Adaptive Designs for Likert-Type Data: An Approach for Implementing Marketing Surveys

Within the marketing profession, there is a growing concern for the quality of survey data. As marketing problems become more complex, collecting more information and complex data appears desirable. Though surveys can be (and are being) designed to satisfy this demand for more information, researchers have begun to recognize that such "information-dense" surveys may not only be inefficient (i.e., because of low response rate), but also may compromise the quality of data because of their length and complexity (Childers and Ferrell 1979). New approaches for data collection have been proposed to address this problem. For instance, the computerized interviewing method is reported to be more “efficient” and “interesting” to respondents. Such approaches are useful improvements, but little progress has been made to address the length versus information tradeoff dilemma. For most research involving attitudinal constructs, researchers have no choice but to have respondents respond to all items on a predesigned questionnaire.

Now, imagine this novel method in the context of a scenario. The SMART interviewer (residing in the computer) is given an assignment that involves a lengthy (e.g., more than 200 questions) and complex survey of, say,
consumers’ attitude toward business, advertising, and retail stores. Undaunted, the SMART interviewer calls up a random respondent and asks just a few (typically fewer than five) initial questions. SMART listens carefully to the answers, makes a few mental calculations, and very quickly searches through its item pool for the next item to be asked of the respondent. However, SMART does not pick the next item at random. Instead, it chooses the item that is most “informative” to administer to that particular respondent, given his or her previous responses. In this way SMART continues to select additional items, in the process adapting the survey to that particular (and in general to each and every) respondent. Furthermore, SMART does not ask all 200 questions to assess a respondent’s attitude. Instead, by asking fewer than 30 items, SMART can estimate the attitude level with nearly the same fidelity as a paper-and-pencil method with all 200 questions. Obviously, SMART asks different questions of different respondents, yet it estimates all respondents’ locations along the same attitude continuum. Hence SMART takes less than 15 minutes to complete each interview. In contrast, a paper-and-pencil survey of the same complexity would require at least 45 minutes.

Although SMART-type systems may appear mystical and far-fetched, the underlying theory and principles have been developed for some time and are available for exploitation. These systems often are referred to as “adaptive designs” (or “tailored” testing) and are rooted in the Latent Trait Theory (LTT) (or Item Response Theory; Lord 1980; Weiss 1982). Unfortunately, most advances in the area have been in education and psychology, though a few applications to marketing problems have been reported (e.g., Kamakura and Srivastava 1983). More recently, Balasubramanian and Kamakura (1989) have successfully applied LTT models for dichotomous data (e.g., yes/no responses) to adaptive testing within a marketing context. This work underscores the growing recognition of LTT within marketing and its potential for wide-ranging applications.

However, much prior research on adaptive designs has focused on dichotomous responses; the researchers either collected binary data (e.g., yes/no scale) or dichotomized Likert-type data, thus relying solely on binary LTT models. Though these studies have been useful in the initial stages, advances in the application to multivariate, Likert-type data are important because (1) most attitude scales in marketing have Likert-type response categories (and) for a fixed set of items, the use of Likert-type data is known to yield more information about a respondent’s attitude level than dichotomous data (Samejima 1969). Cohen (1983, p. 253) goes so far as to say that the loss of information due to dichotomization “cannot be justified, given the methods that fully exploit the original measurement information.”

Dodd, Koch, and De Ayala (1989) have recently argued that adaptive designs yielding sizable efficiencies without significant sacrifices in measurement precision can be implemented successfully with “substantially smaller” item pools (e.g., smaller than 30), provided such designs are based on Likert-type instead of binary data. The reason is that Likert-type items yield more “information” (i.e., than binary items) throughout the range of attitude. This possibility of smaller item pools is critical for marketing applications, because current operational measures for most marketing constructs are likely to contain fewer than 30 items. Unfortunately, adaptive designs for Likert-type data with fewer than 30 items have not yet been documented.

The specific aim of our article is to take an initial step into examining adaptive designs for Likert-type data and illustrate their advantages and disadvantages with “very small” item banks (i.e., significantly smaller than 30 items). Because adaptive designs are rooted in Latent Trait Theory, we initially discuss the underlying theory, estimation, and interpretation issues for the graded-response LTT model developed by Samejima (1969). To illustrate the graded-response model, we utilize data obtained for 12 items from the consumer discontinue scale (Lundstrom and Lamont 1976). Then we identify and discuss the key decisions involved in designing an adaptive survey. In addition, we describe an implementation (i.e., simulation) of the adaptive design by utilizing the discontinue items and evaluate the precision and efficiency of the adaptive design. We close with limitations and concluding notes on adaptive designs in particular and the use of LTT models in general.

**THE LATENT TRAIT THEORY APPROACH TO LIKERT-TYPE MEASURES**

For clarity and better appreciation of LTT principles, we begin by discussing some key similarities and differences between the common factor model and the LTT model, and highlight the point that the LTT model is a generalization of the common factor model. Consider the case of multiple Likert-type measures of inherently continuous latent constructs (e.g., attitudes). Typically, the items are assessed by using ordinal-level, discrete categories corresponding to such phrases as (1) strongly disagree, (2) disagree, (3) neither agree nor disagree, (4) agree, and (5) strongly agree (or some other variation of this general theme). Whether 3-, 4-, 5-, or 6-point scales are used is of no importance to the method discussed here. What is important is that each respondent has selected one and only one of the scale categories as best reflecting his or her location on a specific item and that, consequently, the categories have at least a rank order. Most marketers model the relationship between these observable measures and the underlying latent construct by using a common factor model.

---

1We focus on the graded-response model because it explicitly models responses obtained on an ordered, multiple-category scale (e.g., Likert scale). We believe that such models have a much wider application in marketing.
The Common Factor Model

For the case of \( n \) items, \( x_1, x_2, \ldots, x_n \), hypothesized to measure a single latent trait (i.e., common factor or latent construct), \( \theta \), the common factor model posits:

\[
x_i = \lambda_i \theta + e_i
\]

(1)

where \( i \) subscripts the items, \( \lambda_i \) refers to the loading of the \( i \)th item on the latent trait, \( \theta \), and \( e_i \) represents the specific and random error for the \( i \)th item. For convenience, we consider the observable variables, \( x_i \), to be mean centered. Though marketers are very familiar with the model in equation 1, it is useful to highlight certain aspects of this model for further exposition. First, note that equation 1, and hence the factor model, is linear in both the parameters (i.e., \( \lambda \)'s) and the traits (i.e., \( \theta \)). Second, this model also assumes that, for any fixed value of the observable variables, \( x_i \), have at least an interval property. Of particular interest to us is the assumption that for Likert-type data the distances between any two adjacent categories are uniform (i.e., the same as those between any other adjacent categories) and identical (i.e., the same for any two given categories across all items).

Latent Trait Theory: A Generalization of the Common Factor Model

The Latent Trait Theory (LTT) affords a probabilistic approach in modeling the relationship between the observable variables and the underlying construct. Specifically, LTT posits that, for a given trait value, respondents have varying probabilities for checking a category on a Likert scale. Obviously, the higher the standing on \( \theta \), the greater the probability for responding on the highly positively anchored categories (e.g., strongly agree). However, the probability for responding on lower categories is not zero, but finite. The probability of checking the \( k \)th category or higher of the \( i \)th item is referred to as the item characteristic curve (ICC) and denoted by \( P^*_k(\theta) \).

