

A Simulation Study of the Difference Chi-Square Statistic for Comparing Latent Class Models Under Violation of Regularity Conditions

Judith A. Holt, Gallaudet University
George B. Macready, University of Maryland

This study explored the robustness of the likelihood ratio difference statistic to the violation of a regularity condition when used to assess differences in fit provided by pairs of latent class models. Under regularity conditions, the additive property of the likelihood ratio statistic can be used to assess the statistical difference between pairs of hierarchically related models (i.e., one model is a constrained form of the other). However, when one of the two models being compared is obtained by fixing parameters of the other model at boundary values (i.e., 0 or 1), a regularity condition is violated and the difference statistic is not necessarily distributed as χ^2 . The effects of three independent variables on the distribution of the difference statistics were studied for two generation models and a variety of subsuming models. Differential effects in terms of the direction and the extent of deviation were produced according to the types of model comparisons; these effects negate the application of a simple correction to the statistic to achieve a χ^2 distribution. Recommendations are made regarding how this statistic might reasonably be used under violation of the regularity condition. *Index terms: latent class model, likelihood ratio chi-square, mixture model, regularity conditions, tests of fit.*

Latent class modeling is a probabilistic modeling strategy that can be used for establishing the relation between a response pattern on a set of manifest variables and a latent categorical variable. This

probabilistic approach allows for the statistical assessment of model fit to observed data (Bergan, 1982; Clogg, 1977; Dayton & Macready, 1976; Duncan & Sloane, 1982; Goodman, 1974a, 1979; Macready & Dayton, 1977, 1980).

Dichotomous manifest variables, which are applicable to a wide range of content areas, are frequently used with latent class models. Equation 1 defines a general latent class model that may be used with dichotomously scored variables:

$$P(\mathbf{r}; \theta_1 \dots \theta_J, \alpha_{11} \dots \alpha_{IJ}) = \sum_{j=1}^J \left[\theta_j \prod_{i=1}^I (\alpha_{ij})^{r_i} (1 - \alpha_{ij})^{1-r_i} \right], \quad (1)$$

where \mathbf{r} = vector of $r_i = \{0, 1\}$ manifest responses for variables $i = 1, \dots, I$,
 θ_j = probability of membership in the j th latent class, and
 α_{ij} = probability of a level 1 (e.g., positive or correct) response to the i th manifest variable, given that the assessed element (e.g., respondent) is a member of the j th latent class.

This latent class model is based on the following assumptions. First, it is assumed that the latent classes are mutually exclusive and exhaustive. Second, it is assumed that local independence (Goodman, 1974a) is present within latent classes. This

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 13, No. 3, September 1989, pp. 221-231
© Copyright 1989 Applied Psychological Measurement Inc.
0146-6216/89/030221-11\$1.80

means that the manifest variables conditional on latent class membership are independent, thus

$$P(\mathbf{r}|\theta = j) = \prod_{i=1}^I (\alpha_{ij})^{r_i} (1 - \alpha_{ij})^{1-r_i} \quad (2)$$

Third, it is assumed that the response options to each manifest variable i are mutually exclusive and exhaustive. Fourth, it is assumed that the estimation of the model parameters conforms to certain regularity conditions. Six regularity conditions are required for certain repeated sampling distributions to apply to the maximum likelihood estimator and the likelihood ratio test statistic (Birch, 1964). Steiger, Shapiro, and Browne (1985, pp. 509–511) described all six regularity conditions. The condition relevant to this study is described below.

The likelihood function for the estimation of latent class model parameters is

$$\lambda = \prod_{n=1}^N P(\mathbf{r}_n; \theta_1 \dots \theta_J, \alpha_{11} \dots \alpha_{IJ}) \quad (3)$$

where n is the respondent index.

In maximum likelihood estimation, estimates of $\hat{\theta}_j$ and $\hat{\alpha}_{ij}$ are chosen that maximize λ . A fit statistic may then be derived to test the ratio of the hypothesized model to the observed responses. This likelihood ratio statistic is

$$L^2 = -2 \ln \lambda \quad (4)$$

with λ maximized with respect to θ and α . L^2 is asymptotically distributed as χ^2 (Rao, 1973) with degrees of freedom = $2I - I - (J - 1) - IJ$, where I is the number of dichotomous manifest variables and J is the number of latent classes.

