

# ON THE ASSIGNMENT OF INDIVIDUALS TO LATENT CLASSES

*Leo A. Goodman\**

*Consider an  $m$ -way cross-classification table (for  $m = 3, 4, \dots$ ) of  $m$  dichotomous variables that describes (1) the  $2^m$  possible response patterns to a set of  $m$  questions (where the response to each question is binary), and (2) the number of individuals whose responses to the  $m$  questions can be described by a particular response pattern, for each of the  $2^m$  possible response patterns. Consider the situation where the data in the cross-classification table are analyzed using a particular latent class model having  $T$  latent classes (for  $T = 2, 3, \dots$ ), and where this model fits the data well. With this latent class model, it is possible to estimate, for an individual who has a particular response pattern, what is the conditional probability that this individual is in a particular latent class, for each of the  $T$  latent classes. In this article, the following question is considered: For an individual who has a particular response pattern, can we use the corresponding estimated conditional probabilities to assign this individual to one of the  $T$  latent classes? Two different assignment procedures are considered here, and for each of these procedures, two different criteria are introduced to help assess when the assignment procedure is satisfactory and when it is not. In addition, we describe here the particular framework and context in which the two assignment procedures, and the two criteria, are considered. For illustrative purposes, the latent class analysis of a*

For helpful comments, the author is indebted to Mike Hout, Yu Xie, and an anonymous reviewer. Direct correspondence to Leo Goodman, Department of Sociology and Department of Statistics, Barrows Hall 410, University of California, Berkeley, CA 94720-1980; e-mail: lgoodman@berkeley.edu

\*University of California, Berkeley

*classic set of data, a four-way cross-classification of some survey data, obtained in a two-wave panel study, is discussed; and the two different criteria introduced herein are applied in this analysis to each of the two assignment procedures.*

## 1. INTRODUCTION

For expository purposes, we consider first the two-class latent class model (with, say, latent classes 1 and 2) applied to the observed data in a four-way  $2 \times 2 \times 2 \times 2$  cross-classification of four dichotomous variables (say, variables  $A$ ,  $B$ ,  $C$ , and  $D$ ). There are 16 possible response patterns in the four-way cross-classification table; and we let  $f_{ijkl}$  denote the observed number (i.e., the observed frequency) of individuals whose response pattern is  $(i, j, k, l)$ , with response  $i$  (for  $i = 1, 2$ ) on variable  $A$ , response  $j$  (for  $j = 1, 2$ ) on variable  $B$ , response  $k$  (for  $k = 1, 2$ ) on variable  $C$ , and response  $l$  (for  $l = 1, 2$ ) on variable  $D$ . When the latent class model is applied to the cross-classification table, the observed data in the table are used to estimate the corresponding expected frequency  $F_{ijkl}$  under the latent class model; and, for an individual whose response pattern is  $(i, j, k, l)$ , we can also estimate the conditional probabilities,  $\pi_{ijkl1}^{ABCD\bar{X}}$  and  $\pi_{ijkl2}^{ABCD\bar{X}}$ , of this individual being in latent class 1 and latent class 2, respectively. When the latent class model fits well the observed data in the cross-classification table (i.e., when the estimated  $F_{ijkl}$  are sufficiently close to the corresponding observed  $f_{ijkl}$ , for  $i = 1, 2, j = 1, 2, k = 1, 2$ , and  $l = 1, 2$ ), the following question can arise: For an individual whose response pattern is  $(i, j, k, l)$ , how can we use the corresponding estimated conditional probabilities,  $\pi_{ijkl1}^{ABCD\bar{X}}$  and  $\pi_{ijkl2}^{ABCD\bar{X}}$ , to assign this individual to one of the two latent classes?

The above question was considered very briefly in a short two-page subsection of a long (81-page) article (Goodman 1974a) that was focused on many other matters pertaining to latent class analysis. More needs to be said on the assignment of individuals to latent classes.

In the short subsection of the long article cited above, an assignment procedure was described, but no criterion was introduced there to help assess when the assignment procedure is satisfactory and when it is not. In the present article, an additional assignment procedure is introduced, and two different criteria are also introduced that can be applied to each of the two assignment procedures in order to help assess each of them.

When the two-class latent class model is applied to the observed data in a four-way  $2 \times 2 \times 2 \times 2$  cross-classification of four dichotomous variables, the 16 possible response patterns correspond to the 16 cells in the cross-classification table. We can think of these cells as the 16 possible observed classes into which each of the individual respondents can be classified in accordance with his or her response pattern. The two-class latent class analysis seeks to determine whether the observed classification of the individual respondents into the 16 observed classes can be described in terms of just two latent classes (say, latent classes 1 and 2), where the proportion of respondents in latent class 1 is  $\pi_1^X$  and the proportion in latent class 2 is  $\pi_2^X$  (with  $\pi_1^X + \pi_2^X = 1$ ); and where these two proportions in the latent class model, and some other characteristics of the respondents in each of the two latent classes in the model, are estimated using the observed data. The estimates are then used to estimate the corresponding expected frequency  $F_{ijkl}$  under the latent class model. When the estimated  $F_{ijkl}$  are sufficiently close to the corresponding observed  $f_{ijkl}$  (for  $i = 1, 2; j = 1, 2; k = 1, 2; l = 1, 2$ ), we can view the latent class model with its two latent classes as possibly an underlying description, and a basic explanation, of the subject under study—that is, the observed classification of the respondents in the 16 observed classes. In this case, the latent class model with its two latent classes can be viewed as possibly more fundamental than the observed classification of the respondents in the 16 observed classes. This article considers two possible procedures for assigning the respondents in the 16 observed classes to the two latent classes.

In these introductory comments we have, for expository purposes, considered the two-class latent class model applied to the four-way cross-classification table of four dichotomous variables. These introductory comments can be directly generalized to the case where the  $T$ -class latent class model (for  $T = 2, 3, \dots$ ) is applied to the  $m$ -way cross-classification table of  $m$  dichotomous variables (for  $m = 3, 4, \dots$ ).

The two assignment procedures that will be considered here for assigning individuals to latent classes, and the two criteria that will be introduced here for assessing the assignment procedures, will be viewed as a contribution to the methodology of the subject of “classification.” In the present context, a classification study would determine whether a set of individuals, who have responded to a set of questions, can validly be described in terms of a small number of classes (or clusters) of similar

individuals, whose responses to the questions are similar in some sense. (For views of the subject of classification applied in many other contexts, see, e.g., Gordon 1999.)