In addition, LTT views the relationship between the observable variables and the latent construct as nonlinear in both the parameters and the underlying trait. However, several different LTT models, each positing a different nonlinear function, are available (Thissen and Steinberg 1986). Following Samejima, we focus on the graded-response LTT model. The graded-response model is appropriate for instances in which the item responses can be evaluated according to the magnitude of agreement with given statements (e.g., attitude). Thus, this model is suited to Likert-type data. Under this model, \( P^*_k(\theta) \), the probability of checking the \( k \)th or higher category of the \( i \)th item, is defined as:

\[
P^*_k(\theta) = \frac{1}{1 + \exp(-D a_i(\theta - b_k))}
\]

where \( a_i \) and \( b_k \) are the discrimination and threshold parameters, respectively, for the \( i \)th item and \( D \) is a scaling constant (value = 1.7). The \( a_i \) parameter specifies the slope of \( P^*_k(\theta) \) at the point of its inflection. Hence this parameter is indicative of the sensitivity (i.e., discriminating power) of the item to variation in trait values in the neighborhood of the inflection point. The interpretation of the \( b_k \) parameter is facilitated by noting that it...
is the value of \( \theta \) for which \( P^*_k(\theta) = 0.5 \). As such, \( b_k \) represents the threshold point (i.e., 50% probability) for checking the \( k \)th or higher categories. Note that unlike the common factor model, equation 5 represents a specific nonlinear function relating \( P^*_k(\theta) \) to item (\( a_i \)), category (\( b_k \)), and trait (\( \theta \)) parameters. Whether or not this function provides an adequate fit of observed responses is clearly an empirical question and can be tested as such (Thissen 1985). However, several studies indicate that for Likert-type data the observed responses show nonlinearity with respect to the latent trait (Hulin, Dragow, and Parsons 1983).

This LTT approach can be viewed as a generalization of the common factor model. First, the LTT model analyzes item responses at the level of the individual response categories for each item. That is, the unit of analysis is a response category (equation 5). In contrast, for the common factor model, an item is the unit of analysis (equation 1). Hence the LTT model operates at a more micro level than the common factor model. Second, LTT models do not assume that Likert-type data have interval properties. Instead, such data are assumed to have just ordinal characteristics. In the common factor model, however, the observable variables are assumed to have the interval property. Note that the ordinal assumption is more general than the premise of interval characteristics. Third, for the LTT model, the regression of the observed responses on \( \theta \) is nonlinear (equation 5). For the common factor model, the regression of \( x \) on \( \theta \) is linear both in \( \theta \) and in the parameters (equation 1). Because the nonlinear function of LTT subsumes the linear common factor model as a special case, the LTT model is more general. Some researchers (e.g., McDonald 1982) therefore refer to LTT models as methods for "nonlinear factor analysis."

Though in some instances a more general model may be desirable per se, the graded LTT model has several additional features (e.g., information functions) that make it desirable for specific applications (e.g., adaptive surveys). To ensure full appreciation of these features, we next discuss the LTT models (e.g., basic equations, estimation issues) and interpret LTT characteristics (e.g., operating curves).

The Graded LTT Model: Basic Equations, Estimation, and Fit Issues

A technical description of LTT models is given by Samejima (1969) and Lord (1980). A somewhat less technical presentation of LTT models and methods is given by Hulin, Dragow, and Parsons (1983). The procedures for parameter estimation used in the following examples are described by Bock and Aitken (1981). Because our objective is to introduce the graded-response LTT model and suggest its implications for adaptive survey designs, the presentation here is necessarily brief. Advanced discussions can be found in the sources cited.

To introduce the graded-response LTT model, we discuss the data obtained on a 6-point Likert format (coded as 1 = strongly disagree and 6 = strongly agree) for 12 items from the consumer discontent scale (Lundstrom and Lamont 1976). These items are listed in the Appendix. Data were available from two independent samples, hereafter referred to as phases 1 and 2.\(^5\) Initially, we discuss phase 1 data, which also were used as the calibration study for phase 2 data.

Basic equations. For the graded LTT model, the probability of checking the \( k \)th or higher category of the \( i \)th item, \( P^*_i(\theta) \), is defined as

\[
P^*_i(\theta) = \frac{1}{1 + \exp[-D_{a_i}(\theta - b_i)]}.
\]

Because five thresholds result from a 6-category Likert scale (e.g., one "threshold" between category 1 and 2, and so on), there are five nontrivial \( b \) parameters for each of the 12 items. Consequently, there are five equations in the form of equation 5 for each item. Figure 1 displays the five \( P^*_i(\theta) \)'s, also referred to as the item characteristic curves (ICC's), for a typical discontent item. Note that these ICC's are nonlinear monotonic functions bounded by 0 and 1.

The probability of an individual responding in the \( k \)th (and only the \( k \)th) category of the \( i \)th item, \( P_{a_i}(\theta) \), is defined as the operating characteristic (OC) (Samejima 1969). In general, the operating characteristics can be derived from the ICC's as follows:

\[
P_{a_i}(\theta) = P^*_i(\theta) - P^*_{a+1}(\theta).
\]

That is, the OC's for a graded-response item are obtained by computing the differences between successive ICC's for each category. Note, however, that for \( k = 1 \), \( P^*_1(\theta) = 1 \) because every respondent will respond in category 1 or higher. Figure 1 also displays the corresponding OC's. Unlike \( P^*_i(\theta) \), the OC's are nonmonotonic functions of \( \theta \). Consider the "agree" category OC in Figure 1. Starting from the least discontent people, as respondents with increasing discontent respond to the item, the probability of checking the "agree" category increases. However, as the discontent level continues to increase, the probability of responding "agree" reaches a maximum and then declines. This decline occurs because respondents with greater discontent tend to respond "strongly agree" instead of "agree."

Next, consider the LTT approach to modeling the joint

---

\(^5\)The data collection consisted of two independent phases. In phase 1, 1,000 randomly selected households received a questionnaire containing the 82 discontent items. In all, 512 responses were received, of which 429 were usable. The respondent profile is generally consistent with the census profile of the geographic area. The responses were obtained on a strongly agree/strongly disagree, 6-point Likert scale. Higher numbers indicate agreement with the item and therefore greater discontent. Six months after the phase 1 study, a phase 2 survey was mailed to a fresh sample that excluded all respondents selected in phase 1. The questionnaire consisted of a subset of discontent items (about 15) and several other measures of dissatisfaction. In all, 235 usable responses were available for analysis.
distribution of responses to all of the 12 discontent items. Let $V = (k_1, k_2, \ldots, k_{12})$ be the observed response pattern for the $r$th respondent ($r = 1, 2, \ldots, N$). Each $k_i$, here, can have any value between 1 and 6 because of the 6-category Likert scale. Then $P_v(\theta)$, the probability of response pattern $V$ given that respondent $r$ has the $\theta$, trait level, is:

$$P_v(\theta) = \text{Prob}(k_1, k_2, \ldots, k_{12} | \theta).$$

To simplify equation 7, LTT models utilize the principle of local independence in a way similar to the common factor model. As noted before, the key implication of this assumption is that all of the common variance among the items can be accounted for by $\theta$. Consequently, though item responses are expected to be correlated, these responses tend to be statistically independent when the trait they purport to measure is held constant. Thus, under local independence,

$$P_v(\theta) = \text{Prob}(k_1 | \theta) \cdot \text{Prob}(k_2 | \theta) \cdot \ldots \cdot \text{Prob}(k_{12} | \theta).$$