It is possible for these degrees of freedom to have a negative value, but a model will not be identified in that situation. Positive degrees of freedom is a necessary (but not sufficient) condition for model identification (Goodman, 1974b).

In addition to the fit of a selected model under a null hypothesis, it is possible to test in the same manner the fit of one or more constrained versions of that selected model under a set of alternate hypotheses. The fit statistic for testing the likelihood ratio of a hypothesized constrained model to the observed responses is

$$L_c^2 = -2 \ln \lambda_c \quad (5)$$

with λ_c maximized with respect to θ and α , under constraints. For this constrained model, there is a gain in degrees of freedom equal to the number of non-redundant constraints placed on the parameters in the more general model in defining the subsumed model.

There are several ways that a model may be constrained. Three basic methods involve the imposition of linear, linear logistic, and other functional constraints (Dayton & Macready, 1988). In addition, recent developments in latent class modeling have incorporated outside variables within the model (Clogg & Goodman, 1984, 1985, 1986; Dayton & Macready, 1988; Formann, 1982, 1984, 1985; Macready & Dayton, 1986). However, in most applications of latent class models, linear constraints have been the most widely reported (Bergan, 1982; Clogg, 1981; Dayton & Macready, 1976; Macready & Dayton, 1980). Two special cases of linear constraints are frequently used: fixing parameters at user-specified values, and setting the values of two or more parameters equal to one another.

Although the value of the maximized likelihood for a given model is greater than or equal to that of any model it subsumes (i.e., a model that is obtained by placing one or more constraints on the parameters of the initial model), sometimes a more constrained model will provide fit approaching that obtained under the more general subsuming model. The constrained model also has an advantage of increased parsimony, providing a gain in degrees of freedom as a result of the reduced number of independent parameters to be estimated. The final selection of a preferred model should be based on a balance between fit and parsimony. Fit requires that a model is able to effectively explain the data, whereas parsimony requires that a model is as simple (i.e., constrained) as possible. Thus, a preferred model is one that is as simple as possible yet provides acceptable fit to the data.

The Likelihood Ratio for Model Comparison

The statistical fit provided by a given latent class model may be compared with other models that

are constrained forms of that model. The statistical difference in fit may be assessed for pairs of models with a subsuming relation. The test statistic used to assess such differences may be expressed in terms of Equations 4 and 5:

$$\begin{aligned} G^2 &= -2 \ln \left(\frac{\lambda_c}{\lambda} \right) \\ &= -2 \ln \lambda_c - (-2) \ln \lambda \\ &= L_c^2 - L^2 \quad , \end{aligned} \quad (6)$$

where the subscript *c* indicates that the statistic corresponds to the constrained model.

Under certain assumptions, G^2 is asymptotically distributed as χ^2 with degrees of freedom K , the number of non-redundant constraints placed on the parameters in the less constrained model used in defining the subsumed model.

Two basic assumptions underlie the use of G^2 . First, it is assumed that in the estimation of both models, the above-mentioned four basic assumptions underlying the latent class model have been met. Second, it is assumed that the more general of the two models (i.e., the model with fewer imposed constraints) provides a proper specification for the data.

Study Description

This study focused on statistical testing of hypotheses regarding the differences in fit provided by selected pairs of models for which the more constrained model was defined by placing one or more constraints on the parameters of the other model. Each of the model comparisons considered violates the first regularity condition, and the distribution of G^2 was examined under this violation. The regularity condition considered in this study requires that in defining the parameters of the constrained model in terms of the parameters of the subsuming model, constraints may not be imposed by fixing parameters at boundaries of the parameter space (i.e., 0 or 1). This ensures that the maximum likelihood estimates have an asymptotic normal distribution about their true values. Otherwise, G^2 may fail to be distributed as χ^2 .