The classification issues considered in the present article are viewed here in the situation in which the  $T$ -class latent-class model (for  $T = 2, 3, \dots$ ) is applied to the  $m$ -way cross-classification of  $m$  dichotomous variables (for  $m = 3, 4, \dots$ ). The same kinds of classification issues could also be considered in the situation in which other finite mixture models are applied in the analysis of various kinds of data; but this would be beyond the scope of the present article. (For views of other mixture models, and some extensions of latent class models, see, e.g., MacLachlan and Peel 2000; Magidson and Vermunt 2001.)

## 2. AN EXAMPLE

We shall consider in this section a latent class model applied by Goodman (1974a, 1974b), in analyzing data presented earlier by Coleman (1964). Table 1 presents the observed cross-classification of individuals in interviews at two successive points in time, with respect to two dichotomous variables. Here we have data from a two-wave panel study of individuals interviewed with respect to two questions in the first wave and the same two questions in the second wave. The data in Table 1 are presented as a two-way  $4 \times 4$  table, with the four row categories corresponding to the four possible responses in the first wave, and the four column categories corresponding to the four possible responses in the second wave. The data in the 16 cells of the  $4 \times 4$  table can also be viewed as data in the corresponding 16 cells of a  $2 \times 2 \times 2 \times 2$  table  $\{A, B, C, D\}$ , with  $A$  and  $B$  corresponding to the first two questions in the first wave, and  $C$  and  $D$  corresponding to the same two questions in the second wave.

When the two-class latent class model was applied to the data, the model did not fit the data. (The goodness-of-fit chi-square value was 251.17 on 6 degrees of freedom.) But when a special four-class latent-class model was applied, the model fit the data very well indeed. (The corresponding chi-square value was 1.28 on 4 degrees of freedom.) The estimated conditional probabilities,  $\pi_{ijkl}^{ABCD\bar{X}}$  (for  $t = 1, 2, 3, 4$ ), of being in the corresponding latent classes (say, latent classes 1, 2, 3, 4) for those individuals whose response pattern is  $(i, j, k, l)$  in the four-way  $2 \times 2 \times$

TABLE 1  
Observed Cross-Classification of 3398 Schoolboys in Interviews at Two Successive  
Points in Time, with Respect to Two Dichotomous Variables\*

		Second Interview			
Membership	Attitude	1	1	2	2
		1	2	1	2
First Interview					
Membership	Attitude				
1	1	458	140	110	49
1	2	171	182	56	87
2	1	184	75	531	281
2	2	85	97	338	554

Source: Coleman (1964).  
*Note:* With respect to self-perceived membership in the “leading crowd,” being in it is denoted 1 and being out of it is denoted 2; with respect to attitude concerning the “leading crowd,” a favorable attitude is denoted 1 and an unfavorable attitude is denoted 2. Membership and attitude will be denoted by the letters *A* and *B*, respectively, in the first interview; and by the letters *C* and *D*, respectively, in the second interview.  
\*Variables = (1) self-perceived membership in the “leading crowd,” and (2) favorableness of attitude concerning the “leading crowd.”

$2 \times 2$  table  $\{A, B, C, D\}$  are presented in Table 2. (This table differs from the corresponding table in Goodman [1974a] in that four decimal places are included here; the table in the earlier article included two decimal places. The four decimal places will be needed in the calculations that will be presented later herein.)

For readers who would like to know more about the special four-class latent class model that was applied to the data (Table 1), we include here the following brief description: The four latent classes (say, latent classes 1, 2, 3, 4) can be viewed as the latent classes of a latent variable *X*, and they can also be viewed as the four classes in a  $2 \times 2$  table describing the joint distribution of two dichotomous latent variables (say, latent variable *Y* and *Z*). With respect to the observed data in Table 1, the observed variables *A* and *C* correspond to the first question (on membership) in the first and second interview, respectively; and the observed variables *B* and *D* correspond to the second question (on attitude) in the first and second interview, respectively. Under the special four-class latent class model considered here, the observed variables *A* and *C* are viewed as indicators of the dichotomous latent variable *Y*, and the observed variables *B* and *D* are viewed as indicators of the dichotomous

latent variable  $Z$ . In other words, under this model, variable  $Y$  is the intrinsic latent variable on membership, and variable  $Z$  is the intrinsic latent variable on attitude; and the proportion of individuals estimated to be in each of the four latent classes in the  $2 \times 2$  table describes the estimated joint distribution of the intrinsic latent variables on membership and attitude.

When the four-class latent class model considered here was applied to the data in Table 1, the proportion of individuals in each of the four latent classes was estimated under the latent class model, and the following results were obtained:  $\pi_1^X = .2720$ ,  $\pi_2^X = .1284$ ,  $\pi_3^X = .2315$ , and  $\pi_4^X = .3680$ , for latent classes 1, 2, 3, and 4, respectively. (These proportions differ from the corresponding proportions in Goodman (1974a, 1974b) in that four decimal places are included here; whereas the Goodman [1974a] article included two decimal places, and the Goodman [1974b] article included three decimal places. The four decimal places included here will be needed in the calculations that will be presented later herein.)

With respect to the above four proportions estimated under the latent class model, when these proportions are viewed as proportions in the  $2 \times 2$  table describing the estimated joint distribution of the intrinsic latent variables  $Y$  and  $Z$ , we have  $\pi_1^X = \pi_{11}^{YZ}$ ,  $\pi_2^X = \pi_{12}^{YZ}$ ,  $\pi_3^X = \pi_{21}^{YZ}$ ,  $\pi_4^X = \pi_{22}^{YZ}$ . In other words, the latent variable  $X$  here denotes the joint latent variable  $(Y, Z)$ , and the four latent classes of latent variable  $X$  correspond to the four levels with respect to this joint latent variable—that is, (1, 1), (1, 2), (2, 1), and (2, 2), respectively. Using the corresponding four proportions in the preceding paragraph, the corresponding odds-ratio in the  $2 \times 2$  table is  $(\pi_1^X \pi_4^X) / (\pi_2^X \pi_3^X) = 3.37$ . So we see that there is a strong positive relationship between the intrinsic latent variable on membership and the intrinsic latent variable on attitude.

With the latent class analysis of the data in Table 1 presented in the present section, we shall next consider the two different assignment procedures for assigning individuals to latent classes, and then the two criteria for assessing these assignment procedures. The procedures and criteria will make use of the information in Table 2, and the second criterion will also make use of the information presented earlier in this section on the proportion of individuals (i.e.,  $\pi_t^X$ , for  $t = 1, 2, 3, 4$ ) estimated to be in each of the four latent classes under the latent class model.