That is, the probability of the joint distribution of responses can be written as the product of the conditional probabilities for individual items. Noting that $\text{Prob}(k_i)$ is the OC for the $k_i$th category of the $i$th item, $P_a(\theta)$, we see that equation 8 becomes

$$P_v(\theta) = P_a(\theta) | \theta \cdot P_{a2}(\theta) | \theta \cdot \ldots \cdot P_{a12}(\theta) | \theta,$$

where $P_{a}(\theta)$ is a function of $a_i$, $b_\theta$, and $\theta$ parameters (see equations 5 and 6). The marginal probability for obtaining a response pattern $V$ is obtained by integrating over $\phi(\theta)$, the distribution of $\theta$ in the population of interest.

$$P_v = \int_{-\infty}^{\infty} \prod_{i=1}^{12} P_{a}(\theta) \cdot \phi(\theta) \cdot d\theta$$

Under specified conditions for the dimensionality of the latent space, certain consequences of local independence can be tested. Bejar (1980) provides one such test based on LTT models. In Bejar’s method for unidimensional latent space (i.e., $q = 1$), the items are divided initially into an arbitrary number of subsets such that each subset contains items that are relatively more similar to each other than to other items. Then the $b_{\theta}$ estimates are obtained for each item in two ways: (1) by analyzing each subset individually and (2) by computing estimates from the total set of items. Bejar suggests that the null hypothesis is supported (i.e., $H_0: q = 1$) if the plot of the two $b_{\theta}$ estimates (i.e., subset-based vs. total-based) falls on a straight line with a slope of 1.

Do the 12 discontent items yield evidence of unidimensionality (e.g., $H_0: q = 1$)? To test this hypothesis, we divided the 12 items into four subsets of three items each. The threshold estimates were obtained for the subsets individually, as well as for the total set of 12 items.

The resulting plots of estimated threshold values are in Figure 2. For each of the four subsets, the threshold estimates are mostly along a straight line, though for a few subsets (e.g., items 2, 9, and 10) some deviations occur among the low threshold values (e.g., $b_\theta < -2.5$). This violation is not serious, however, because at low values of $b_\theta$, the standard error of the estimates is significantly higher. Viewed in this context, Figure 2 suggests that unidimensionality of the 12 items is a tenable proposition.

Estimation and goodness-of-fit issues. Several procedures are available for the estimation of LTT parameters, such as the conditional maximum likelihood (CML), joint maximum likelihood (JML), and marginal maximum likelihood (MML) methods (e.g., see Hambleton and Swaminathan 1985). In general, both the item (i.e., $a_i$, $b_\theta$) and trait (i.e., $\theta$) parameters are unknown. However, the estimation problem is viewed as either one of fixed- or random-regressors treatments. In the fixed-regressors model, $\theta$ for $N$ respondents are assumed as $N$ fixed unknown values. For this case, the item and trait parameters are estimated jointly and sometimes are referred to as structural and incidental parameters, respectively (Baker 1987). The JML estimation method is appropriate here.

Under the random-regressors model, $\theta$ is assumed to be a random variable. As such, with appropriate distributional assumptions about $\theta$ (usually normal), the trait estimates are removed from the estimation of item parameters by integrating them out of the likelihood function. This approach is consistent with the common factor model (McDonald 1982). In addition, for "calibration" samples, to which the LTT model is initially fit (as is the case in this study), the trait estimates are rarely of interest in and of themselves. Rather, item parameters are required from the calibration sample to estimate the location of future respondents on the underlying latent construct. For the preceding reasons, we focus on the random regressors model. The MML method is appropriate for parameter estimation under the random regressors model.

Several computer programs are now available to estimate LTT parameters. Thiessen’s (1985) MULTILOG is particularly flexible as well as comprehensive for MML estimators. In addition, MULTILOG allows both unrestricted and restricted LTT models to fit. The term “restricted” is used to refer to any LTT model that imposes constraints on item parameters (i.e., $a_i$ and $b_\theta$). For instance, $a_i$ can be constrained to be equal for all the items considered. This particular constraint results in a special class of LTT models termed “Rasch” models. Furthermore, for each category $k$, $b_\theta$ can be set equal for all the items. Note that this particular restriction corresponds to an assumption (i.e., categories are equivalent across items) commonly utilized for most, if not all, marketing research today. Andrich (1982) has discussed a variation of such LTT models in detail.

MULTILOG was used to fit a graded-response LTT
model to the 12 discontent items for the phase 1 sample. Because these items have not been calibrated previously, the estimation focused on obtaining MML estimators. The results obtained are reported in Table 1. Does the graded-response LTT model provide an adequate fit of the data? MULTILOG provides one index of fit—the likelihood ratio chi square statistic—computed as

\[ G^2 = 2N \sum_{q} P_q \log \left( \frac{P_q}{\hat{P}_q} \right) \]

where \( N \) = number of respondents, \( P_q \) = observed proportion in the \( q^{th} \) cell, and \( \hat{P}_q \) = expected proportion in the \( q^{th} \) cell. As usual, the degrees of freedom associated with the \( G^2 \) statistic are the number of free cells less the number of parameters to be estimated. Because there is one linear dependency in a multiway table, the number of free cells associated with a 12-item scale of six categories is \( 6^{12} - 1 \) = 2,176,782,335. The number of item parameter estimates required for the graded model is 12 \( a_i \), plus \( (12 \times 5) = 60 \) \( b_{ik} \) estimates, resulting in 2,176,782,263 degrees of freedom. Correspondingly, the \( G^2 \) statistic is 9449.