There is evidence in the literature (Bergan, 1982;

Clogg & Goodman, 1984, 1985, 1986; Macready, 1982; Macready & Dayton, 1980) that researchers have implemented tests of the relative fit of latent class models while violating the above-mentioned regularity condition. This has been the case even though G^2 is not necessarily χ^2 distributed and thus should not formally be used for assessing relative fit (Aitkin, Anderson, & Hinde, 1981; Aitkin & Rubin, 1985; Binder, 1978; Bowman & Shenton, 1973; Everitt & Hand, 1981; McLachlan & Basford, 1987; Quinn, McLachlan, & Hjort, 1987; Titterton, Smith, & Makov, 1985). In some instances researchers have acknowledged the violation of this underlying assumption while explaining that a more effective method did not exist at the time for accomplishing the systematic comparison of models (Dayton & Macready, 1988). However, the implications of using this statistic in the selection of a preferred latent class model have not previously been studied.

The purpose of this empirical examination of the distribution of G^2 was to investigate the direction and magnitude of its departure from χ^2 for a variety of latent class model comparisons. This examination of the robustness of G^2 will likely prove useful to applied researchers in two basic ways. First, it will be helpful in deciding whether the use of this statistic is warranted under any circumstances in the assessment of relative model fit when the first regularity condition is violated. Second, if the statistic is used in the selection of a preferred model, the results of this study will provide researchers with a better understanding of the effects on the null sampling distribution and thus give them a better idea of how to moderate their conclusions regarding the relative preferability of compared models.

Previous researchers have conducted empirical simulation studies with small samples from mixtures of normal distributions under violation of the first regularity condition. Based on a post hoc examination of their data, a number of these researchers have suggested modifications for G^2 and/or the degrees of freedom to improve the conformity to a χ^2 distribution (Aitkin & Rubin, 1985; Bartlett, 1938, 1947; Box, 1949; Everitt, 1981; Hartigan, 1975; Hogg, 1956; Quinn et al., 1987;

Wolfe, 1970, 1971). However, attempts at cross-validation have not produced the same desired results of improved conformity.

The present study differs from previous studies in three ways. Mixtures of latent binomial distributions were examined, both large and small sample sizes were considered, and two additional independent variables were included.

Method

This study was based on mixtures of latent binomial distributions using data generated from five dichotomous manifest variables. The number of response patterns for those variables was $2^5 = 32$. A total of 16 datasets were generated, representing two generation models under eight data conditions resulting from all combinations of three independent variables at two levels each.

Two latent class models were specified for the data generation. The first model was a general unconstrained two-class model defined by Equation 1. The second model was a three-class constrained model with equality constraints imposed on the conditional probabilities. The values of the conditional probabilities for these generation models are listed in Tables 1 and 2.

The first criterion used in the selection of these models was their applicability to a wide range of research in education, psychology, and other social sciences. The second criterion was that they were as unconstrained as was feasible. An unconstrained

three-class model was originally considered for data generation, but was rejected. Although it provided sufficient degrees of freedom for identifiability with five dichotomous items, a comparable four-class unconstrained model that would have been appropriate for comparison did not.

The first two independent variables varied the levels of the two types of parameters used in defining latent class models. This included the magnitude of the disparity of α_{ij} values across latent classes and the proportion of elements from each latent class.

The first level of the disparity of α_{ij} , which was "small" disparity, placed the response probabilities near the center of the parameter space. The second level, which was "large" disparity, placed α_{ij} values nearer the boundaries of the space. The specific values of α_{ij} for the two-class generation model are designated in Table 1. This type of model has been proposed for use in the assessment of mastery (Macready & Dayton, 1980) as well as a wide variety of other uses (Bergan, 1982; Clogg, 1981). Note from Table 1 that there is no overlap in the ranges of α_{ij} values across latent classes, and there is equal variability of α_{ij} values within each latent class.

