

SAJEEV VARKI, BRUCE COOIL, and ROLAND T. RUST*

In marketing, qualitative data are used in theory development to investigate marketing phenomena in more depth. After qualitative data are collected, the judgment-based classification of items into categories is routinely used to summarize and communicate the information contained in the data. In this article, the authors provide marketing researchers with a method that (1) provides useful substantive information about the proportion and degree to which items belong to several categories and (2) measures the classification accuracy of the judges. The model is called the fuzzy latent class model (FLCM), because it extends Dillon and Mulani's (1984) latent class model by freeing it from the restrictive assumption that all items are crisp for a given categorization. Instead, FLCM allows for items to be either crisp or fuzzy. Crisp items belong exclusively to one category, whereas fuzzy items belong—in varying degree—to multiple categories. This relaxation in the assumption about the nature of qualitative data makes FLCM more widely applicable: Qualitative data in marketing research are often fuzzy, because they involve open-ended descriptions of complex phenomena. The authors also propose a moment-based measure of overall data fuzziness that is bounded by 0 (completely crisp) and 1 (completely fuzzy).

Modeling Fuzzy Data in Qualitative Marketing Research

In marketing, qualitative data are used in theory development for investigating marketing phenomena in more depth. In a review of all articles appearing in *Journal of Marketing Research*, *Journal of Marketing*, and *Journal of Consumer Research* over the ten-year period from 1988 to 1997, researchers in almost 100 articles have content-classified qualitative data into nominal categories to help summarize and communicate the information they contain (Spiggle 1994). Specific examples include Hastak and Olson's (1989) analysis of respondents' verbal protocols in cognitive response research; Bitner, Booms, and Tetreault's (1990) use of the critical incident technique to determine customers' reasons for satisfaction and dissatisfaction; and Gross and Sheth's (1989) content analysis of advertisements to identify salient advertising themes (see Table 1).

*Sajeev Varki is an assistant professor, University of Rhode Island (e-mail: svarki@uriacc.uri.edu). Bruce Cooil is an associate professor, Owen Graduate School of Management, Vanderbilt University (e-mail: bruce.cooil@owen.vanderbilt.edu). Roland T. Rust is David Bruce Smith Chair in Marketing, Robert H. Smith School of Business, University of Maryland (e-mail: rrust@rhsmith.umd.edu). The authors thank Max Woodbury, Gautam Biswas, Terry Elrod, and Susan Keaveney for their help. Financial support for the project was provided by the Center for Service Marketing, Vanderbilt University.

Percentage agreement is currently the most popular method among marketers for assessing the quality of categorization employed. However, the drawbacks of percentage agreement, and many other methods that summarize reliability as a single statistic, are that they (1) do not correct for chance agreements, which can be quite substantial (Perreault and Leigh 1989); (2) do not provide information on whether interjudge reliability is low because of judges' lack of expertise or because of ill-defined categories; and (3) do not provide sufficient diagnostic detail to help researchers improve the categorization scheme. Thus, when interjudge reliability is low, these methods do not provide the researcher with guidance on how to train the judges better, nor do they even indicate whether the true character of the items is accurately summarized in the selected classification scheme.

A model that addresses some of the shortcomings of percentage agreement is Dillon and Mulani's (1984) latent class model (LCM). It treats judges' classifications as probabilistic outcomes and provides estimates of the probability with which each judge classifies an item in a given category, conditional on the true (latent) category to which that item belongs. However, LCM presupposes that every item is crisp; that is, it belongs exclusively to one of the several categories employed, though often the sample of items classified is a mixture of crisp and fuzzy items. Here, fuzzy items

Table 1
MARKETING APPLICATIONS OF JUDGMENT-BASED NOMINAL CLASSIFICATION IN THREE MAJOR JOURNALS BETWEEN 1988
AND 1997^a

<i>Substantive Area^b</i>	<i>Journal of Marketing</i>	<i>Journal of Marketing Research</i>	<i>Journal of Consumer Research</i>
Sales management	Leigh and McGraw (1989)	Sujan, Sujan, and Bettman (1988); Brown and Peterson (1993)	Munch and Swasy (1988)
Advertising and media strategy	Gross and Sheth (1989); Mackenzie and Lutz (1989); Weinberger and Spotts (1989); Mazis et al. (1992); Drumwright (1996)	Finn (1988); Edell and Keller (1989); Janiszewski (1990); Keller (1991); Unnava and Burnkrant (1991a, b); Meyers-Levy and Peracchio (1992); Singh et al. (1994); Derbaix (1995); Pechmann (1996); Baumgartner, Sujan, and Padgett (1997)	Cox and Cox (1988); Hastak and Olson (1989); Jacoby and Hoyer (1989); Ford, Smith, and Swasy (1990); Maheswaran and Sternthal (1990); Scott, Klein, and Bryant (1990); Olney, Holbrook, and Batra (1991); McQuarrie and Mick (1992); Mick (1992); Barlow and Wogalter (1993); Brown and Rothschild (1993); Meyers-Levy and Peracchio (1995)
International marketing/services	Bitner, Booms, and Tetreault (1990); Samiee and Roth (1992); Bitner, Booms, and Mohr (1994); Keaveney (1995)	Maheswaran and Meyers-Levy (1990)	—
Judgment and decision making/information processing/variety seeking/choice	Curren, Folkes, and Steckel (1992)	Beihal and Chakravarti (1989), Corfman (1991); Mahajan (1992); Armstrong and Collopy (1996)	Celsi and Olson (1988); Johnson (1989); Park and Smith (1989); Macinnis and Park (1991); Meyers-Levy (1991); Park, Milberg, and Lawson (1991); Menon (1993); Peracchio and Meyers-Levy (1994); Greenleaf and Lehmann (1995); Mitchell, Kahn, and Knasko (1995); Meyers-Levy and Peracchio (1996)
Pricing and distribution	Butaney and Wortzel (1988); Lichtenstein et al. (1990); Dant and Schul (1992)	Sethuraman (1996)	Yadav (1994)
Product/brand/marketing management/choice	Aaker and Keller (1990)	Ferrell and Skinner (1988); Leong, Busch, and John (1989); Broniarczyk and Alba (1994); Gundlach and Cadotte (1994)	Johnson (1988); Park, Iyer, and Smith (1989); Coupey (1994)
Organizational behavior/consumer behavior attitude theory/measurement/meta-analysis/survey design	Drumwright (1994); Frankwick et al. (1994)	Perreault and Leigh (1989); Hughes and Garrett (1990); Lastovicka, Murry, and Joachimsthaler (1990); Miniard, Bhatla, and Rose (1990); Buchanan (1992); Sujan, Bettman, and Baumgartner (1993); Block and Keller (1995); Menon et al. (1995)	Sheppard, Hartwick, and Warshaw (1988); Kolbe and Burnett (1991); Brown and Stayman (1992); Babin, Darden, and Griffin (1994); Richins (1994); Netemeyer, Burton, and Lichtenstein (1995)
Cross-cultural research/group and interpersonal influence/motivation/involvement/consumer expertise/persuasion/child consumers	—	—	Bearden, Netemeyer, and Teel (1989); Tse, Belk, and Zhou (1989); Mick and DeMoss (1990); Moorman (1990); Miniard et al. (1991); Meyers-Levy and Maheswaran (1992); Peracchio (1992, 1993); Rose, Bearden, and Teel (1992); Huffman and Houston (1993); Mitchell and Dacin (1996)