Though the fit may appear to be adequate, such an interpretation is not technically valid. The reason is that the observed responses are spread so sparsely over the \( 6^{12} \) cells that there is no chi square statistic against a "general multinomial alternative"; consequently, the \( G^2 \) statistic in this case does not provide an absolute test for the goodness of fit. This problem is not likely to arise
Table 1
ESTIMATED MML ITEM PARAMETERS FOR THE 12 DISCONTENT ITEMS

<table>
<thead>
<tr>
<th>Item number*</th>
<th>Discrimination parameter</th>
<th>Threshold parameters*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_i$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>1</td>
<td>1.21</td>
<td>-.307</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
<td>(.38)</td>
</tr>
<tr>
<td>2</td>
<td>1.01</td>
<td>-.389</td>
</tr>
<tr>
<td></td>
<td>(.12)</td>
<td>(.61)</td>
</tr>
<tr>
<td>3</td>
<td>2.06</td>
<td>-.266</td>
</tr>
<tr>
<td></td>
<td>(.23)</td>
<td>(.30)</td>
</tr>
<tr>
<td>4</td>
<td>.83</td>
<td>-.470</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
<td>(.82)</td>
</tr>
<tr>
<td>5</td>
<td>1.70</td>
<td>-.225</td>
</tr>
<tr>
<td></td>
<td>(.15)</td>
<td>(.38)</td>
</tr>
<tr>
<td>6</td>
<td>1.23</td>
<td>-.377</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
<td>(.44)</td>
</tr>
<tr>
<td>7</td>
<td>1.68</td>
<td>-.310</td>
</tr>
<tr>
<td></td>
<td>(.15)</td>
<td>(.39)</td>
</tr>
<tr>
<td>8</td>
<td>1.49</td>
<td>-.351</td>
</tr>
<tr>
<td></td>
<td>(.15)</td>
<td>(.44)</td>
</tr>
<tr>
<td>9</td>
<td>1.44</td>
<td>-.387</td>
</tr>
<tr>
<td></td>
<td>(.14)</td>
<td>(.45)</td>
</tr>
<tr>
<td>10</td>
<td>1.58</td>
<td>-.337</td>
</tr>
<tr>
<td></td>
<td>(.16)</td>
<td>(.41)</td>
</tr>
<tr>
<td>11</td>
<td>1.20</td>
<td>-.371</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
<td>(.51)</td>
</tr>
<tr>
<td>12</td>
<td>1.03</td>
<td>-.438</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
<td>(.30)</td>
</tr>
</tbody>
</table>

*The estimates were obtained by the method of marginal maximum likelihood using MULTILOG. The asymptotic standard error of the estimate is in parentheses.

*The items are listed in the order in which they appear in the Appendix.

*Read the threshold parameter, say $b_2$, as the "cutting point" between the strongly disagree (coded as 1) and disagree (coded as 2) categories.

When either the sample size is very large (as is rarely the case in marketing) or the number of items is small (as is usually the case in marketing), in particular, for the three-item subsets used for testing unidimensionality (see Figure 2), most of the $G^2$ cells are nonempty and consequently the $G^2$ statistic is likely to yield relatively reliable results. In fact, the $G^2$ values obtained by fitting a graded-response model to each of the four subsets (all with d.f. = 197) were: subset 1, 2, 9, 10 = 246 ($p > .01$), subset 4, 5, 8 = 195 ($p > .1$), subset 1, 7, 11 = 253 ($p > .005$), and subset 3, 6, 12 = 226 ($p > .05$). Consequently, at a .01 level of significance, the graded LTT model cannot be rejected in three of the four subsets. In the fourth, the fit is borderline. This finding suggests a reasonably good fit to the data. The $a_i$ and $b_k$ parameter estimates along with their standard errors for each item are given in Table 1. An interpretation of these parameters as well as other features (e.g., information) of the graded-response model follows.

The Graded LTT Model: Interpretation of Item Parameters and Information Functions

Item parameters. The discrimination parameter, $a_i$, is the slope of $P_k^*(\theta)$ at the point of its inflection (see Figure 1). The steeper the slope (i.e., higher the value of $a_i$), the less the measurement error in the neighborhood of the inflection point. Note that in Figure 1 all $P_k^*(\theta)$'s for a given item have the same slope at their point of inflection, consistent with the notion that measurement error is a function of item wording and composition, and thus independent of category labels. Samejima refers to this as the "homogeneous" thinking process with an item. Consistent with this, only one $a_i$ value is utilized for each item. Lord (1980) and others have likened the $a_i$ parameter to the factor loading in a common factor model. The lower bound for $a_i$ is zero. Consequently, the $a_i$ estimates in Table 1, ranging from .83 to 2.06, suggest that the items are reasonably good discriminators at their point of inflection.

The threshold estimates can be interpreted by noting that, in the ICC depicted in Figure 1, $b_k$ is that value of $\theta$ for which $P_k^*(\theta) = 5$. Consider, for instance, $b_2$ for item 1. For this item, respondents with a $-3.07$ discontent level are expected to have a 50% chance of checking "disagree" (instead of "strongly disagree") or higher. Obviously, for respondents with $\theta > -3.07$, this probability increases whereas the opposite happens for still less discontent respondents. Though there is a direct counterpart for the discrimination parameter in factor analysis, the threshold parameters are unique to LTT models. The estimates in Table 1 clearly suggest that the thresholds can neither be treated as equal across items.
nor perceived as equidistant from each other. Thus, critical information about item responses is missing in most current measurement techniques in marketing.

Lord (1980), Samejima (1969), and others posit that under LTT, item parameters are invariant properties of items and ought not to be different when estimated in different samples and/or populations. As a direct consequence, LTT permits delineation of the confounding effects due to measurement bias when one is conducting comparisons across different populations of interest. Because estimation methods impose a standardization on $\theta$, in practice item parameters estimated from two different samples would not be directly comparable (Lord 1980). Instead, these estimates would be related by multiplicative or additive factors. In particular, the $b_k$ estimates are expected to be related across samples by additive constants (i.e., linearly). In contrast, the $a_i$ estimates are expected to be related across samples by multiplicative factors (Lord 1980).

Scale and item information functions. Samejima extended Birnbaum’s and Lord’s notions of “information” to items with graded response categories. She defines the information function for the $k^{th}$ category of item $i$ as

\begin{equation}
I_k(\theta) = \left( \frac{P'_k(\theta)}{P_k(\theta)} \right)^2 / \sigma_k(\theta)
\end{equation}

where $P'_k(\theta)$ is the first derivative of the OC for the $k^{th}$ category. These “information” functions (i.e., $I(\theta)$) characterize the standard error of the $\theta$ estimates. Mathematically, the standard error for $\theta$ equals $I(\theta)^{-1/2}$. Thus, the more the “information” about $\theta$, the less the measurement error and more precise the estimate of a respondent’s location on the latent trait. In addition, information functions have several distinct features under LTT.

First, information functions have an additive property. That is, the information function for any given item is the sum of the information obtained from each of its response categories. Likewise, the information function for a scale of $n$ items is the sum of $n$ individual item information functions.

\begin{equation}
I_i(\theta) = \sum_{k=1}^{m} I_k(\theta)
\end{equation}

\begin{equation}
I(\theta) = \sum_{i=1}^{n} I_i(\theta) = \sum_{i=1}^{n} \sum_{k=1}^{m} I_k(\theta)
\end{equation}

where $I_i(\theta)$ and $I(\theta)$ are the item and scale information functions, respectively. Because of this additive property, the exact contribution of any given item or additional category to the “information” about $\theta$ can be ascertained. Conversely, items and/or categories can be added or deleted to achieve desired levels of measurement error.

Second, information for any category, item, or scale is not constant. Instead, it has a functional form varying with the value of $\theta$. This feature is evident from Figure 3, which depicts the item ($I_i$) information functions and the standard error of measurement for two typical dis- content items. The corresponding functions for all of the 12 items also are displayed. A clear implication of this property is that any given item performs effectively (i.e., low measurement error) only for a specific range of $\theta$, and outside that range the item behaves rather poorly (i.e., high measurement error). For instance, note that item 10 behaves poorly for discontent values greater than 1.5. When the 12 items are considered together, the measurement of consumer discontent is more accurate in the range −3.0 to 1.5. Outside this range the information falls off sharply, implying greater measurement error. This feature suggests that when the goal of research is to measure effectively throughout the range of $\theta$, it is inappropriate to have items that are worded similarly or with merely minor variations. Rather, scales should be composed of items with nonoverlapping information functions.