TABLE 2  
Estimated Conditional Probability  $\pi_{ijklt}^{ABCD\bar{X}}$  That an Individual Will Be in Latent Class  $t$  (for  $t = 1, 2, 3, 4$ ), Given That the individual's Response Pattern Is  $(i, j, k, l)$  on Variables  $(A, B, C, D)$

Variable				Observed Frequency	Estimated Conditional Probability for Latent Class $t$			
A	B	C	D		t = 1	t = 2	t = 3	t = 4
1	1	1	1	458	.9355	.0529	.0097	.0019
1	1	1	2	140	.5937	.3866	.0062	.0136
1	1	2	1	110	.3864	.0219	.4970	.0947
1	1	2	2	49	.1737	.1131	.2234	.4899
1	2	1	1	171	.5959	.3844	.0062	.0135
1	2	1	2	182	.1150	.8539	.0012	.0299
1	2	2	1	56	.1747	.1127	.2247	.4880
1	2	2	2	87	.0239	.1773	.0307	.7681
2	1	1	1	184	.7349	.0416	.1878	.0358
2	1	1	2	75	.4053	.2640	.1036	.2271
2	1	2	1	531	.0259	.0015	.8170	.1556
2	1	2	2	281	.0098	.0064	.3082	.6757
2	2	1	1	85	.4072	.2627	.1041	.2260
2	2	1	2	97	.0664	.4927	.0170	.4240
2	2	2	1	338	.0098	.0063	.3101	.6737
2	2	2	2	554	.0012	.0090	.0381	.9518

*Note:* The conditional probability is estimated under a special four-class latent class model applied to the four-way cross-classification (Table 1).

Before closing this section, we note that more about the special four-class latent class model considered here is included in Section A.1 of the Appendix. We also refer interested readers to Goodman (1974a, 1974b) for additional material on this special latent class model and on other latent class models. Reference is also made here to, e.g., Clogg and Goodman (1984, 1985), Goodman (1987, 2002, 2007), and Hagenaars and McCutcheon (2002).

### 3. TWO ASSIGNMENT PROCEDURES

In this section, we first consider the assignment procedure described briefly in Goodman (1974a). This procedure is closely related to the measure of association  $\lambda$ , which is based on optimal prediction, presented in Goodman and Kruskal (1954). We then introduce the second assign-

ment procedure. This is closely related to the measure of association  $\tau$ , which is based on proportional prediction, also presented in Goodman and Kruskal (1954).

For an individual whose response pattern is  $(i, j, k, l)$ , let us now consider the first assignment procedure. This procedure assigns the individual to a corresponding modal latent class—that is, to a particular latent class for which the corresponding estimated conditional probability  $\pi_{ijkl}^{ABCDX}$  (considering  $t = 1, 2, 3, 4$ ) is the largest. If this assignment procedure is applied to all of the individuals whose response pattern is  $(i, j, k, l)$ , using the estimated conditional probabilities in Table 2, we see, for example, that the 458 individuals whose response pattern is  $(1, 1, 1, 1)$  would be assigned to latent class 1; and the 554 individuals whose response pattern is  $(2, 2, 2, 2)$  would be assigned to latent class 4. This simple assignment procedure minimizes the number of incorrect assignments. Applying this procedure to the individuals whose response pattern is  $(i, j, k, l)$  for each of the 16 response patterns, we find that the number of individuals assigned to the latent classes 1, 2, 3, and 4 is 1113, 279, 641, and 1365, respectively, and the corresponding proportions assigned to the latent classes are as follows:  $p_1^X = .3275$ ,  $p_2^X = .0821$ ,  $p_3^X = .1886$ , and  $p_4^X = .4017$ . When these four proportions assigned to the latent classes are compared with the corresponding proportions presented in the preceding section—that is, the proportion of individuals in each of the four latent classes ( $\pi_1^X, \pi_2^X, \pi_3^X, \pi_4^X$ ) estimated under the latent class model—we see that the proportions obtained with the first assignment procedure considered in this section (i.e., the assignment procedure that minimizes the number of incorrect assignments) can be very different from the estimated proportions obtained under the latent class model.

Under the four-class latent class model considered here, we noted in the preceding section that the particular ratio  $(\pi_1^X \pi_4^X) / (\pi_2^X \pi_3^X)$  was of special interest. Using the estimated proportions  $\pi_i^X$  (for  $t = 1, 2, 3, 4$ ) in the preceding section, the corresponding estimated ratio was 3.37; whereas, using the corresponding proportions obtained when individuals are assigned to latent classes by applying the assignment procedure considered above, the corresponding ratio,  $(p_1^X p_4^X) / (p_2^X p_3^X)$ , is 8.49.

We shall now consider a different procedure for assigning individuals to latent classes. This assignment procedure is designed in a way so that the expected proportion of individuals assigned to each of the



four latent classes can approximate the corresponding proportion,  $\pi_t^X$ , for  $t = 1, 2, 3, 4$ , estimated under the latent class model.

With this assignment procedure, each of the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$  would be assigned at random to one of the latent classes 1, 2, 3, 4, using the corresponding estimated probability distribution,  $\pi_{ijklt}^{ABCD\bar{X}}$ , for  $t = 1, 2, 3, 4$ , to make the random assignments. For example, with this assignment procedure, using the estimated conditional probabilities in Table 2, each of the 458 individuals whose response pattern is (1, 1, 1, 1) would be assigned at random to one of the four latent classes, using the probability distribution .9355, .0529, .0097, .0019 to make the random assignments to latent classes 1, 2, 3, 4, respectively; each of the 554 individuals whose response pattern is (2, 2, 2, 2) would be assigned at random to one of the four latent classes, using the probability distribution .0012, .0090, .0381, .9518 to make the random assignments to latent classes 1, 2, 3, 4, respectively. So we see here that each of the 458 individuals sharing a given response pattern also share a given assignment probability distribution; each of the 554 individuals sharing a very different given response pattern also share a very different given assignment probability distribution.