^aThe complete bibliography of Table 1 is available separately from the authors on request.

^bEach journal's coding scheme was used to identify the substantive area of an article. In some cases, an article was classified into more than one category, in which case we made a determination of the most appropriate categorization after reading through the article.

are defined as items that belong to multiple categories to varying extents.

Fuzzy data, or data sets that include fuzzy items, are common in content classification. This occurs whenever items vary in the intensity to which they belong to different categories, even when the categories are distinct and well defined. Note that a conceptual distinction is made here

between undesirable fuzziness that occurs because of imprecisely defined categories and the fuzziness that arises because items intrinsically belong to multiple, well-defined categories.¹ We focus on a way of modeling the latter situa-

¹We gratefully acknowledge an anonymous reviewer for making this distinction.

tion. For example, consider a comprehensive “pick any” coding scheme in which each judge indicates whether each of M unique attributes is present or absent in an item. Viewed within the classification framework, the judges simply pick from among the possible 2^M profiles (or categories) the profile (or category) that best matches the item profile. In such a scheme, it is still possible for some items to be fuzzy, in that they are more appropriately represented as mixtures of the 2^M pure profiles. Thus, fuzziness is an item characteristic for a given categorization or classification scheme.

In this article, we develop the fuzzy latent class model (FLCM), which provides the researcher with estimates of (1) the proportions of fuzzy and crisp items, (2) the proportions of crisp items in each category, (3) the mixture distribution of the fuzzy items, and (4) the probabilities with which judges classify any type of item to each category. Thus, FLCM not only provides the researcher with a way of identifying judges’ areas of least accuracy in classifying items but also provides useful substantive information about the mixture character of the data. Another strength of this model is that it only requires each judge to classify an item into the single nominal category that is deemed most appropriate and thus does not require any more information than is already collected in applications in which LCM is used.²

Because FLCM subsumes LCM as a special case, it can be applied to data sets in which all items are crisp. In contrast to LCM, however, FLCM is also appropriate when some proportion of items are fuzzy. Note that LCM does not provide consistent estimates of the classification probabilities if the data set includes fuzzy items.

Our model is related to Manton, Woodbury, and Tolley’s (1994) general grade-of-membership (GoM) model, which has been used to identify the extent to which people suffer from various mental disorders, on the assumption that patients’ responses to standard psychiatric scales are influenced by the extent to which the patients suffer from multiple diseases. Our model is possibly among the first applications of GoM in marketing. However, two points need to be made. First, the model parameterizations are anchored in different substantive contexts, which alters their interpretation. In GoM, patients are considered fuzzy, whereas in FLCM, items are considered fuzzy. In GoM, patient’s fuzziness influences their probabilities of responding to an ordered categorical scale, whereas in FLCM, item fuzziness influences the classification probabilities of an external set of judges. Second, in GoM, there is no relationship between the various mental states (categories) a patient can be in and the number of scale points of the Likert scale used, whereas in FLCM, there is an equivalence between the categories of membership and the categories of classification—a restriction imposed by the logic of the classification context.

The rest of the article is organized as follows: First, we develop our model (FLCM) and show how it subsumes Dillon and Mulani’s (1984) LCM approach as a special case when all items are crisp. Second, we develop a measure of the fuzziness level of the data set and show how it can be

obtained from FLCM estimates. This measure and the estimated proportion of fuzzy items, which are also provided by FLCM, both provide ways of testing whether a data set is significantly fuzzy. Third, we test FLCM on marketing data and compare the results with those of LCM. We conclude with a discussion of the results and suggestions for further research.

FLCM

In this section, we relate the GoM parameters to the context of judgment-based classification.

Assumption 1: Each judge j is assumed to classify item i independently of the other judges. That is, judges are assumed to not discuss an item’s classification among themselves; this is an assumption that is relatively easy to ensure in practice. Similar assumptions are made by Cooil and Rust (1995), Dillon and Mulani (1984), Perreault and Leigh (1989), and Rust and Cooil (1994).