Third, information is a function of item parameters. That is, the higher the $a_i$ value, the greater the information maxima. Note in Figure 3 that item 10 provides more information than item 1. The reason is that the former item discriminates better ($a_1 = 1.58$) than the latter item ($a_1 = 1.21$). However, the width of the information functions is related to the spread in the threshold parameters. Item 1 yields a much wider information function as a result of its greater range of threshold parameters. Thus, the amount of scale information is influenced by the quality and number of items in that scale. These information curves are particularly useful in adaptive survey designs.

IMPLEMENTING ADAPTIVE DESIGNS FOR LIKERT-TYPE DATA

The adaptive survey capitalizes on the differences in information functions for the individual items. It is obvious from Figure 3 that, in general, different items provide different amounts of information at different levels of $\theta$. Consequently, in principle, if a priori we have a rough idea about a respondent’s $\theta$ (say $\theta_r$), we can adapt our questionnaire by asking only those questions that provide significant information about $\theta$. Conversely, those items that do not provide much information about $\theta$ could easily be dropped from respondent $r$’s questionnaire, without significantly compromising the quality of measurement. In this way, surveys can be conducted in which the questionnaires are adapted to each and every respondent. A distinct feature of these designs is that different respondents are usually given different items, which in turn are generally subsets of the entire scale; yet their location on the same underlying construct is assessed.

The preceding principle can be implemented effectively when the survey is conducted via a computerized facility, such as a telephone interview and computer-based questionnaire and response-entry facility. In addition, parameter estimates (i.e., $a_i$ and $b_k$) for all of the scale items (the “item bank”) must be known from previous studies (i.e., “calibrated”). Weiss (1985) suggests that
Figure 3
ITEM INFORMATION FUNCTIONS FOR SELECTED DISCONTENT ITEMS AND THE SCALE INFORMATION FUNCTION FOR ALL DISCONTENT ITEMS, WITH THE CORRESPONDING STANDARD ERROR OF MEASUREMENT

Item Information Function and Standard Error of Measurement for Discontent Item #1

Information Standard Error

Discontent

--- Information --- Standard Error

Item Information Function and Standard Error of Measurement for Discontent Item #10

Information Standard Error

Discontent

--- Information --- Standard Error

Total Scale Information Function And Standard Error Of Measurement For All Discontent Items

Information Standard Error

Discontent

--- Information --- Standard Error
the adaptive designs can be implemented by programming the computer to:

—utilize and store a previously calibrated item bank containing items of sufficient number, and with appropriate characteristics (the item bank decision),
—start by asking a few questions (typically between 1 and 5) that are typically of medium “difficulty” (the entry level decision),
—make effective use of knowledge of the respondent’s responses to the previous items to select the specific item to be administered next (the item selection decision),
—continue the preceding process and compute an estimate of the respondent’s θ, as well as the standard error for this estimate, at each stage (the trait estimation decision), and
—terminate the preceding process when some predetermined criterion is satisfied (the termination decision).

Examples of such implementations for dichotomous items have been provided by Weiss (1982) and Balasubramanian and Kamakura (1989). As noted before, little attention has been directed to implementing such procedures for attitude scales assessed by using Likert-type response categories. In this section, we discuss a simulation study that is an initial step toward addressing this gap. First, we describe the adaptive design, detailing the decision used at each step of the process. We then present the strategy for evaluating the results from the adaptive design procedure. Finally, we present the results and examine them in light of the evaluation strategy.

Key Decisions in Adaptive Designs

Weiss and Kingsbury (1984) discussed decision rules for adaptive designs with dichotomous items. Similar rules for polytomous items have not yet emerged. As a result, the rules we enumerate here are based on basic principles of adaptive designs, and in some cases on reasonable extensions of previous work.

Item bank. The quality and size of the item bank are critical elements in the potential payoffs from adaptive designs. Urry (1977) has suggested that, in terms of quality, the item bank must contain items with a_ values of at least .80, and the threshold values should be distributed evenly and widely (e.g., over a −2.0 to +2.0 range of θ values). The first condition guarantees that each item provides some minimum information about θ and the latter assures that this information is available across a broad range of θ values. Such an item bank allows individuals with varying trait values to be estimated reliably and efficiently.

The preceding conditions are equally valid for item banks based on the graded-response LTT model. For the discontent items, Urry’s condition for the a_j parameter is met because all estimates in Table 1 exceed .80. However, for the threshold estimates, the table reveals that mostly negative threshold values dominate. In fact, the threshold estimates are evenly and widely distributed only for item 1. For most of the remaining items, the threshold estimates cover a much wider range for negative values of θ. This finding is similar to that of a recent study in marketing showing that the discontent measure was more “informative at the middle attitude levels” (Balasubramanian and Kamakura 1989, p. 316).

Urry also suggests that, for dichotomous data, item banks must contain at least 100 items for implementing adaptive designs. Because Likert-type data yield more information, smaller item banks are likely to suffice for adaptive designs based on the graded-response model. Few researchers have examined adaptive designs for Likert-type data, but Dodd, Koch, and De Ayala (1989, p. 141) have conducted some initial studies indicating that “substantially smaller item pools can be used successfully,” and in particular item banks consisting of “30 or fewer items” can be implemented by utilizing the graded LTT model. Though banks of this size can be developed for marketing constructs, most current marketing scales are likely to consist of far fewer items, typically between 10 and 20. Adaptive designs with item banks of this size have not yet been documented.

Entry level. Ideally, any item can be used to start the adaptive design procedure. However, efficient designs (i.e., designs that achieve the termination criterion in fewer items) usually employ some information about the sample or the purpose of the study to decide the entry level. For instance, if the discontent items are to be administered to a sample of business students who are likely to be less discontent, items that provide more information in the −3.0 to 0.0 range of θ are good candidates for the entry level. In situations where the samples are random, Weiss suggests that items with medium threshold values could be utilized. This suggestion can be easily implemented with dichotomous items because only a single threshold value is estimated for each item. Multiple thresholds estimated in graded LTT models do not allow a straightforward extension of this rule. Note, however, that the third threshold in our study, b_3, is the cutting point between all of the “disagree” and “agree” categories. For this reason it is reasonable to utilize b_3 as the focal point for the entry level decision. Specifically, the 12 items were ordered according to their estimated b_3 values, and the item(s) at the median of this distribution was selected as the potential candidate(s) for the entry level. Accordingly, items 9 and 11 with b_3 values of −1.322 and −1.247, respectively, were selected. The choice between these items was guided by the discrimination parameter. Recall that the higher the a_j, the higher the information maxima. Because it is desirable to prefer an item that provides more information, item 9 was the preferable choice for the entry level.