The assignment procedure described in the preceding paragraph is such that, for the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$ , the expected proportion of individuals assigned to latent classes 1, 2, 3, 4 will be equal to the corresponding estimated probabilities  $\pi_{ijklt}^{ABCD\bar{X}}$ , for  $t = 1, 2, 3, 4$ . We explain in the next section why the expected proportion of individuals assigned to each of the four latent classes using this assignment procedure, applied to the individuals sharing a given response pattern for each of the 16 possible response patterns, can approximate the corresponding proportion  $\pi_t^X$ , for  $t = 1, 2, 3, 4$ , estimated under the latent class model.

The first assignment procedure considered in this section (and earlier in Goodman [1974a]), used a *modal* latent class based on the estimated probability distribution of the four latent classes corresponding to each of the 16 response patterns; we shall call this assignment procedure *M*. (For related material on this procedure, see Vermunt and Magidson [2005].) The second assignment procedure uses *random* assignments based on the estimated probability distribution of the latent classes corresponding to each of the 16 response patterns; we shall call this assignment procedure *R*.

#### 4. TWO CRITERIA FOR ASSESSING ASSIGNMENT PROCEDURES $M$ AND $R$

With respect to the first assignment procedure considered in the preceding section (i.e., assignment procedure  $M$ , the procedure that minimizes the number of incorrect assignments), here is a simple way to estimate the proportion of incorrect assignments obtained when this assignment procedure is used:

With  $f_{ijkl}$  denoting the observed number of individuals whose response pattern is  $(i, j, k, l)$ , we let  $n$  denote the total number of observed individuals in the  $2 \times 2 \times 2 \times 2$  table  $\{A, B, C, D\}$ . Now we let  $E_M$  denote the proportion of incorrect assignments obtained with assignment procedure  $M$ . The following formula can be used to estimate  $E_M$ :

$$E_M = 1 - \left\{ \left[ \sum_{i,j,k,l} (f_{ijkl} \cdot \mu_{ijklm}^{ABCD}) \right] / n \right\}, \quad (1)$$

where

$$\mu_{ijklm}^{ABCD} = \max_t \pi_{ijklt}^{ABCD\bar{X}}.$$

Using the corresponding information in Table 2, we find with formula (1) that the estimate of  $E_M$  is .24. (More on formula [1] will be included in Section A.2 of the Appendix.)

Strictly speaking, a small improvement in formula (1) can be obtained by reducing slightly each product  $(f_{ijkl} \cdot \mu_{ijklm}^{ABCD})$  in the formula, by using the corresponding whole number without the decimal that follows the whole number (i.e., by replacing the product with the largest integer that is less than or equal to the product). This replacement of the product is equal to the number of (whole) individuals estimated to be in the modal latent class corresponding to response pattern  $(i, j, k, l)$ . With this small improvement, again using the corresponding information in Table 2, we find that the estimate of  $E_M$  becomes .25.

Now let us consider the second assignment procedure—assignment procedure  $R$ —described in the preceding section. With this procedure, the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$  are assigned at random to latent classes 1, 2, 3, 4 using the corresponding estimated probability distribution,  $\pi_{ijklt}^{ABCD\bar{X}}$ , for  $t = 1, 2, 3, 4$ , to make

the random assignments. The estimate of the expected proportion  $E_R$  of incorrect assignments obtained with assignment procedure  $R$  can be described simply as follows:

$$E_R = 1 - \left\{ \left[ \sum_{i,j,k,l} (f_{ijkl} \cdot \omega_{ijkl}^{ABCD}) \right] / n \right\}, \quad (2)$$

where

$$\omega_{ijkl}^{ABCD} = \sum_{t=1}^4 \left( \pi_{ijklt}^{ABCD\bar{X}} \right)^2.$$

Using the corresponding information in Table 2, we find with formula (2) that the estimate of  $E_R$  is .34. (More on formula [2] will be included in Section A.3 of the Appendix.)

For those readers who may be interested in analyzing in more detail the proportion  $E_M$  of incorrect assignments obtained with assignment procedure  $M$ , and the expected proportion  $E_R$  of incorrect assignments obtained with assignment procedure  $R$ , a method of partitioning these error rates will be presented in the next section.

We have been considering in this section the estimate of the proportion (or the expected proportion) of incorrect assignments obtained with an assignment procedure as our first criterion for assessing the assignment procedure. We noted earlier that assignment procedure  $M$  minimizes the proportion of incorrect assignments. (More on this in Section A.2 of the Appendix.) On the other hand, we also found that the proportion of individuals assigned to each of the latent classes by assignment procedure  $M$  can be very different from the corresponding estimated proportions obtained under the latent class model.

Our second criterion for assessing an assignment procedure is the comparison of the proportion of individuals assigned to each of the latent classes by the assignment procedure with the corresponding estimated proportions obtained under the latent class model. We shall now see why the expected proportion of individuals assigned to each of the four latent classes using assignment procedure  $R$  can approximate the corresponding proportion,  $\pi_t^X$  for  $t = 1, 2, 3, 4$ , estimated under the latent class model.

To explain the above phenomenon, we make use of the following formula:

$$\sum_{i,j,k,l} \left( F_{ijkl} \cdot \pi_{ijkl}^{ABCD\bar{X}} \right) / n = \pi_t^X, \quad (3)$$

for  $t = 1, 2, 3, 4$ , where  $F_{ijkl}$  is the expected number of individuals whose response pattern is  $(i, j, k, l)$  under the latent class model, and where  $n$  is the total number of individuals in the cross-classification table. (More on this formula will also be included in Section A.3 in the Appendix.) When the latent class model fits well the observed data in the cross-classification table (i.e., when the estimated  $F_{ijkl}$  are sufficiently close to the corresponding observed  $f_{ijkl}$ ), we can approximate the quantity in parentheses on the left side in formula (3) by  $(f_{ijkl} \cdot \pi_{ijkl}^{ABCD\bar{X}})$ , and formula (3) will then be approximated by

$$\sum_{i,j,k,l} \left( f_{ijkl} \cdot \pi_{ijkl}^{ABCD\bar{X}} \right) / n = \tilde{\pi}_t^X, \quad (4)$$

where  $\tilde{\pi}_t^X$  denotes an approximation to  $\pi_t^X$ , for  $t = 1, 2, 3, 4$ . Since assignment procedure  $R$  assigns each of the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$  at random to latent classes 1, 2, 3, 4, using the estimated probability distribution  $\pi_{ijkl}^{ABCD\bar{X}}$  (for  $t = 1, 2, 3, 4$ ), we see that the expected number of these individuals who are assigned to latent class  $t$  is simply  $(f_{ijkl} \cdot \pi_{ijkl}^{ABCD\bar{X}})$ , for  $t = 1, 2, 3, 4$ . When assignment procedure  $R$  is applied to each of the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$ , for each of the 16 possible response patterns, the expected proportion of these  $n$  individuals who are assigned to latent class  $t$  is given by formula (4), which is an approximation to  $\pi_t^X$ , for  $t = 1, 2, 3, 4$ .