Assumption 2: Items to be classified may be fuzzy (complex). The extent/degree to which an item i is a member of category k is captured by g_{ik} , the grade of membership of i in k , subject to the following constraints:

$$(1) \quad \sum_{k=1}^K g_{ik} = 1, \quad 0 \leq g_{ik} \leq 1.$$

Thus, g_{ik} indicates the extent to (or intensity with) which each item i is a member of category k . This is the GoM parameterization for fuzziness and is adopted by Wedel and Steenkamp (1989, 1991). The reader may note that the g_{ik} are unknown (latent) and represent crisp items whenever they take on values in the set $\{0, 1\}$.

Assumption 3: Judges are assumed to make classification errors as a function of their expertise, which is measured by their (conditional) classification probabilities for each of the categories. This assumption is similar to Dawid and Skene’s (1979) and Dillon and Mulani’s (1984) probabilistic assumption. Accordingly, the conditional probability that judge j will classify an item i that lies exclusively in category k ($g_{ik} = 1$) into category k^* is λ_{jkk^*} .

Assumption 4: The probability that judge j will classify item i , a partial member of the K categories, into a specific category k^* is a convex combination of the probabilities λ_{jkk^*} defined in Assumption 3:

$$(2) \quad p_{ijk^*} = \sum_{k=1}^K g_{ik} \lambda_{jkk^*}.$$

Note that Equation 2 models the effect of fuzziness on judges’ classifications and subsumes the LCM case, because when the GoM value, g_{ik} , for item i is 1 (so that $g_{is} = 0$ when $s \neq k$), p_{ijk^*} is simply λ_{jkk^*} , which is the LCM conditional probability. This parameterization is borrowed from the GoM model and related by us to the substantive context of judgment-based classification and implies that the GoM value, g_{ik} , of item i represents the frequency with which each judge will classify that item as if it were a crisp item from category k . It is important to emphasize that the GoM values, g_{ik} , are item-specific for a given classification scheme

²Conceivably, judges could be asked to classify items using constant sum allocation. However, this would require considerably greater effort from the judges than the simple (and conventional) coding scheme of pick just one category.

and are not psychometric values that vary from judge to judge. Thus, g_{ik} is not subscripted with the index j .

Assumption 5: Each item's GoM vector $(g_{i1}, \dots, g_{iK}), 1 \leq i \leq I$, comes from the same discrete-Dirichlet mixture population, in which a proportion of (q) items are crisp (within this segment, items from category k occur with probability of $p_k, 1 \leq k \leq K, \sum p_k = 1$) and the remaining proportion $(1 - q)$ are fuzzy items from a Dirichlet distribution with parameters $\alpha_1, \alpha_2, \dots, \alpha_K$. Formally, the GoM vectors have the common probability function

$$(3) \quad f(x) = qf_1(x) + (1 - q)f_2(x), 0 \leq q \leq 1,$$

where f_1 is the discrete probability mass function $f_1(x) = f_1(x_1, x_2, \dots, x_k) = p_k$ when $x_k = 1 (x_s = 0, s \neq k)$, for $1 \leq k \leq K$; $f_1(x) = 0$ otherwise; f_2 is the Dirichlet density

$$(4) \quad f_2(x) = D(\alpha_1, \alpha_2, \dots, \alpha_K) \prod_{k=1}^K x_k^{\alpha_k - 1},$$

where

$$(5) \quad D(\alpha_1, \alpha_2, \dots, \alpha_K) \equiv \frac{\Gamma\left(\sum_{k=1}^K \alpha_k\right)}{\sum_{k=1}^K \Gamma(\alpha_k)}.$$

This GoM distribution (Equation 3) provides a flexible way of representing the possible presence of both crisp and fuzzy items. When all items are crisp ($q = 1$), this reduces to the LCM case (see Equation A6 of the Appendix). In marketing, the Dirichlet distribution has been used to model such things as advertising exposures (Rust and Leone 1984), consumer preference heterogeneity (Jain, Bass, and Chen 1990), and constant sum allocations (DeSarbo, Ramaswamy, and Chatterjee 1995).

The FLCM Likelihood

As shown in the Appendix, the FLCM likelihood may be written as

$$(6a) \quad L_{FLCM} = \prod_{i=1}^I \sum_{k_1=1}^K \sum_{k_2=1}^K \dots \sum_{k_J=1}^K \left[\prod_{j=1}^J \lambda_{jk_j k_{ij}} E \left(\prod_{j=1}^J g_{ik_j} \right) \right],$$

where

- i indexes the items in the data set, $1 \leq i \leq I$;
- j indexes the judges, $1 \leq j \leq J$;
- $k_j, 1 \leq k_j \leq K$, indexes the category from which judge j perceives GoM value g_{ik_j} when classifying item i ;
- $\lambda_{jk_j k_{ij}}$ is the (conditional) probability that judge j will classify an item that belongs exclusively to category k_j into category k_{ij} ;
- $k_{i,j}$ denotes the actual category to which judge j assigns item i ; and

$E(\bullet)$ denotes the expectation with respect to the distribution of the GoM vectors (g_{i1}, \dots, g_{iK}) .

Given the assumption that the GoM distribution is the discrete-Dirichlet mixture (Equation 3), the J th-order cross-product moments in Equation 6a are of the form

$$(6b) \quad E \left(\prod_{j=1}^J g_{ik_j} \right) = E \left(\prod_{k=1}^K g_{ik}^{m_k} \right) = (1 - q) \frac{D(\alpha_1, \dots, \alpha_K)}{D(\alpha_1 + m_1, \dots, \alpha_K + m_K)} + q \sum_{k=1}^K p_k I(m_k = J),$$

where $D(\alpha_1, \alpha_2, \dots, \alpha_K)$ is the Dirichlet function of Equation 5; $m_k, 1 \leq k \leq K$, represents the number of times the GoM value g_{ik} is the stimulus perceived by the J judges when they classify item i ; and $I(\bullet)$ is the indicator function ($I[m_k = J] \equiv 1$ if $m_k = J, 1 \leq k \leq K$, and otherwise zero). When there are no crisp items ($q = 0$), Equation 6b reduces to just the Dirichlet moments (Bishop, Feinberg, and Holland 1975, p. 405). At the other extreme, if all the items are crisp ($q = 1$), Equation 6a reduces to the LCM likelihood (see Equation A6 in the Appendix).