Item selection. This decision centers on defining a rule for selecting the next item to be administered to the re-
respondent, given the responses to previous items. Early adaptive designs were based on variations of an intuitively appealing principle—"if pass (i.e., agree with) previous item, give the next item with higher difficulty (i.e., threshold), or else choose the one with lower difficulty." This method often is referred to as the fixed-branching method. Weiss has noted that this principle has been utilized to develop a highly efficient adaptive design called the "stratified adaptive" or "straddaptive" test. However, a problem with this procedure is that it utilizes only a single threshold value to select the next item and thus ignores additional information from the \( a_i \) and other \( b_a \) parameters.

For the case of dichotomous items, a popular procedure is the maximum information method (Weiss 1982). It involves two steps. First, the respondent's \( \theta \) is estimated from his or her responses to all of the previous items. In the second step, the next item is selected so that it provides the maximum information at the estimated value of \( \theta \). In addition, a restriction is imposed such that a particular item should not be administered twice to the same respondent. This method can be extended easily to graded-response adaptive designs by using the item information as the focal point for the selection decision. An attractive feature of this method is that it utilizes all of the LTT parameters in the item selection process, because item information is a function of \( a_i \) and \( b_a \) parameters. For this reason we selected this method to simulate the adaptive design. Note that the selection procedure lies at the heart of the adaptive design because it ensures that different respondents are likely to receive different items in different sequences, and items in turn are selected on the basis of the respondent's responses to all of the previous items.

**Trait estimation.** An estimate for each respondent's \( \theta \) and the associated standard error is obtained at each stage (i.e., after the respondent has answered an item) of the adaptive design. In general, the item parameter estimates are assumed to be known from some previous calibration study. In this situation the estimation of \( \theta \) is fairly straightforward and can be implemented by maximum likelihood or Bayesian methods (Weiss 1982). Typically, the maximum likelihood estimates (MLE) are obtained by numerically finding the value of \( \theta \) that maximizes the natural logarithm of the likelihood function for a given response pattern, \( V \), of an individual respondent. Maximum likelihood (MLE) estimates are asymptotically unbiased (Samejima 1969). However, unique MLE estimates are not available for patterns of responses that are all "strongly agree" (i.e., 666 \ldots) or "strongly disagree" (i.e., 111 \ldots). Such patterns are likely to occur for some respondents during the initial stages of the adaptive design process. The use of Bayesian methods can address this problem. Unfortunately, the Bayesian methods yield estimates that are not unbiased and equi-precise for different levels of \( \theta \) (Weiss and McBride 1984). Because of this limitation, we use the MLE method for trait estimation in our simulation study.

An important property of LTT models is that the estimation of \( \theta \) can be obtained independently of the specific items answered by the respondent. In particular, Lord (1980) has noted that when an LTT model fits the data, the MLE estimates are asymptotically unbiased estimates of their "true" values, irrespective of the number of items answered. Some researchers refer to this as the "item-free" property. It is put to use effectively in adaptive designs where different respondents are expected to answer different questions (usually in terms of both number and type) and yet each respondent's location along the same trait continuum is to be estimated. In sharp contrast, in conventional survey designs all respondents must respond to the same questions in order to determine their comparative standing on the underlying construct.

**Termination rule.** Researchers can choose from several different termination criteria. The adaptive design can be terminated after a respondent's \( \theta \) has been estimated to a prespecified degree of precision. That is, one could initially choose the amount of standard error that would be tolerated in estimating \( \theta \) and then terminate the adaptive design when this target is attained or when the item bank is exhausted. Alternatively, a minimum information criterion can be used whereby the adaptive design will terminate unless the next selected item yields some minimum level of information about \( \theta \). In addition, different rules could be employed at different ranges of \( \theta \), if needed—for instance, if more precise estimates are needed in some prespecified range of \( \theta \) values (e.g., identifying high discontent respondents). Thus, one could utilize the "precision" criterion and specify a lower standard error for, say, high discontent levels (e.g., \( \theta > 1.0 \)) and a higher value for all other discontent levels. Clearly, the nature of the sample and purpose of the study should guide the choice of the termination criterion.

Because the purpose of our study is to illustrate the adaptive design via a simulation study, the design was continued until there were no additional items to be administered. This approach allows the examination of the precision of the adaptive-design-based \( \theta \) estimates as a respondent progresses through the various stages. Also, the adaptive design can be evaluated from several perspectives.

**Method and Evaluation Strategy**

In accord with the preceding decisions, an adaptive design was simulated for phase 2 data. These data consist of 235 observations. A PC program ADLMAX (Adaptive Design for Likert data by the method of Maximum information) was written to simulate the responses for each respondent as she or he progressed through the various stages of the adaptive design.\(^6\) The stages were

\(^6\)Note, however, that the actual responses were obtained by asking the discontent items in some fixed sequential order. To the extent that "order" bias is present in the data, the simulation reported here would be limited in its ability to mimic actual adaptive designs.
numbered to reflect the number of questions answered by the simulated respondent (e.g., at stage 5, the respondent would have answered five questions). At each stage the simulated respondent’s \( \theta \) was estimated along with its standard error and stored in a file for further analysis.

The results of the preceding simulated adaptive design were evaluated at both the individual and the aggregate level. At the individual level, plots of the \( \theta \) estimate and the corresponding standard error at different stages were examined for at least 50 respondents. These respondents were selected systematically so that the impact of the adaptive design could be ascertained for individuals with varying levels of discontent. At the aggregate level, fidelity correlations were examined initially at each stage. Fidelity correlations are simply the Pearson moment correlations between estimated values of \( \theta \) and its corresponding “true” values. In practice, the “true” values on a latent construct are, by definition, unobservable. Therefore, for the purposes of our study, the MLE estimate of \( \theta \) based on all of the 12 items was utilized as the (imperfect) surrogate for the “true” value. Because in reality this surrogate “true” value is also contaminated by measurement error, the fidelity correlations reported here should be viewed as lower bounds for their corresponding “true” values. For the adaptive design to be efficient, it should yield “high” fidelity correlations by using fewer items, that is, in fewer than 12 stages. Clearly, it is nearly impossible to obtain fidelity correlations of 1.00 with fewer items; however, it is desirable to expect values greater than .90. Furthermore, the fewer the stages required to achieve this level, the more efficient the adaptive design.

Following Crichthon (1981), we also evaluated the adaptive design by computing the bias \( B(\theta) \), root mean square error \( \text{RMSE}(\theta) \), and inaccuracy of the \( \theta \) estimates \( \text{IA}(\theta) \):

\[
B(\theta) = \frac{\sum(\hat{\theta}_i - \theta)}{N},
\]

\[
\text{RMSE}(\theta) = \sqrt{\frac{\sum(\hat{\theta}_i - \theta)^2}{N}},
\]

\[
\text{IA}(\theta) = \sqrt{\frac{\sum(\hat{\theta}_i - \theta)}{N}},
\]

where \( \hat{\theta}_i \) represents the estimates of \( \theta \) for each individual, at each stage of the process, and \( \theta \) denotes a surrogate for the “true” value for that individual. These statistics then were plotted against the stages of the adaptive design. As in the case of fidelity correlations, the adaptive design would be efficient to the extent that it can yield negligible bias (B), root mean square error (RMSE), and inaccuracy (IA) by utilizing fewer than 12 items (i.e., stages). Following Weiss, we considered RMSE and IA values of .20 or less as negligible. Likewise, B values of \( \pm .10 \) or less were deemed insignificant.