Using the corresponding information in Table 2, we find with formula (4) that, despite the fact that the  $f_{ijkl}$  differ somewhat from the corresponding  $F_{ijkl}$ , the  $\tilde{\pi}_t^X$  (for  $t = 1, 2, 3, 4$ ) turned out to be equal to the corresponding  $\pi_t^X$  to four decimal places. (For greater accuracy, the information actually used here with formula (4) used eight decimal places for the  $\pi_{ijkl}^{ABCD\bar{X}}$  rather than the corresponding four decimal places presented in Table 2.)

Assignment procedure  $R$  replaces the  $F_{ijkl}$  in formula (3) with the corresponding observed number  $f_{ijkl}$ , and it replaces the  $\pi_{ijkl}^{ABCD\bar{X}}$  with a random assignment procedure in which the expected proportion of the  $f_{ijkl}$  individuals who are assigned to latent class  $t$  is equal to  $\pi_{ijkl}^{ABCD\bar{X}}$ , for  $t = 1, 2, 3, 4$ . Since assignment procedure  $R$  is a random assignment

procedure, the actual proportion of the  $f_{ijkl}$  individuals assigned to each latent class will differ in a random way from the corresponding expected proportion. Also, the number of individuals  $f_{ijkl}$  whose response pattern is  $(i, j, k, l)$  will differ from the corresponding  $F_{ijkl}$  in a random way, given that the latent class model applied to the data in Table 1 is a correct model for describing these data. The actual proportion of individuals assigned to each latent class using the assignment procedure  $R$  should approximate the corresponding  $\pi_t^X$ , for  $t = 1, 2, 3, 4$ , estimated under the latent class model; but this approximation will depend on the magnitude of the differences between the  $f_{ijkl}$  and the corresponding  $F_{ijkl}$ , and on the size of the  $f_{ijkl}$ , and on the conditional probabilities  $\pi_{ijkl}^{ABCD\bar{X}}$ , for  $t = 1, 2, 3, 4$ .

## 5. THE PARTITIONING OF ERROR RATES

We first consider the proportion  $E_M$  of incorrect assignments using assignment procedure  $M$ . For the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$ , using assignment procedure  $M$ , the proportion of these individuals who are assigned correctly is  $\mu_{ijklm}^{ABCD}$ , where  $m$  is a modal latent class, and where

$$\mu_{ijklm}^{ABCD} = \max_t \pi_{ijklt}^{ABCD\bar{X}} = \pi_{ijklm}^{ABCD\bar{X}}.$$

The proportion of these individuals who are assigned incorrectly is  $\pi_{ijklt}^{ABCD\bar{X}}$  for all  $t \neq m$ . And so, for the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$ , we find that  $f_{ijkl} \cdot \pi_{ijklm}^{ABCD\bar{X}}$  of them are assigned correctly, and  $f_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}}$  are assigned incorrectly for each latent class  $t \neq m$ .

Now let us partition the 16 response patterns into four separate sets; with  $\Delta m$  denoting the set of those response patterns that have a latent class  $m$  as their modal class, for  $m = 1, 2, 3, 4$ . For example, we see from the information in Table 2 that the set  $\Delta 1$  consists of the response patterns  $(1, 1, 1, 1)$ ,  $(1, 1, 1, 2)$ ,  $(1, 2, 1, 1)$ ,  $(2, 1, 1, 1)$ ,  $(2, 1, 1, 2)$ ,  $(2, 2, 1, 1)$ ; and the set  $\Delta 4$  consists of the response patterns  $(1, 1, 2, 2)$ ,  $(1, 2, 2, 1)$ ,  $(1, 2, 2, 2)$ ,  $(2, 1, 2, 2)$ ,  $(2, 2, 2, 1)$ ,  $(2, 2, 2, 2)$ . Let  $f_{\Delta m}$  denote the total number of individuals whose response pattern is one of the response patterns in the set  $\Delta m$ , for  $m = 1, 2, 3, 4$ . (From the information in

Table 2, we see that  $f_{\Delta 1} = 1113$ ,  $f_{\Delta 2} = 279$ ,  $f_{\Delta 3} = 641$ ,  $f_{\Delta 4} = 1365$ , as was noted earlier in Section 3.) Using assignment procedure  $M$ , the proportion of the  $n$  individuals in the cross-classification table who are assigned to latent class  $m$  is simply  $f_{\Delta m}/n$ . And the proportion of the  $f_{\Delta m}$  individuals who are assigned correctly is  $\sum_{\Delta m} (f_{ijkl} \cdot \pi_{ijklm}^{ABCD\bar{X}}) / f_{\Delta m}$ , for  $m = 1, 2, 3, 4$ , where the summation is made over all response patterns in the set  $\Delta m$ . Similarly, the proportion of the  $f_{\Delta m}$  individuals who are assigned incorrectly is  $\sum_{\Delta m} (f_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}}) / f_{\Delta m}$ , for all  $t \neq m$ , for  $m = 1, 2, 3, 4$ . Using the information in Table 2, we now present in Table 3 a partitioning of the proportion of correct assignments and the proportion of incorrect assignments using assignment procedure  $M$ . (For related material on this procedure, see Vermunt and Magidson [2005].)

We noted in the preceding section, using the information in Table 2, that the proportion of incorrect assignments obtained with assignment procedure  $M$  was .24, and so the proportion of correct assignments was .76. With Table 3, we see (1) how the proportion of correct assignments varies depending on which assignment latent class is considered, and (2) how the proportion of incorrect assignments varies depending on which assignment latent class is considered and on which latent class the individuals are in.