Necessary and Sufficient Condition for Identifiability of FLCM

A necessary condition for the identification of Equation 6 is that the number of possible classifications for each item must exceed the number of parameters to be estimated; that is, for $J \geq 2$ and $K \geq 2$,

$$(7) \quad K^J > J \times K \times (K - 1) + 2K.$$

The term on the left-hand side is the number of possible ways that J judges may classify an item among the K categories. On the right-hand side, the first term is the number of independent classification probabilities, $\lambda_{jk_k k_{ij}}$, to be estimated, and the second term is the combined number of g_{ik} distribution parameters to be estimated (there are K Dirichlet parameters in Equation 5 and K degrees of freedom for the remaining parameters of Equation 3, because the $q + p_1 + \dots + p_K = 1$). For example, the constraint in Equation 7 implies that at least four judges ($J = 4$) are needed if there are only two categories ($K = 2$). In addition to Equation 7, the number of J th-order moments of the g_{ik} distribution that is used in the FLCM likelihood must be at least as large as the number of parameters used to identify the discrete-Dirichlet distribution mixture (i.e., $2K$):

$$(8) \quad \frac{(J + K - 1)!}{J!(K - 1)!} \geq 2K.$$

A closed-form sufficient condition is not available, though identifiability is ensured (ex post) if the Hessian is positive definite (Dillon and Mulani 1984). The Hessian is not positive definite if some parameters are linear functions of other parameters or when the number of parameters exceeds the number of observations.

MEASURING THE FUZZINESS OF A DATA SET

The FLCM model provides an estimate of the proportion of fuzzy items, and this serves as one direct measure of data

fuzziness. Another useful measure of fuzziness is based on the cross-product moments of the g_{ik} distribution in Equation 3. In FLCM, we allow for the possibility that different judges may classify the same item as if it were from different categories because of the mixed nature (fuzziness) of the item, and therefore the cross-product moments are not generally zero. In contrast, LCM posits that all moments that involve cross-products (mixed moments) are zero, because each item is assumed to belong entirely in one category. The expectation of the product of all K components of the g_{ik} vectors (g_{i1}, \dots, g_{iK}), that is,

$$(9) \quad E \left(\prod_{k=1}^K g_{ik} \right) = E (g_{i1} \times g_{i2} \times \dots \times g_{iK}),$$

is the lowest-order cross-product moment that will be nonzero when (and only when) there are g_{ik} vectors that mix all K categories (i.e., $g_{ik} > 0$ for all $k, 1 \leq k \leq K$). Consequently, one natural measure, M_K , of the degree to which items are mixtures of all K categories is the K th root of the K th product moment in Equation 9 multiplied by the number of categories (K):

$$(10) \quad M_K \equiv K \cdot \left[E \left(\prod_{k=1}^K g_{ik} \right) \right]^{\frac{1}{K}}$$

$$= K \cdot \left[E (g_{i1} \times g_{i2} \times \dots \times g_{iK}) \right]^{\frac{1}{K}}.$$

In Table 2, we present the theoretical limits of M_K for several familiar distributions. This measure is zero when all the

items are crisp and close to zero when items are drawn from a bowl-shaped distribution in which most items are crisp or nearly crisp. At the other extreme, when items are equal mixtures of all the categories (i.e., all g_{ik} values are $1/K$), the measure is one, and if they are nearly equal mixtures, its value is close to one.³

When the underlying g_{ik} distribution is the Dirichlet, the cross-product moment is

$$(11) \quad D_K \equiv E \left(\prod_{k=1}^K g_{ik} \right)$$

$$= \left[\frac{\Gamma(\alpha_1 + \dots + \alpha_K)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_K)} \right] \left[\frac{\Gamma(\alpha_1 + 1) \Gamma(\alpha_2 + 1) \dots \Gamma(\alpha_K + 1)}{\Gamma(\alpha_1 + \dots + \alpha_K + K)} \right].$$

Consequently, when the g_{ik} distribution is a discrete-Dirichlet mixture (Equation 3), the measure of fuzziness is

$$(12) \quad M_K \equiv K \left[(1 - q) D_K \right]^{\frac{1}{K}}.$$

EMPIRICAL DEMONSTRATION OF FLCM

The data set consists of classifications by four judges of 314 customer descriptions of service incidents. These open-ended responses from Keaveney (1995) detail each cus-

³This measure can be evaluated over any subgroup of categories, and it would be appropriate to consider evaluating $M_L, L < K$, over a specific subset $S_L = \{k_1, k_2, \dots, k_L\}$ of L categories, where we condition on that part of the g_{ik} distribution in which $g_{ik} = 0$ whenever $k \notin S_L$. If we find the conditional measure in this way, the benchmarks of Table 2 still apply.