The specific values of α_{ij} for the three-class constrained generation model are listed in Table 2. The constraints imposed on the second latent class specified that $\alpha_{i2} = \alpha_{i3}$ for $i = 1, 2$ and $\alpha_{i2} = \alpha_{i1}$ for $i = 3, 4, 5$. Constraints of this type are applicable when the three latent classes represent progressive states of acquisition (i.e., non-acquisition,

Table 1
Conditional Probabilities for the Two-Class
Generation Model for Small and Large
Disparities of α_{ij} Values

Variable	Small Disparity		Large Disparity	
	Class 1	Class 2	Class 1	Class 2
1	$\alpha_{11} = .49$	$\alpha_{12} = .79$	$\alpha_{11} = .29$	$\alpha_{12} = .99$
2	$\alpha_{21} = .47$	$\alpha_{22} = .77$	$\alpha_{21} = .27$	$\alpha_{22} = .97$
3	$\alpha_{31} = .45$	$\alpha_{32} = .75$	$\alpha_{31} = .25$	$\alpha_{32} = .95$
4	$\alpha_{41} = .43$	$\alpha_{42} = .73$	$\alpha_{41} = .23$	$\alpha_{42} = .93$
5	$\alpha_{51} = .41$	$\alpha_{52} = .71$	$\alpha_{51} = .21$	$\alpha_{52} = .91$

Table 2
Conditional Probabilities for the Three-Class Generation Model
for Small and Large Disparities of α_{ij} Values

Variable	Small Disparity			Large Disparity		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
1	$\alpha_{11} = .49$	$\alpha_{12} = .79$	$\alpha_{13} = .79$	$\alpha_{11} = .29$	$\alpha_{12} = .99$	$\alpha_{13} = .99$
2	$\alpha_{21} = .47$	$\alpha_{22} = .77$	$\alpha_{23} = .77$	$\alpha_{21} = .27$	$\alpha_{22} = .97$	$\alpha_{23} = .97$
3	$\alpha_{31} = .45$	$\alpha_{32} = .45$	$\alpha_{33} = .75$	$\alpha_{31} = .25$	$\alpha_{32} = .25$	$\alpha_{33} = .95$
4	$\alpha_{41} = .43$	$\alpha_{42} = .43$	$\alpha_{43} = .73$	$\alpha_{41} = .23$	$\alpha_{42} = .23$	$\alpha_{43} = .93$
5	$\alpha_{51} = .41$	$\alpha_{52} = .41$	$\alpha_{53} = .71$	$\alpha_{51} = .21$	$\alpha_{52} = .21$	$\alpha_{53} = .91$

partial acquisition, and acquisition), as suggested by Dayton and Macready (1976).

The next manipulated variable considered in this study was the relative proportion of elements from each latent class. The first level of latent mixture was equal latent proportions, where $\theta_1 = \theta_2 = .5$ for the two-class generation model and $\theta_1 = \theta_2 = \theta_3 = .333$ for the three-class generation model. The second level was unequal latent class mixture, where $\theta_1 = .8$ and $\theta_2 = .2$ for the two-class generation model and $\theta_1 = .6$, $\theta_2 = .3$, and $\theta_3 = .1$ for the three-class generation model. The first level of this variable placed the parameters near the center of the parameter space; the second level placed them nearer the boundaries of that space.

The last independent variable was sample size, which has been recognized as an important factor affecting the distribution of the logarithm of the likelihood ratio. The first level of this variable was $N = 160$, and the second level was $N = 960$. The values selected were based in part on the results of Hayek (1978), who found that $N = 800$ with five dichotomous items was not sufficiently large for the likelihood ratio statistic L^2 to have a distribution approximating χ^2 .

The IMSL (1980) subroutine GGUBS was used to generate random numbers from a uniform distribution, ranging from 0 to 1. Sixteen seeds, one for each of the 16 data generation conditions, were used. The seeds were examined to verify that they produced uniform distributions (Sedlacek, 1983). The random numbers were then used to create the raw data for the study. These numbers were compared to each of 32 cumulative probability intervals to produce the observed frequencies for the item

response patterns. This was repeated for 1,000 replications for each of the 16 data conditions.

Parameter Estimation

Parameters were estimated for the two generation models and for a set of subsuming models. The two criteria used in the selection of these subsuming models were their applicability for research and their relation to the generation models.