Now let us consider assignment procedure  $R$ . For the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$ , using assignment procedure  $R$ , we find that the expected number of these individuals who are assigned to latent class  $t$  is  $f_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}}$ , for  $t = 1, 2, 3, 4$ ; and the expected

TABLE 3  
Partition of the Proportion of Correct Assignments and the Proportion of  
Incorrect Assignments Using Assignment Procedure  $M$

Latent Class	Proportion Assigned to Latent Class*	Proportion of Those Who Were in Column Latent Class Among Those Who Were Assigned to Row Latent Class*			
		1	2	3	4
1	.3275	.7311	.1742	.0517	.0430
2	.0821	.0981	.7283	.0067	.1669
3	.1886	.0878	.0050	.7621	.1452
4	.4017	.0199	.0265	.1749	.7788

\* See corresponding comments in this section.

number of these individuals who are assigned correctly to this latent class is  $f_{ijkl} \cdot (\pi_{ijklt}^{ABCD\bar{X}})^2$ . The following formula is the expected number who are assigned incorrectly:

$$f_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}} \cdot \pi_{ijkls}^{ABCD\bar{X}},$$

the expected number who are assigned to latent class  $t$  and are in latent class  $s$ , where  $s \neq t$ .

Now using assignment procedure  $R$ , the following formula is the expected proportion of the  $n$  individuals in the cross-classification table who are assigned to latent class  $t$ :

$$\sum_{i,j,k,l} \left( f_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}} \right) / n. \quad (5)$$

Formula (6) is the proportion of the expected number of individuals correctly assigned to latent class  $t$  among the expected number assigned (correctly or incorrectly) to that latent class:

$$\sum_{i,j,k,l} \left[ f_{ijkl} \cdot (\pi_{ijklt}^{ABCD\bar{X}})^2 \right] / \sum_{i,j,k,l} \left( f_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}} \right). \quad (6)$$

Formula (7) is the proportion of the expected number of individuals in latent class  $s$  among the expected number assigned (correctly or incorrectly) to latent class  $t$ , where  $s \neq t$ :

$$\sum_{i,j,k,l} \left( f_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}} \cdot \pi_{ijkls}^{ABCD\bar{X}} \right) / \sum_{i,j,k,l} \left( f_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}} \right). \quad (7)$$

Using the information in Table 2, we now present in Table 4 a partitioning of the expected proportion of correct assignments and the expected proportion of incorrect assignments using assignment procedure  $R$ . The entries in the column of the expected proportions assigned to each latent class were obtained using formula (5), for  $t = 1, 2, 3, 4$ ; and the entries on the main diagonal of the part of Table 4 that can be viewed as a  $4 \times 4$  table were obtained using formula (6), for  $t = 1, 2, 3, 4$ . The entries that are not on the main diagonal of the  $4 \times 4$  table (say, in row  $t$  and column  $s$ , for  $s \neq t$ ), were obtained using formula (7), for all  $s \neq t$ , in row  $t = 1, 2, 3, 4$ .

TABLE 4  
Partition of the Expected Proportion of Correct Assignments and the Expected  
Proportion of Incorrect Assignments Using Assignment Procedure  $R$

Latent Class	Expected Proportion Assigned to Latent Class*	Proportion of the Expected in Column Latent Class Among the Expected Assigned to Row Latent Class†			
		1	2	3	4
1	.2720	.7135	.1527	.0822	.0515
2	.1284	.3235	.5025	.0324	.1415
3	.2315	.0966	.0180	.5787	.3066
4	.3680	.0381	.0494	.1929	.7196

\* See formula (5) and corresponding comments in this section.

† See formulas (6) and (7), and corresponding comments in this section.

We noted in the preceding section, using the information in Table 2, that the expected proportion of incorrect assignments obtained with assignment procedure  $R$  was .34, and so the expected proportion of correct assignments was .66. With Table 4, we see how the expected proportion of correct assignments varies, and how the expected proportion of incorrect assignments varies. In addition, with the entries in Table 4 in the column of the expected proportion of individuals assigned to each latent class using assignment procedure  $R$ , we also see how very different these expected proportions are from the entries in Table 3 in the corresponding column of the proportion assigned to each latent class using assignment procedure  $M$ . In addition, we note here, as we did earlier in the preceding section, that the expected proportion assigned to each latent class using assignment procedure  $R$  turned out, in this example, to be the same, to four decimal places, as the proportion of individuals in each of the latent classes estimated under the latent class model.

## 6. CONCLUSION

We have considered here the following two procedures for assigning individuals to latent classes: (1) assignment procedure  $M$  minimizes the number of incorrect assignments, and (2) assignment procedure  $R$  assigns individuals to the different latent classes in a way so that, for



those individuals whose response pattern is  $(i, j, k, l)$ , the expected proportion of these individuals assigned to each latent class will be equal to the corresponding conditional probability  $\pi_{ijklt}^{ABCD\bar{X}}$  estimated to be in that latent class under the latent class model. Formulas (1) and (2) in this paper can be used to estimate the proportion of incorrect assignments using assignment procedure  $M$  and the expected proportion of incorrect assignments using assignment procedure  $R$ , respectively. In addition, with assignment procedure  $M$ , we can determine the proportion of individuals assigned to each of the latent classes; with assignment procedure  $R$ , we can determine the expected proportion of individuals assigned to each of the latent classes. And these proportions and expected proportions can be compared with the corresponding proportion estimated to be in that latent class under the latent class model. The two criteria have been introduced here to help assess when assignment procedures  $M$  and/or  $R$  are satisfactory and when they are not.

## APPENDIX

### A.1. More on the Special Four-Class Latent Class Model

We shall begin this section by first considering the conditional probabilities  $\pi_{1t}^{\bar{A}X}, \pi_{1t}^{\bar{B}X}, \pi_{1t}^{\bar{C}X}, \pi_{1t}^{\bar{D}X}$  (for  $t = 1, 2, 3, 4$ ), which are parameters in the special four-class latent-class model applied in Section 2. The  $\pi_{1t}^{\bar{A}X}$  denotes the conditional probability of a response 1 on variable A, given that the respondent is in latent class  $t$  (for  $t = 1, 2, 3, 4$ ); and so  $1 - \pi_{1t}^{\bar{A}X}$  is the conditional probability of a response 2 on variable A, given that the respondent is in latent class  $t$ . The other three conditional probabilities,  $\pi_{1t}^{\bar{B}X}, \pi_{1t}^{\bar{C}X}, \pi_{1t}^{\bar{D}X}$ , are defined similarly. Table 5 presents each of these four conditional probabilities (for  $t = 1, 2, 3, 4$ ) estimated under the special four-class latent class model.