Table 2
FUZZINESS BENCHMARKS

	<i>Crisp Case</i>	<i>Intermediate Levels of Fuzziness</i>			<i>Maximal Fuzziness</i>
Verbal description	Probability mass concentrated in a proper subset of the K categories				Probability mass equally distributed across K categories
Mathematical description	$g_{ik} = 0$ for some k	U-shaped $\alpha_k = \delta$ for all k	Flat $\alpha_k = 1$ for all k	Peaked $\alpha_k = \delta^{-1}$ for all k	$g_{ik} = \frac{1}{K}$ for all k
$\left\{ E \left[\left(\prod_{k=1}^K g_{ik} \right) \right]^{\frac{1}{K}} \right\}$	0	$\delta \left[\frac{\Gamma(K\delta)}{\Gamma(K\delta + K)} \right]^{\frac{1}{K}}$	$\left[\frac{(K-1)!}{(2K-1)!} \right]^{\frac{1}{K}}$	$\delta^{-1} \left[\frac{\Gamma(K\delta^{-1})}{\Gamma(K\delta^{-1} + K)} \right]^{\frac{1}{K}}$	$\frac{1}{K}$
$\left\{ E \left[\left(\prod_{k=1}^K g_{ik} \right) \right]^{\frac{1}{K}} \right\}$ as $K \rightarrow \infty$	0	$\frac{e\delta}{K^{1+\delta}}$	$\frac{e}{4K}$	$\frac{1}{K}$	$\frac{1}{K}$
Fuzzy measure as $K \rightarrow \infty$ (largest-order term)	0	$\frac{e\delta}{K^\delta}$	$\frac{e}{4}$	1	1

^a δ denotes a positive constant close to zero, fixed with respect to K .
^b e denotes the transcendental number 2.718....

tomer’s reasons for switching services. Because most customers described some type of service failure, we classified responses into three categories: basic service failure; service encounter failure; and other, a category for all other responses. The basic service failure category includes core service failures (e.g., billing errors, service catastrophes) and poor responses to service failures; service encounter failure includes impolite, uncaring, and unresponsive customer service provider interactions; and other includes inconvenience, involuntary switching, and so forth. We used Keaveney’s (1995) definitions to train the judges before they classified the customer responses. We then analyzed the data with both the LCM (which assumes all items are crisp) and the FLCM models, using the constrained maximum likelihood procedure in Gauss (Aptech Systems Inc. 1995).⁴

Model Selection

To determine which model provides a more accurate interpretation, we used several model selection criteria to compare LCM and FLCM: the likelihood ratio (LR) test, the Akaike information criterion (AIC; Akaike 1974), and the Bayesian information criterion (BIC; Schwarz 1978). Note that because LCM is a nested case of FLCM, the LR test is an appropriate test statistic. In Table 3, we compare the two models according to the three criteria. The LR test indicates that the superior fit of FLCM is statistically significant ($\chi_4^2 = 122, p = .001$). Also, both AIC and BIC, each of which imposes penalties for overparameterization, take on lower values for FLCM, which indicates that it is the preferred model. In general, AIC and BIC are used to evaluate the relative predictive (Atkinson 1980) and scientific (Rust et al. 1995; Woodroffe 1982) value, respectively, of competing models.

Overall Description of the Items

In Table 4, we present an overall description of the customer responses. Both FLCM and LCM indicate that roughly one-quarter of the customer responses falls into each of the categories service encounter failure (LCM: 27%; FLCM: 25%) and other (LCM: 26%; FLCM: 28%). But though LCM indicates that nearly half (47%) the items are in the basic service failure category, FLCM indicates that

only 19% are in this category. Furthermore, according to FLCM, 28% of the customer responses are fuzzy, and the Dirichlet parameter estimates indicate that they are primarily mixtures of the basic service failure and other categories: Among these fuzzy items, the average g_{ik} value is approximately 54% basic service failure and 46% other (see Table 4).

Overall Estimates of Fuzziness

The fuzziness measure \hat{M}_3 (see Equation 12) is .112, which is only one-ninth what it would be for a maximally fuzzy data set. This is because even the fuzzy items are primarily mixtures of only two categories. Nevertheless, Monte Carlo simulations indicate that the true fuzziness level is greater than .10, with greater than 99% confidence (each of 100,000 replications was sampled from the asymptotic multivariate normal distribution of the maximum likelihood estimations of the mixture proportion and Dirichlet parameters reported in Table 4). Thus, we reject the hypothesis that the data set is crisp at the .01 level. Finally, although \hat{M}_3 indicates a relatively low level of fuzziness, an estimated 28% (standard error: 1.6%) of the items are fuzzy (Table 4) and this presumably accounts for the superior overall performance of FLCM (Table 3).

Judges’ Classification Probabilities

In Table 5, we report the LCM (outside parentheses) and the FLCM (inside parentheses) estimates of the classification probabilities of each judge for the crisp items in the data set.⁵ For each model, the estimated probabilities of correct classification are reported on the diagonals of each judge’s summary, and the estimated probabilities of misclassification are on the off-diagonals. A comparison of these estimates reveals interesting differences. The LCM results suggest that Judges 1, 3, and 4 have less than a 70% probability of correctly identifying items from the “other” category and perhaps need to be retrained. In contrast, FLCM indicates that each judge correctly identifies items in this category more than 80% of the time.

⁴This Gauss program is available from the second author.

⁵For reasons of space and clarity, we do not report the standard errors. They are available on request.