Each of the comparisons of the generation models with their subsuming models violated the first regularity condition by placing one or more latent proportions at a boundary value (i.e., 0). The data generated from the two-class unconstrained model were used to estimate parameters for that model, as well as for an unconstrained three-class model, a constrained three-class model, and a constrained four-class model. Similarly, the data generated from the three-class constrained model were used to estimate parameters for that model, as well as for a constrained four-class model.

The constrained three-class models estimated under both levels of data generation were identical. The constrained four-class model was specified with the same constraints on the first three classes as those imposed for the constrained three-class model. In addition, constraints on the fourth latent class were $\alpha_{i4} = \alpha_{i1}$ for $i = 1, 2$ and $\alpha_{i4} = \alpha_{i3}$ for $i = 3, 4, 5$. This model corresponds to an acquisition model in which item subsets $\{2,3\}$ and $\{3,4,5\}$ are at the same latent level of acquisition, namely, "masters" or "non-masters" (Macready, 1982).

Maximum likelihood estimation was used to estimate the model parameters. Clogg's (1977)

FORTRAN program, which is based on the iterative proportional fitting algorithm (Goodman, 1979), was adapted and used as a subroutine for this purpose. All calculations were done in double precision to minimize rounding error.

Note that the Newton-Raphson procedure could alternatively have been used in parameter estimation. This is the algorithm used in Formann's (1984) general-purpose computer program. An advantage of this alternative approach is that convergence is generally reached in fewer iterations. However, the time in which each iteration is completed is usually somewhat longer.

Analysis

The distributions of G^2 statistics were analyzed according to their overall fit to χ^2 distributions with the appropriate degrees of freedom. Each distribution of 1,000 observed statistics was divided into 100 intervals, based on the P values of a central χ^2 distribution (i.e., the cell lower boundaries correspond to $P = .00, .01, \dots, .99$). Each of the intervals, therefore, had an expected cell frequency of 10. A Pearson χ^2 statistic with 99 degrees of freedom was calculated to assess the fit of the observed G^2 statistics to a central χ^2 distribution.

In addition to overall fit, each of the distributions was examined in terms of its tail weights (i.e., the observed proportion of cases falling at or above those values of G^2 for which the expected proportions are .20, .10, .05, and .01). The tail weights for the distributions were compared to the distribution means and to the overall fit statistics.

Results

An examination of the effects of sample size indicated that the data generated under the smaller sample size did not result in accurate parameter estimates for any of the generation models. In addition, the likelihood ratio statistic L^2 for testing the fit of the generation models to the observed data also showed substantial deviation from a χ^2 distribution. Thus it might be expected that both small sample size and violation of a regularity con-

dition may result in G^2 that provides particularly poor fit to χ^2 .

At the large sample size, parameter estimates were reasonably accurate for all of the generation models considered. Additionally, the fit statistic L^2 was found to approximate a χ^2 distribution for all generation models.

Although all parameter estimates for the generation models at the large sample size were good, the small disparity in conditional response probabilities and the equal mixture proportions produced slightly better estimates. (This is not surprising, given the more central location of those parameters within the parameter space.) However, the more accurate reproduction of the generated parameters did not necessarily result in a better fit for the distributions of G^2 . Indeed, it had the opposite effect in some cases.

The fit statistics provide support for the contention that none of the distributions of G^2 are distributed as χ^2 . The average values of the fit statistics with 99 degrees of freedom were: 358.6 ($N = 960$) and 300.4 ($N = 160$) for the comparison of the unconstrained two-class model with the unconstrained three-class model; 13,850.4 ($N = 960$) and 13,827.2 ($N = 160$) for the comparison of the unconstrained two-class model with the constrained four-class model; and 26,279.4 ($N = 960$) and 31,551.0 ($N = 160$) for the combination of the comparisons of the unconstrained two-class model with the constrained three-class model and the constrained three-class model with the constrained four-class model.

The first three moments (i.e., mean, variance, and skewness) were also examined for the distributions of G^2 statistics. These moments are reported in Tables 3 and 4. Table 3 contains the moment information related to G^2 (with $df = 6$) resulting from the comparison of the unconstrained two-class model with the unconstrained three-class model. The means of these distributions are all larger than their expected values, with comparable results observed across the two sample sizes. However, the direction of the deviation is inconsistent for the other two moments across the four considered designs.