Under this latent class model, for each respondent in latent class  $t$ , the respondent's responses on variables  $A, B, C$ , and  $D$  are assumed to be mutually independent of each other; so the conditional probability of each of the 16 possible response patterns  $(i, j, k, l)$ , for the respondents in latent class  $t$ , can thus be expressed directly in terms of the conditional probabilities described in the preceding paragraph; and the corresponding estimate of the conditional probability of each of the 16

TABLE 5  
Conditional Probabilities,  $\pi_{1t}^{\bar{A}X}$ ,  $\pi_{1t}^{\bar{B}X}$ ,  $\pi_{1t}^{\bar{C}X}$ ,  $\pi_{1t}^{\bar{D}X}$ , Estimated Under the Special  
Four-Class Latent Class Model

Latent Class $t$	$\pi_{1t}^{\bar{A}X}$	$\pi_{1t}^{\bar{B}X}$	$\pi_{1t}^{\bar{C}X}$	$\pi_{1t}^{\bar{D}X}$
1	.7543	.8056	.9098	.8325
2	.7543	.2665	.9098	.3015
3	.1112	.8056	.0755	.8325
4	.1112	.2665	.0755	.3015

*Note:* The  $\pi_{1t}^{\bar{A}X}$  is the conditional probability of a response 1 on variable A, given that the respondent is in latent class  $t$  (for  $t = 1, 2, 3, 4$ ). The  $\pi_{1t}^{\bar{B}X}$ ,  $\pi_{1t}^{\bar{C}X}$ ,  $\pi_{1t}^{\bar{D}X}$  are defined similarly.

possible response patterns  $(i, j, k, l)$ , for the respondents in latent class  $t$ , can be calculated directly using the estimated conditional probabilities in Table 5.

Now let  $\pi_{ijklt}^{\bar{A}\bar{B}\bar{C}\bar{D}X}$  denote the conditional probability of a response pattern  $(i, j, k, l)$ , given that the respondent is in latent class  $t$ ; and so  $\pi_t^X \cdot \pi_{ijklt}^{\bar{A}\bar{B}\bar{C}\bar{D}X}$  is the expected proportion of individuals in the four-way cross-classification table  $\{A, B, C, D\}$  who are in latent class  $t$  and whose response pattern is  $(i, j, k, l)$ . The estimate of the proportion  $\pi_t^X$  of individuals in the four-way table  $\{A, B, C, D\}$  who are in latent class  $t$  was given earlier in Section 2 (for  $t = 1, 2, 3, 4$ ), and the conditional probability  $\pi_{ijklt}^{\bar{A}\bar{B}\bar{C}\bar{D}X}$  can be estimated for each of the 16 response patterns  $(i, j, k, l)$  using the estimated conditional probabilities in Table 5, as we noted in the preceding paragraph. For each response pattern  $(i, j, k, l)$ , the sum of the  $\pi_t^X \cdot \pi_{ijklt}^{\bar{A}\bar{B}\bar{C}\bar{D}X}$ , summed from  $t = 1$  to  $t = 4$ , is the expected proportion of individuals in the four-way table  $\{A, B, C, D\}$  whose response pattern is  $(i, j, k, l)$ . And the expected frequency of response pattern  $(i, j, k, l)$  under the latent class model is obtained by multiplying each of the expected proportions by the total number of individuals in the table. The corresponding estimated expected frequencies for the 16 possible response patterns are presented in Table 6.

The expected proportion  $\pi_t^X \cdot \pi_{ijklt}^{\bar{A}\bar{B}\bar{C}\bar{D}X}$  considered in the preceding paragraph can also be viewed as the probability that an individual in the four-way table  $\{A, B, C, D\}$  is in latent class  $t$  and the individual's response pattern is  $(i, j, k, l)$ . Thus, for each individual whose response pattern is  $(i, j, k, l)$ , the conditional probability that the individual is in latent class  $t$  can be calculated by dividing  $\pi_t^X \cdot \pi_{ijklt}^{\bar{A}\bar{B}\bar{C}\bar{D}X}$  by the sum

TABLE 6  
Comparison of Observed Frequencies in Table 1 with the Expected Frequencies  
Estimated Under the Special Four-Class Latent Class Model

Variable				Observed Frequency	Estimated Expected Frequency
A	B	C	D		
1	1	1	1	458	454.7695
1	1	1	2	140	144.2282
1	1	2	1	110	109.1153
1	1	2	2	49	48.8634
1	2	1	1	171	172.3030
1	2	1	2	182	179.6693
1	2	2	1	56	58.2627
1	2	2	2	87	85.7585
2	1	1	1	184	188.5942
2	1	1	2	75	68.8156
2	1	2	1	531	530.5210
2	1	2	2	281	283.0927
2	2	1	1	85	82.1401
2	2	1	2	97	101.4501
2	2	2	1	338	337.2942
2	2	2	2	554	553.0921

of the  $\pi_t^X \cdot \pi_{ijklt}^{ABCDX}$ , summed from  $t = 1$  to  $t = 4$ . The corresponding estimated conditional probability that an individual is in latent class  $t$  (for  $t = 1, 2, 3, 4$ ), given that the individual's response pattern is  $(i, j, k, l)$ , was presented earlier in Table 2.

#### A.2. On Assignment Procedure $M$

We first consider formula (1) in Section 4. For the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$  on variables  $A, B, C, D$ , we find that, when assignment procedure  $M$  is applied, the number of individuals assigned correctly is  $f_{ijkl} \cdot \mu_{ijklm}^{ABCD}$ , where  $\mu_{ijklm}^{ABCD}$  is defined just below formula (1), and  $(f_{ijkl} \cdot \mu_{ijklm}^{ABCD})/n$  is the proportion of the total number of the  $n$  individuals in the cross-classification table whose response pattern is  $(i, j, k, l)$  and who have been assigned correctly by assignment procedure  $M$ . Thus, the proportion of the total number of individuals in the cross-classification who have been assigned correctly is the quantity in braces on the right side of formula (1). And the proportion assigned incorrectly is one minus the proportion assigned correctly.

We noted earlier herein that assignment procedure  $M$  minimizes the proportion of incorrect assignment. If we first consider only assignment procedures in which all of the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$  are assigned to the same latent class, then it is clear from formula (1) that assignment procedure  $M$  minimizes the proportion of incorrect assignments, since  $\mu_{ijklm}^{ABCD}$  is the largest of the  $\mu_{ijklt}^{ABCD\bar{X}}$  for  $t = 1, 2, 3, 4$ . (If there are two or more latent classes that are modal, then assignment of the  $f_{ijkl}$  individuals to any one of the modal latent classes will still provide an assignment procedure that minimizes the proportion of incorrect assignments.)

