper Business.

Moore, William L. (1982). Concept Testing. *Journal of Business Research* 10(3):279–94.

Moreau, Page C., Lehmann, Donald R. and Markman, Arthur B. (2001). Entrenched Knowledge Structures and Consumer Response to New Products. *Journal of Marketing Research* 38:14–29 (February).

Ostrom, T. (1969). The Relationship between the Affective, Behavioral, and Cognitive Components of Attitude. *Journal of Experimental Social Psychology* 5(1):12–30.

Page, Albert L. (1993). Assessing New Product Development Practices and Performance: Establishing Crucial Norms. *Journal of Product Innovation Management* 10(4):273–90.

Page, Albert L. and Rosenbaum, Harold F. (1992). Developing an Effective Concept Testing Program for Consumer Durables. *Journal of Product Innovation Management* 9:267–77 (December).

Rangan, V. Kasturi and Bartus, Kevin (1995). New Product Commercialization: Common Mistakes. In: *Business Marketing Strategy*. V.K. Rangan, et al. (eds.). Chicago: Irwin.

Reid, Susan E. and de Brentani, Ulrike (2004). The Fuzzy Front End of New Product Development for Discontinuous Innovations: A Theoretical Model. *Journal of Product Innovation Management* 21:170–84 (May).

Rogers, Everett M. (1995). *Diffusion of Innovations*. New York: Free Press.

Schmidt, Jeffrey B. and Calantone, Roger J. (1997). Decision-Making in New Product Development Projects: A Conceptualization and Empirical Investigation. In: *PDMA Conference Proceedings*. L.W. Murray (ed.). Monterey, CA: Product Development and Management Association, 215–44.

Schmidt, Jeffrey B. and Calantone, Roger J. (2002). Escalation of Commitment during New Product Development. *Journal of the Academy of Marketing Science* 30(2):103–18.

Schoormans, Jan P.L., Ortt, Roland J. and de Bont, J.P.M. (1995). Enhancing Concept Test Validity by Using Expert Consumers. *Journal of Product Innovation Management* 12:153–62 (March).

Trebbi Jr., G. and Flesch, Edward J. (1993). Single versus Multiple Concept Tests. *Journal of Advertising Research* 23:21–6 (June–July).

Zien, Karen Anne and Buckler, Sheldon A. (1997). Dreams to Market: Crafting a Culture of Innovation. *Journal of Product Innovation Management* 14:274–87 (July).

Appendix A. Stimuli

Lawnmower

Did you know that, after automobiles, gasoline-powered lawn and garden equipment is the second leading cause of damage to the ozone layer? Introducing a new line of lawnmowers with the environment and convenience in mind—battery-operated lawnmowers. While the battery-operated lawnmower is environmentally friendly, it is also extremely convenient. Unlike conventional electric lawnmowers, you don't have to deal with the hassle of an electric cord out in the yard. And, of course, it does not harm the ozone layer like gasoline lawnmowers. Think about saving the environment and saving the hassle when it comes to lawn care . . . try a battery-operated lawnmower.

Smoke Detector

Introducing an advanced new smoke detector. Whereas conventional home smoke detectors only alert household members when activated, the new smoke detector alerts household members, as well as your local fire department. Patterned after the popular home security alarm system which alerts police when an intruder is detected, the new smoke detector places an immediate call to your local fire department when activated. Similar fire alarm or smoke detector systems have been used for years in apartment buildings and commercial establishments. Now in your own home you can have the added protection of smoke detectors that instantly call your local fire department to action.

Dry Cleaning Unit

Tired of ridiculously high dry cleaning bills? Tired of making those trips to and from the local dry cleaner every week? Introducing a patented new product for your home . . . the home dry cleaning unit. It cleans your clothes the same way as professional services, but without the hefty bills or hassle. The home dry cleaning unit is about the same size as your current washer or dryer. It's easy to use . . . all you have to do is add the special cleaning packet at the beginning of the cycle and your clothes come out with a professional look. Look your best without the expensive dry cleaner bills or inconvenience . . . try the home dry cleaning unit.

Laptop Computer

Tired of limited battery life with your laptop computer? Introducing a new line of advanced laptop computers with solar panels. These solar panels can double the running time of your computer's battery by providing a supplemental source of energy to your computer. Located on the keyboard, these solar panels draw on normal room lighting to extend the computer's battery life when the laptop is open. Even better—the time needed to recharge the battery is less than that of other leading laptops. Available with the latest technology in processing speed, RAM memory, CD ROM, and hard disk drive space.