Table 3
Mean, Variance, and Skewness of G^2 Distributions for
the Comparison of the Unconstrained Two-Class Generation
Model vs. Unconstrained Three-Class Model and the
Constrained Four-Class Model

Disparity	Mixture	Mean		Variance		Skewness	
		N=160	N=960	N=160	N=960	N=160	N=960
Unconstrained Three-Class Model With Theoretical Mean = 6.0, Variance = 12.0, and Skewness = 48.0							
Small	Equal	7.02	7.70	14.21	12.68	74.81	48.77
Large	Equal	7.74	7.75	13.85	12.65	56.73	42.54
Small	Unequal	6.62	7.38	8.98	11.60	31.74	36.35
Large	Unequal	7.87	7.74	23.81	11.45	42.58	32.53
Combined	Combined	7.31	7.64	12.46	12.10	51.46	40.05
Constrained Four-Class Model With Theoretical Mean = 2.0, Variance = 4.0, and Skewness = 16.0							
Small	Equal	.80	.76	2.07	1.77	8.83	6.65
Large	Equal	.80	1.00	2.18	2.93	11.48	15.50
Small	Unequal	.65	.75	1.00	2.06	2.24	10.00
Large	Unequal	.76	.88	1.99	2.24	9.04	12.58
Combined	Combined	.75	.85	1.81	2.25	7.90	11.18

The moment information related to G^2 (with $df = 2$) resulting from the comparison of the unconstrained two-class model with the constrained four-class model is also reported in Table 3. All three moments for those distributions are consis-

tently smaller than their expected values. Additionally, the observed discrepancies are somewhat larger for $N = 160$.

Table 4 contains the moment information related to G^2 (with $df = 1$) resulting from the comparison

Table 4
Mean, Variance, and Skewness for G^2 Distributions for Comparisons
of the Unconstrained Two- and Constrained Three-Class Generation
Models vs. Constrained Three-Class and Four-Class Subsuming Models,
with Theoretical Mean = 1.0, Variance = 2.0, and Skewness = 8.0

Disparity	Mixture	Mean		Variance		Skewness	
		N=160	N=960	N=160	N=960	N=160	N=960
Unconstrained Two-Class Generating Model vs. Constrained Three-Class Subsuming Model							
Small	Equal	.53	.48	1.24	1.14	4.08	3.98
Large	Equal	.51	.53	1.53	1.48	8.46	7.01
Small	Unequal	.40	.51	.61	1.39	1.40	7.04
Large	Unequal	.49	.47	1.47	1.33	7.82	9.35
Combined	Combined	.48	.50	1.21	1.34	5.44	6.84
Constrained Three-Class Generating Model vs. Constrained Four-Class Subsuming Model							
Small	Equal	.45	.51	.96	1.04	3.52	3.26
Large	Equal	.33	.35	.75	.74	2.84	2.63
Small	Unequal	.41	.59	1.37	1.29	12.91	4.49
Large	Unequal	.27	.36	.56	.87	1.96	3.61
Combined	Combined	.36	.45	.91	.99	5.31	3.50

of the unconstrained two-class model with the constrained three-class model and the comparison of the constrained three-class model with the constrained four-class model. All of the observed means are smaller than their expected values for those distributions. With two exceptions (both related to skewness), the observed values for the other two moments are also smaller than their expected values.

In addition to the assessment of overall fit, the tails of the distributions of G^2 were examined for the adequacy of the proportion of rejections at specified critical values. Tables 5 and 6 list the observed proportion of rejections for each of the distributions at the critical values of .01, .05, .10, and .20.

The comparison of the unconstrained two-class model with the unconstrained three-class model is reported in Table 5. This comparison resulted in observed tail weights that are heavier than expected (i.e., the observed probability is greater than that corresponding to the central χ^2 distribution), which is consistent with the larger observed means found for these distributions. The one exception to this finding corresponds to the small sample size with unequal mixture, where the proportion of rejections is more accurate than expected.