We now consider an assignment procedure in which the proportion of the  $f_{ijkl}$  individuals assigned to latent class  $t$  is  $v_{ijklt}^{ABCD\bar{X}}$  for  $t = 1, 2, 3, 4$ , where  $\sum_t v_{ijklt}^{ABCD\bar{X}} = 1$ . The proportion of the  $f_{ijkl}$  assigned correctly will then be  $\sum_t (\pi_{ijklt}^{ABCD\bar{X}}) \cdot (v_{ijklt}^{ABCD\bar{X}})$ , and this quantity will be maximized when

$$v_{ijklt}^{ABCD\bar{X}} = \begin{cases} 1, & \text{for } t = m, \\ 0, & \text{for } t \neq m \end{cases}$$

where  $m$  is a modal latent class. (If there are two or more latent classes that are modal, then the assignment procedure will minimize the proportion of incorrect assignments when  $v_{ijklt}^{ABCD\bar{X}} = 0$  for each latent class  $t$  that is not modal, and the  $f_{ijkl}$  individuals are distributed among the latent classes that are modal in any manner.)

### A.3. On Assignment Procedure $R$

We now consider formula (2) in Section 4. For the  $f_{ijkl}$  individuals whose response pattern is  $(i, j, k, l)$  on variables  $A, B, C, D$ , we find that, when assignment procedure  $R$  is applied to these individuals, the expected number assigned correctly is  $f_{ijkl}[\sum_t (\pi_{ijklt}^{ABCD\bar{X}})^2]$ , and  $f_{ijkl}[\sum_t (\pi_{ijklt}^{ABCD\bar{X}})^2]/n$  is the expected proportion of the total number of the  $n$  individuals in the cross-classification table whose response pattern is  $(i, j, k, l)$  and who have been assigned correctly by assignment procedure  $R$ . Thus, the expected proportion of the total number of individuals in the cross-classification who have been assigned correctly is the quantity in braces on the right side of formula (2). And so the

expected proportion assigned incorrectly is one minus the expected proportion assigned correctly.

We now consider formula (3) in Section 4. This formula was obtained in the following way: We first note that the formula for the conditional probability  $\pi_{ijklt}^{ABCD\bar{X}}$  of being in latent class  $t$ , given that the response pattern was  $(i, j, k, l)$  on variables  $A, B, C, D$ , is the following:

$$\pi_{ijklt}^{ABCD\bar{X}} = (\pi_{ijklt}^{ABCDX}) / (\pi_{ijkl}^{ABCD}), \quad (\text{A-1})$$

where  $\pi_{ijkl}^{ABCD}$  is the probability that the response pattern will be  $(i, j, k, l)$ , and  $\pi_{ijklt}^{ABCDX}$  is the joint probability that the response pattern will be  $(i, j, k, l)$  and the latent class will be  $t$ . Also, the formula for the expected frequency  $F_{ijkl}$  of response pattern  $(i, j, k, l)$  on variables  $A, B, C, D$ , can be written as follows:

$$F_{ijkl} = n\pi_{ijkl}^{ABCD}. \quad (\text{A-2})$$

Now using formulas (A-1) and (A-2), we see that

$$(F_{ijkl} \cdot \pi_{ijklt}^{ABCD\bar{X}}) / n = \pi_{ijkl}^{ABCD} \pi_{ijklt}^{ABCD\bar{X}} = \pi_{ijklt}^{ABCDX}; \quad (\text{A-3})$$

and so, formula (3) in Section 4 can be rewritten as follows:

$$\sum_{i,j,k,l} \pi_{ijklt}^{ABCDX} = \pi_t^X.$$

This formula states that the joint probability that the response pattern will be  $(i, j, k, l)$  on variables  $A, B, C, D$ , and the latent class will be  $t$ , when summed over variables  $A, B, C$ , and  $D$ , is simply equal to the probability that the latent class will be  $t$ . In other words, the probability distribution that the latent class will be  $t$  (for  $t = 1, 2, 3, 4$ ) is simply equal to the marginal distribution obtained when the joint probability distribution that the response pattern will be  $(i, j, k, l)$  on variables  $A, B, C, D$ , and the latent class will be  $t$  (for  $t = 1, 2, 3, 4$ ), is summed over variables  $A, B, C$ , and  $D$ .

## REFERENCES

- Clogg, C. C., and L. A. Goodman. 1984. "Latent Structure Analysis of a Set of Multidimensional Tables." *Journal of the American Statistical Association* 79:762–771.
- . 1985. "Simultaneous Latent Structure Analysis in Several Groups." Pp. 81–110 in *Sociological Methodology*, vol. 15, edited by Nancy B. Tuma. San Francisco, CA: Jossey-Bass.
- Coleman, J. S. 1964. *Introduction to Mathematical Sociology*. New York: Free Press.
- Goodman, L. A. 1974a. "The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. Part I—A Modified Latent Structure Approach." *American Journal of Sociology* 79:1179–259.
- . 1974b. "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215–31.
- . 1987. "New Methods for Analyzing the Intrinsic Character of Qualitative Variables Using Cross-Classified Data." *American Journal of Sociology* 93:529–583.
- . 2002. "Latent Class Analysis: The Empirical Study of Latent Types, Latent Variables, and Latent Structures." Pp.3–53 in *Applied Latent Class Analysis*, edited by J. A. Hagenaars and A. L. McCutcheon. Cambridge, England: Cambridge University Press.
- . 2007. "Statistical Magic and/or Statistical Serendipity: An Age of Progress in the Analysis of Categorical Data." *Annual Review of Sociology* 33:1–19.
- Goodman, L. A., and W. H. Kruskal. 1954. "Measures of Association for Cross Classifications." *Journal of the American Statistical Association* 49:732–84.
- Gordon, A. D. 1999. *Classification*, Second Edition. Boca Raton, FL: Chapman & Hall/CRC.
- Hagenaars, J. A., and A. L. McCutcheon, eds. 2002. *Applied Latent Class Analysis*. Cambridge, England: Cambridge University Press.
- MacLachlan, G. J., and D. Peel. 2000. *Finite Mixture Models*. New York: Wiley.
- Magidson, J., and J. K. Vermunt. 2001. "Latent Class Factor and Cluster Models, Bi-Plots and Related Graphical Displays." Pp. 223–64 in *Sociological Methodology*, vol.31, edited by Michael E. Sobel and Mark P. Becker. Boston, MA: Blackwell Publishing.
- Vermunt, J. K. and J. Magidson. 2005. *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont, MA: Statistical Innovations.