Table 3
COMPARISON OF LCM AND FLCM MODELS

	LCM	FLCM	Remarks
Number of items (I)	314	314	
Number of parameters (NP) 1	26	30	$NP_{lcm} = J \times K \times (K - 1) + K - 1$ $NP_{flcm} = J \times K \times (K - 1) + 2K$
Log-likelihood (LL)	-1252	-1191	
LR test $-2 \times (LL_{lcm} - LL_{flcm}) \sim \chi_4^2$		$\chi_4^2 = 122$	FLCM has significant incremental value ($p < .001$).
AIC $-2 \times LL + 2 \times NP$	2556	2442	FLCM is preferred according to the AIC criterion.
BIC	2654	2554	FLCM is preferred according to

Table 4
DESCRIPTION OF THE ITEMS ACCORDING TO BOTH MODELS^a

Category	LCM		FLCM		
	Proportion of Items in Each Category	Proportion of Items in Each Category	Dirichlet Parameter Value	Average g_{ik} Values ^b	Average g_{ik} Values Among Fuzzy Items ^c
Basic service failure	.47 (.060)	.19 (.027)	30 (3.2)	.29	.54
Service encounter failure	.27 (.054)	.25 (.00038)	.04 (.00085)	.18	.00
Other	.26 (.074)	.28 (.011)	26 (.43)	.33	.46
Fuzzy items	0 ^d	.28 (.016)	—	—	—

^aAll values in this table are maximum likelihood estimates. Values in parentheses are heteroscedastic-consistent standard error estimates (White 1982).

^bFollowing the notation of Equations 3 and 5 for FLCM, the average g_{ik} value is

$$E(g_{ik}) = (1 - q) \frac{\alpha_k}{\sum_{k=1}^K \alpha_k} + qp_k$$

^cAverage g_{ik} values among mixtures are given by the expression in footnote b when $q = 0$.

^dLCM assumes that there are no fuzzy items.

There are other important differences between the LCM and FLCM estimates. For example, the LCM results for Judge 1 suggest that misclassifications between Categories 1 and 2 are infrequent (only 7.7% of Category 1 items are misclassified into Category 2, whereas 4.6% of Category 2 items are misclassified into Category 1). However, FLCM reports that Judge 1 has a distinct bias and never misclassifies Category 2 items into Category 1 but misclassifies nearly 20% of crisp Category 1 items into Category 2. Also, FLCM indicates that Judge 2 has a similar misclassification bias, whereas LCM indicates that Judge 2 misclassifies Category 1 items into Category 2 and items from Category 2 into Category 1 with the same relatively high frequency of 19%. For Judges 3 and 4, the LCM results suggest a bias toward Category 1 (because their respective misclassification probabilities show that they classify only 10% and 13% of Category 1 items into Category 2, and yet they classify 26% and 33% of items belonging to Category 2 into Category 1). In contrast, the FLCM results indicate that Judges 3 and 4 misclassify approximately the same proportion of crisp Category 1 items into Category 2 as they do Category 2 items into Category 1. Furthermore, according to FLCM, Judges 3 and 4 are the least accurate overall because of their tendency to misclassify items from the first two categories, and among all judges, these are the only cases in which correct classification probabilities fall below 70%.

DISCUSSION AND CONCLUSION

In marketing, qualitative data are common in the descriptions of complex market stimuli or phenomena and often are found in the form of open-ended responses to survey questions and in behavioral experiments. Consequently, categories that help summarize and communicate this information are integral to theory development efforts. Our

model enables researchers to assess both the classification accuracy of the judges (or instruments) employed and the quality of the categorization framework in terms of its ability to provide crisp classifications. Using this information, researchers may seek to either modify the categorization scheme or retrain the judges in an appropriate manner. Our model is termed the FLCM, because it extends the LCM by allowing for the classification of fuzzy items. Because FLCM subsumes LCM, it is applicable in all cases in which LCM is used and is especially appropriate when the presence of fuzzy items is suspected. Moreover, FLCM provides an estimate of the proportion of fuzzy items and a characterization of the mixture distribution they come from, as well as all the information that is typically available from LCM. Columns 2 through 4 of Table 4 (of the empirical example) summarize the additional information that is provided by FLCM compared with LCM. Also, FLCM can accommodate binary categories that partition an underlying continuum into extremes (e.g., customers are classified as satisfied or dissatisfied). In this type of application, the fuzziness results from not all items being at the extremes (e.g., satisfied, dissatisfied), and the GoM value (g_{ik}) indicates where each customer falls along the underlying continuum on the basis of customers' open-ended responses and judges' classifications. There are many business applications in which FLCM can be used to measure the accuracy of decision-making instruments for making classifications. Rust and Cooil (1994), for example, argue that hiring sales personnel is a case of judgment-based classification, because managers (judges) evaluate applicants (items) and classify them according to whether they are a good match with the firm. In this context, researchers may be interested in evaluating the managers' abilities to spot the right kind of applicant. Other business

Table 5
 CLASSIFICATION PROBABILITY ESTIMATES FOR KEAVENEY'S (1995) DATA RETURNED BY LCM (OUTSIDE PARENTHESES) AND
 FLCM (INSIDE PARENTHESES)

<i>Judge 1</i>			
	<i>Category 1: Basic Service Failure</i>	<i>Category 2: Service Encounter Failure</i>	<i>Category 3: Other</i>
Category 1: basic service failure	.813 (.779)	.077 (.195)	.110 (.025)
Category 2: service encounter failure	.046 (.000)	.864 (.900)	.090 (.100)
Category 3: other	.152 (.038)	.194 (.059)	.654 (.903)
<i>Judge 2</i>			
	<i>Category 1: Basic Service Failure</i>	<i>Category 2: Service Encounter Failure</i>	<i>Category 3: Other</i>
Category 1: basic service failure	.658 (.705)	.192 (.255)	.149 (.039)
Category 2: service encounter failure	.193 (.095)	.778 (.758)	.029 (.148)
Category 3: other	.098 (.000)	.022 (.067)	.880 (.933)
<i>Judge 3</i>			
	<i>Category 1: Basic Service Failure</i>	<i>Category 2: Service Encounter Failure</i>	<i>Category 3: Other</i>
Category 1: basic service failure	.658 (.693)	.100 (.173)	.242 (.135)
Category 2: service encounter failure	.263 (.210)	.513 (.505)	.224 (.285)
Category 3: other	.272 (.054)	.184 (.112)	.544 (.835)
<i>Judge 4</i>			
	<i>Category 1: Basic Service Failure</i>	<i>Category 2: Service Encounter Failure</i>	<i>Category 3: Other</i>
Category 1: basic service failure	.572 (.528)	.134 (.276)	.294 (.196)
Category 2: service encounter failure	.333 (.242)	.481 (.487)	.186 (.271)
Category 3: other	.173 (.000)	.216 (.076)	.611 (.924)

Notes: Each entry is the probability that the particular judge classifies a crisp item belonging to category k (indexed by row) into category k^* (indexed by column).

scenarios include the classification of business prospects/clients by advertising managers in terms of client–agency fit and the identification of potential defaulters by credit managers.