Finally, the extent to which the preceding results are robust to different values of \( \theta \) was examined. This issue is important because a design is less attractive if it is efficient for only a limited range of \( \theta \) values. In fact, outside that range the design may be highly inefficient.

In contrast, designs that are efficient for a wide range of \( \theta \) values are clearly preferable. Such designs are said to have wide “bandwidth” (McBride 1976). Adaptive designs are likely to be efficient over a wide bandwidth because items are adapted to each and every individual, presumably representing different values of \( \theta \). To examine this issue, we divided the sample into three groups corresponding to low, medium, and high levels of discontent based on the 33 and 66 percentile cutpoints. The fidelity correlations, B, RMSE, and IA then were evaluated separately for each of the three discontent groups. The notion that adaptive designs have wide bandwidth would be supported if this design yielded efficient estimation of \( \theta \) irrespective of the discontent level.

**Results**

**Nonconvergence cases.** For the MLE estimation, the convergence criterion was set at a value of .001 and a maximum of 25 iterations were allowed. However, 17 cases (five each from the low and medium groups and seven from the high discontent group) yielded persistent problems in producing MLE estimates, and convergence was not obtained even after 25 iterations at several stages. For this reason these cases were excluded from subsequent analysis. In addition, convergence could not be obtained in the initial stages for respondents who consistently responded at the scale extremities (i.e., 111 . . . or 666 . . . ). After the first stage, nonconvergence was obtained for 55 respondents. By stage 4, convergence was obtained for all respondents, with the exception of four respondents in the high discontent group. Finally, by stage 7, none of the respondents yielded a nonconvergent solution. These cases were included for analysis when the corresponding MLE estimates could be computed.

**Individual analysis.** The plots for the estimate of \( \theta \) and its standard error versus the adaptive design stages for each of the 50 respondents we examined are too numerous to reproduce here. Representative plots for four respondents are depicted in Figure 4. Note that these respondents have different levels of discontent, respondent 128 being the least (final \( \theta = 1.399 \)) and respondent 159 being the most (final \( \theta = 1.541 \)) discontent. This variation is useful in evaluating the effectiveness of the adaptive design at different levels of \( \theta \). Also note that for respondents 96 and 128, the MLE estimate of \( \theta \) was not available at the first stage because of extreme responses. However, this problem was resolved at the second stage.

In general, Figure 4 indicates that initially the estimate of \( \theta \) is inaccurate, with wide 95% confidence bands. However, as the respondent progresses through the adaptive design, the point estimates as well as the confidence bands quickly approach their “true” values. Specifically, by the eighth stage the adaptive design is able to estimate a respondent’s discontent level with nearly the same accuracy (within \( \pm .20 \) of point estimates) and reliability (within \( \pm .10 \) of standard error) as the entire
Figure 4
INDIVIDUAL-LEVEL ANALYSIS FOR THE SIMULATED ADAPTIVE DESIGN: PLOTS FOR SELECTED RESPONDENTS

Theta Estimate and 95% Confidence Bands for Respondent #128

Theta Estimate and 95% Confidence Bands for Respondent #96

Theta Estimate and 95% Confidence Bands for Respondent #24

Theta Estimate and 95% Confidence Bands for Respondent #159
set of 12 items. Furthermore, the preceding pattern appears to emerge irrespective of the particular discontent level of the individual respondent. For instance, consider respondent 159. After the initial item, this respondent's discontent level was estimated as \( \theta = .45 \), with a standard error of 1.085. At the end of the fifth stage, however, this person's \( \theta \) was revised to 1.613, with the standard error declining by more than 30% to .739. Upon continuing this process, by the eighth stage this respondent's \( \theta \) was estimated at 1.377 in comparison with the "true" value of 1.541. Likewise, the standard error was .417 and the corresponding "true" value was .421. This example suggests that the point estimates and confidence intervals nearly level off by the eighth stage of the adaptive design.

**Aggregate analysis.** The aggregate analysis affords more precise evaluation of the adaptive design. The results are displayed graphically in Figure 5. Note that only 11 stages are displayed, because at the twelfth stage all estimates attain their "true" values.

In terms of the fidelity correlations, note that by stage 8 the correlations exceed .90 for the low and high discontent groups. For the medium discontent group, however, the fidelity correlations achieve the target value only by stage 10. For the case of bias in \( \theta \) estimates, the \( B(\theta) \) values achieve target values of \( \pm .10 \) by stage 5 in all three discontent groups. It is noteworthy that, though for the medium and high discontent groups the adaptive design begins with a fairly high bias, significant reductions are obtained rather quickly as the adaptive design progresses. In a similar way, the inaccuracy in \( \theta \) estimates becomes negligible (i.e., < .20) by stage 8 of the adaptive design, irrespective of the discontent group. Likewise, the target value of < .20 for the RMSE criterion is achieved by stage 8 in each of the three discontent groups. These findings suggest that, irrespective of the discontent group, by stage 8 the adaptive design yields measurement of nearly the same fidelity (the only exception being for the medium group) and with nearly the same bias, inaccuracy, and RMSE as are obtained in 12 stages of a conventional survey design. This outcome represents an efficiency of about 33% (i.e., 8/12).

The results for the medium discontent group are noteworthy. For this group, though the target values for bias, RMSE, and inaccuracy are achieved by stage 7 (the earliest among the three groups), the fidelity correlation attains its target value only at stage 10. More importantly, note that for this group the adaptive design produces the lowest RMSE and inaccuracy at every stage of the design after the initial two stages. This pattern is consistent with our previous conclusion that the discontent items provide more information in the middle range of \( \theta \). Furthermore, for the low discontent group, the RMSE value drops below .20 by stage 8 but rises to .25 in stage 9. Other measures of design efficiency do not change appreciably. The RMSE therefore appears to be a sensitive indicator for the efficiency of the adaptive design. This result, combined with that of the medium discontent group, underscores the need for using multiple criteria to evaluate adaptive designs.

In summary, the results of the simulation indicate that the graded-response adaptive design can be applied successfully to relatively small item banks because target values for most criteria are obtained by the eighth stage for each of the discontent groups.

**DISCUSSION AND CONCLUDING COMMENTS**

We elucidate a measurement model for Likert-type data, the graded-response LTT model, and use this model to illustrate a novel approach for implementing efficient marketing surveys—the adaptive design procedure. The novelty of this approach lies in the fact that the design continually modifies itself as a function of the responses to preceding questions for each individual respondent, with a goal of yielding a more efficient design. The underlying philosophy is likened to the old adage, "From each according to his/her attitude; To each according to his/her attitude" (Wainer 1989). It is in sharp contrast to the philosophy of current approaches in marketing, which can be likened to "one size fits all" (Balasubramanian and Kamakura 1989) because all respondents are required to answer all questions, irrespective of their attitude level.