The maximum likelihood estimations provided by FLCM are only as good as the data used to estimate them, and to that extent, it is important that researchers sample the content domain in an adequate and appropriate manner. However, unlike percentage agreement, which is not model-based, FLCM still provides an assessment of the accuracy of the estimates (through estimated standard errors) and the quality of fit (through LR tests and model selection criteria).

The empirical example illustrates that qualitative marketing data can be fuzzy. This suggests that it is important to model data fuzziness in marketing applications, which requires specific analytical tools. Whenever researchers seek to gain additional insight, they must pay the price of increased model complexity. Such models may also require

richer data sets. Thus, although LCM provides biased estimates of the classification probabilities when the data include items that are mixtures, these estimates must suffice when the sample size is small or when only a few judges are available (see the requirements in Equations 7 and 8). Nevertheless, in our empirical example, AIC and BIC both indicate that FLCM is preferred to LCM. In this case, the estimated degree of fuzziness is low, but FLCM still provides a more detailed description of the items and a dramatically different assessment of judgment accuracy.

The FLCM proposed in this article extends the LCM to situations in which data may be either fuzzy or crisp. Our model can improve the diagnostic value of judgment-based classifications by distinguishing between misclassification error and inherent item fuzziness. Given the frequently fuzzy nature of marketing data, new methods of modeling and measuring fuzzy data promise to be of great value to marketing researchers.

APPENDIX: FLCM LIKELIHOOD

In FLCM, judges' classifications of an item are treated as independent, conditioned on the fuzziness of the item. Because the GoM of items is unknown or latent, this uncertainty about the GoM (g_{ik}) of the items is factored in by integrating over the arbitrary distribution F from which the vectors (g_{i1}, \dots, g_{iK}) are sampled:

$$(A1) L_{FLCM} = \prod_{i=1}^I \int_{\sum_k g_{ik}=1} \prod_{j=1}^J \prod_{k^*=1}^K \left(\sum_{k=1}^K g_{ik} \lambda_{jkk^*} \right)^{y_{ijk^*}} dF.$$

The notation is as follows:

- i indexes the items in the data set, $1 \leq i \leq I$;
- j indexes the judges, $1 \leq j \leq J$;
- k indexes the categories, $1 \leq k \leq K$;
- k^* indexes the category into which item i is classified by judge j ;
- λ_{jkk^*} is the (conditional) probability that judge j will categorize an item that belongs exclusively to category k into category k^* ;
- y_{ijk^*} is a 0–1 indicator variable that indexes the category k^* into which judge j classifies item i ;
- g_{ik} is the GoM of item i in category k ;
- $\int_{\sum_k g_{ik}=1}$ is the integration over the g_{ik} simplex, because the GoM values for a particular item i must sum to 1;
- dF is the probability density function of the distribution from which the g_{ik} are sampled (see Equation 3).

Equation A1 can be simplified further by the introduction of the additional notation of κ_{ij} , which denotes the actual category to which judge j assigns item i such that y_{ijk^*} is zero except when $k^* = \kappa_{ij}$. Accordingly,

$$(A2) L_{FLCM} = \prod_{i=1}^I \int_{\sum_k g_{ik}=1} \prod_{j=1}^J \sum_{k=1}^K g_{ik} \lambda_{jk\kappa_{ij}} dF.$$

Expanding Equation A2,

$$(A3) L_{FLCM} = \prod_{i=1}^I \sum_{k_1=1}^K \sum_{k_2=1}^K \dots \sum_{k_J=1}^K \left\{ \left(\prod_{j=1}^J \lambda_{jk_1 \kappa_{ij}} \right) \left[\int_{\sum_k g_{ik}=1} \left(\prod_{j=1}^J g_{ik_j} \right) dF \right] \right\},$$

where judge j classifies item i as if it were from category k_j . Because $\int_{\sum_k g_{ik}=1} (\prod_{j=1}^J g_{ik_j}) dF$ is nothing but the expectation of $\prod_{j=1}^J g_{ik_j}$, Equation A3 becomes Equation 6a. Readers may note that the expectation in Equation 6a yields moments of the J th order, and consequently, a larger number of judges (J) results in a better resolution of the distribution from which the g_{ik} random vectors are realized. The number of parameters in this model can be reduced if parametric assumptions are made about the nature of the g_{ik} distribu-

tion. We are proposing the discrete-Dirichlet mixture (Equation 3). As shown in Equation 6b, this then provides a representation of all the J th-order moments in terms of the $2 \times K$ parameters of the discrete-Dirichlet distribution. For example, if there are two categories and two judges ($K = 2, J = 2$) and we assume that the g_{ik} come from the discrete-Dirichlet mixture (Equation 3), Equation 6b becomes

$$(A4) L_{FLCM} = \prod_{i=1}^I \left\{ \lambda_{11\kappa_{i1}} \lambda_{21\kappa_{i2}} \mu_{20} + \lambda_{11\kappa_{i1}} \lambda_{22\kappa_{i2}} \mu_{11} + \lambda_{12\kappa_{i1}} \lambda_{21\kappa_{i2}} \mu_{11} + \lambda_{12\kappa_{i1}} \lambda_{22\kappa_{i2}} \mu_{02} \right\},$$

with

$$(A5) \mu_{rs} = E(g_{i1}^r g_{i2}^s) = (1 - q) \frac{D(\alpha_1, \alpha_2)}{D(\alpha_1 + r, \alpha_2 + s)} + q [p_1 I(r = 2) + p_2 I(s = 2)],$$

where r and s are nonnegative integers, $r + s = 2$, and we are using the notation of Equation 6b.