In examining the findings, however, readers should bear in mind two unique aspects of our research. First, we explicitly consider Likert-type data. Previous research has tended to utilize binary data, consequently relying solely on binary LTT models. As a result of this emphasis, procedural guidelines and general recommendations for binary adaptive designs have been established, but similar advances have not been made for graded-response adaptive designs. Preliminary results by Dodd and her associates suggest that important differences are likely to emerge for adaptive designs that use Likert-type data. Second, our study illustrates adaptive designs with as few as 12 items. Unlike binary designs, which require more than 100 items, adaptive designs based on the graded-response model may require substantially smaller item pools because the use of Likert-type data provides more information across the full range of the theta scale. Despite their considerable promise for marketing applications, adaptive designs with fewer than 30 items have not yet been documented. For these reasons, our study should be viewed as an initial one in the application of a graded-response adaptive design for Likert-type marketing data.

Overall, the findings suggest that the graded-response adaptive design performed reasonably well on several different criteria. At the aggregate level, the adaptive design is found to be efficient because by the eighth stage the indices for fidelity, bias, RMSE, and inaccuracy are within predetermined limits, indicating negligible deviation from their corresponding "true" values. There is only one exception (i.e., fidelity for the medium group). This pattern of leveling off by the eighth stage is further reinforced by the findings from individual analysis. Taken
Figure 5
AGGREGATE-LEVEL ANALYSIS FOR THE SIMULATED ADAPTIVE DESIGN: PLOTS FOR THE LOW, MEDIUM, AND HIGH DISCONTENT GROUPS

Performance of Adaptive Design
Low Discontent Group

Performance of Adaptive Design
Medium Discontent Group

Performance of Adaptive Design
High Discontent Group
together, our results are encouraging for applications of adaptive designs to Likert-type data and suggestive of efficiencies, even with relatively small item banks, of the order of 33% in comparison with conventional survey designs. Hence researchers are likely to find it practical to exploit the additional information from Likert-type data to overcome a potential impediment (i.e., item bank size) in the implementation of binary adaptive designs. Furthermore, with better quality item banks, efficiencies of the order of 50% (as reported for binary designs) with significantly smaller item banks appear plausible. We therefore seem to have enough evidence at this point to indicate that efficient marketing surveys can be implemented practically with relatively small item banks and that the graded-response adaptive designs warrant the serious attention of researchers and practitioners.

Such efficiencies in implementing marketing surveys are likely to yield several direct and indirect payoffs. In terms of direct payoffs, cost savings are likely because an efficient adaptive survey promises to reduce interviewing time, thereby reducing survey administration costs (e.g., interviewer and supervisory time). In addition, because usually fewer items need to be answered (for comparable precision), data handling and manipulation costs also are expected to decline as a direct result of adaptive surveys. These direct cost savings are likely to be of the same order of magnitude as efficiencies obtained with simulation studies (e.g., between 33% and 50%); hence the direct payoffs from adaptive designs should be substantial. Furthermore, researchers are also likely to obtain indirect payoffs. Possibly, the quality of survey data might improve because an adaptive design alleviates respondent fatigue and helps make the survey more interesting, as respondents need answer only questions that are relevant to their individual attitude level. In addition, as marketers adopt adaptive designs, they are likely to accelerate the penetration of CAQ/CRT interviewing technology. The result may be added indirect payoffs because a computerized survey approach should yield cost savings (in comparison with conventional methods) from reductions in questionnaire design, coding, data entry, and other associated costs.

However, the exceptions encountered in our study pinpoint areas for followup studies and further research. One problem area is the nonconvergence cases at the early stages of the adaptive design. To mitigate this problem, researchers may consider starting with two items, one that is relatively easy (i.e., low $b_a$ value) and another that is relatively difficult (i.e., high $b_a$ value). The choice of two items differing in difficulty is likely to produce fewer responses at the extremities. An alternative approach is to employ a Bayesian estimation procedure rather than an MLE method, especially in the early stages of the adaptive design (Weiss 1982). In this strategy the Bayesian method is utilized until an MLE procedure is possible, then the MLE is used for the remaining stages. Further simulation research is needed, however, to evaluate the effectiveness of such alternative approaches.

The quality of the item bank is another important concern. The problem seems to be that for some range of $\theta$ values (e.g., high discontent level in our study), no or relatively few items are available in the item bank. Because most, if not all, marketing scales have not been developed or analyzed by utilizing LTT models, we believe that gaps in the scale information characteristics are likely to go unrecognized and may be fairly common. Some construct developmental work therefore may be necessary before optimal efficiencies from adaptive designs can be obtained. For the discontent scale, additional items that provide information throughout the range of $\theta$—thereby providing a bank of items with a wide bandwidth—are clearly needed. Finally, in terms of evaluating adaptive designs, our results appear to suggest that it is useful to (1) apply multiple criteria and (2) consider explicitly different levels of the underlying trait. Further research is needed, however, to study the properties of various evaluative measures. In addition, the use of multiple discontent groups is useful in localizing problem areas and in evaluating whether the adaptive design offers efficiency over a wide bandwidth. The results of the simulation reported here are indicative of efficiencies over a wide bandwidth with the ADLMAX design, as a 33% efficiency is obtained in each of the three discontent groups.

At a more general level, our article introduces marketers to the graded-response LTT model for Likert-type data. As a generalization of the common factor model, the LTT model is shown to make assumptions about the measurement process (e.g., nonlinearity, nonmetric data) that are not unrealistic for most marketing data. Furthermore, the LTT model affords numerous advantages over current measurement approaches in marketing. Perhaps the most important advantage is that one can estimate a respondent’s attitude from any subset of items that have been previously calibrated. The fact that the underlying construct can be estimated independently of the particular choice or number of items represents a major breakthrough in the area of construct measurement. This feature underlies the adaptive design illustrated here and provides a systematic approach for reducing the length of survey instruments without the attendant loss of information. Another significant advantage is captured in the notion of information function, which allows the assessment of the measurement error at each level of the underlying trait. This notion of information lies at the heart of the adaptive design described here because the very feasibility of such designs depends on effectively capitalizing on the differences in information functions for different items. Because of these advantages, and especially considering the significant payoffs from adaptive designs, we believe marketing scientists must examine closely the graded-response LTT model’s potential contribution to implementing efficient marketing surveys in particular and to measurement in marketing in general.
APPENDIX

CONSUMER DISCONTENT ITEMS USED FOR LIT ANALYSIS AND ADAPTIVE DESIGN

1. I am often dissatisfied with a recent purchase.
2. It is hard to understand why some brands are twice as expensive as others.
3. Business profits are high yet they keep on raising their prices.
4. An attractive packagemany times influences a purchase that is not necessary.
5. Companies "jazz up" a product with no real improvement, just to get a higher price or sell more.
6. Companies try to influence government just to better themselves.
7. As soon as they make a sale, most businesses forget about the buyer.
8. Stores advertise "special deals" just to get the shopper into the store to buy something else.
9. All business really wants to do is to make the most money it can.
10. Prices of products are going up faster than the incomes of ordinary customers.
11. Chain stores are getting so big that they don’t treat the consumer personally.
12. Industry is not cleaning up the waste it has been dumping.

REFERENCES


Reprint No. JMR273104