Finally, if the g_{ik} are restricted to the values in the set $\{0,1\}$, Equation 6a reduces to the LCM likelihood

$$(A6) L_{LCM} = \prod_{i=1}^I \sum_{k=1}^K p_k \prod_{j=1}^J \prod_{k=1}^K (\lambda_{jkk^*})^{y_{ijk^*}},$$

where p_k is simply the proportion of crisp items that belong to category k .

REFERENCES

Akaike, H. (1974), "A New Look At the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19 (6), 716–23.

Aptech Systems Inc. (1995), *GAUSS Version 3.2 System and Graphics Manual*. Kent, WA: Aptech Systems.

Atkinson, A.C. (1980), "A Note on the Generalized Information Criterion for Choice of a Model," *Biometrika*, 67 (2), 413–18.

Bishop, Yvonne M., Stephen E. Feinberg, and Paul W. Holland (1975), *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Bitner, Mary Jo, Bernard M. Booms, and Mary Stanfield Tetreault (1990), "The Service Encounter: Diagnosing Favorable and Unfavorable Incidents," *Journal of Marketing*, 54 (January), 71–84.

Cooil, Bruce and Roland T. Rust (1995), "General Estimators for the Reliability of Qualitative Data," *Psychometrika*, 60 (2), 199–220.

Dawid, A.P. and A.M. Skene (1979), "Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm," *Applied Statistics*, 28 (1), 20–28.

DeSarbo, Wayne S., Venkatram Ramaswamy, and Rabikar Chatterjee (1995), "Analyzing Constant Sum Multiple Criterion Data: A Segment-Level Approach," *Journal of Marketing Research*, 32 (2), 222–32.

Dillon, W.R. and N. Mulani (1984), "A Probabilistic Latent Class Model for Assessing Inter-judge Reliability," *Multivariate Behavioral Research*, 19 (4), 438–58.

Gross, Barbara L. and Jagdish N. Sheth (1989), "Time-Oriented Advertising: A Content Analysis of United States Magazine Advertising, 1890–1988," *Journal of Marketing*, 53 (October), 76–84.

Hastak, Manoj and Jerry C. Olson (1989), "Assessing the Role of Brand-Related Cognitive Responses as Mediators of

- Communication Effects of Cognitive Structure," *Journal of Consumer Research*, 15 (March), 444–57.
- Hruschka, H. (1986), "Market Definition and Segmentation Using Fuzzy Clustering Methods," *International Journal of Research in Marketing*, 3 (2), 117–34.
- Hughes, Marie Adele and Dennis E. Garrett (1990), "Intercoder Reliability Estimation Approaches in Marketing: A Generalizability Theory Framework for Quantitative Data," *Journal of Marketing Research*, 27 (May), 185–95.
- Hunt, Shelby D. (1991), *Modern Marketing Theory: Critical Issues in the Philosophy of Marketing Science*. Cincinnati: South-Western Publishing Co.
- Jain, Dipak, Frank M. Bass, and Yu-Min Chen (1990), "Estimation of Latent Class Models with Heterogeneous Choice Probabilities: An Application to Market Structuring," *Journal of Marketing Research*, 27 (February), 94–101.
- Keaveney, Susan (1995), "Customer Switching Behavior in Service Industries: An Exploratory Study," *Journal of Marketing*, 59 (April), 71–82.
- Manton, Kenneth G., Max A. Woodbury, and H. Dennis Tolley (1994), *Statistical Applications Using Fuzzy Sets*. New York: John Wiley & Sons.
- Perreault, William D., Jr., and Laurence E. Leigh (1989), "Reliability of Nominal Data Based on Qualitative Judgments," *Journal of Marketing Research*, 26 (May), 135–48.
- Richins, Marsha L. (1994), "Valuing Things: The Public and Private Meanings of Possessions," *Journal of Consumer Research*, 21 (December), 504–21.
- Rust, R.T. and Bruce Cooil (1994), "Reliability Measures for Qualitative Data: Theory and Implications," *Journal of Marketing Research*, 31 (February), 1–14.
- and Robert P. Leone (1984), "The Mixed-Media Dirichlet Multinomial Distribution: A Model for Evaluating Television-Magazine Advertising Schedules," *Journal of Marketing Research*, 21 (1), 89–100.
- , Duncan Simester, Roderick J. Brodie, and V. Nilikant (1995), "Model Selection Criteria: An Investigation of Relative Accuracy, Posterior Probabilities, and Combinations of Criteria," *Management Science*, 41 (2), 322–33.
- Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6 (2), 461–64.
- Spiggle, Susan (1994), "Analysis and Interpretation of Qualitative Data in Consumer Research," *Journal of Consumer Research*, 21 (3), 491–503.
- Wedel, Michel and Jan Benedict Steenkamp (1989), "A Fuzzy Clusterwise Regression Approach to Benefit Segmentation," *International Journal of Research in Marketing*, 6 (4), 241–58.
- and ——— (1991), "A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation," *Journal of Marketing Research*, 28 (November), 385–96.
- White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.
- Woodroffe, M. (1982), "On Model Selection and the Arc Sine Laws," *Annals of Statistics*, 10 (4), 1182–94.