The remaining three groups of model comparisons involving a generation model and a subsuming model with equality constraints imposed on conditional probabilities are reported in Table 6. These

comparisons resulted in tail weights that are lighter than expected, which is consistent with their smaller observed mean values.

In terms of the magnitude of the absolute deviation in the observed proportion of rejection, it is interesting to note the relatively high similarity across all independent variables for the model comparisons represented in Table 6 (i.e., comparisons for which $df = 1$ or 2). Note further that the absolute magnitude of these discrepancies is inversely related to the extremity of the critical values. This is the same relation that was found for the comparisons considered in Table 5 (except that the direction of the discrepancies is reversed).

Discussion

The G^2 statistic is not particularly robust to violation under any of the conditions studied. Deviations from expected values were manifested in the overall distributions, as well as in the tails. Both the magnitude and the direction of the discrepancies are dependent on the nature of the latent class models being compared. However, the values of the parameters of the generation models and the sample size used in estimation have only a small influence on the observed discrepancies.

There are two possible factors that may provide at least a partial explanation for the disparate results obtained across various classes of compared models.

Table 5
 Proportion of Rejection for G^2 Distributions
 at Specified Values for the Comparison of an
 Unconstrained Two-Class Generation Model vs.
 Unconstrained Three-Class Model, with Small and
 Large Disparity of α_{ij} Values

Mixture	Critical Values	Small Disparity		Large Disparity	
		N=160	N=960	N=160	N=960
Equal	.01	.02	.02	.02	.02
Equal	.05	.08	.10	.10	.10
Equal	.10	.14	.18	.18	.19
Equal	.20	.26	.34	.36	.33
Unequal	.01	.00	.01	.02	.02
Unequal	.05	.04	.08	.10	.08
Unequal	.10	.09	.16	.20	.19
Unequal	.20	.22	.31	.36	.36

Table 6
 Proportion of Rejection for G^2 Distributions at Specified Values for Comparisons of an Unconstrained Two-Class Generation Model vs. Constrained Three-Class Model, Unconstrained Two-Class Generation Model vs. Constrained Four-Class Model, and Constrained Three-Class Generation Model vs. Constrained Four-Class Model, with Small and Large Disparity of α_{ij} Values

Mixture	Critical Values	Small Disparity		Large Disparity	
		N=160	N=960	N=160	N=960
Unconstrained Two-Class Generation Model vs. Constrained Three-Class Model					
Equal	.01	.00	.00	.01	.01
Equal	.05	.03	.03	.03	.03
Equal	.10	.06	.05	.06	.05
Equal	.20	.12	.10	.10	.11
Unequal	.01	.00	.01	.01	.00
Unequal	.05	.00	.03	.01	.02
Unequal	.10	.03	.05	.05	.05
Unequal	.20	.07	.10	.09	.09
Unconstrained Two-Class Generation Model vs. Constrained Four-Class Model					
Equal	.01	.00	.00	.00	.01
Equal	.05	.02	.01	.02	.03
Equal	.10	.04	.03	.03	.05
Equal	.20	.07	.06	.06	.08
Unequal	.01	.00	.00	.00	.00
Unequal	.05	.00	.02	.02	.02
Unequal	.10	.01	.03	.03	.03
Unequal	.20	.04	.06	.06	.07
Constrained Three-Class Generation Model vs. Constrained Four-Class Model					
Equal	.01	.01	.00	.00	.00
Equal	.05	.02	.02	.02	.02
Equal	.10	.03	.05	.03	.03
Equal	.20	.10	.10	.06	.07
Unequal	.01	.00	.00	.00	.00
Unequal	.05	.02	.03	.01	.02
Unequal	.10	.04	.06	.03	.03
Unequal	.20	.07	.13	.05	.08

First, the nature of the deviation of the G^2 statistic from its expected distribution might be related to the degrees of freedom of the statistic. The results related to model comparisons with $df = 1$ and $df = 2$ are quite similar, while both differ substantially in magnitude and direction from the results for model comparisons with $df = 6$. Second, the G^2 statistic corresponding to comparisons between pairs of models with no constraints imposed

on their conditional probabilities might be affected differently than the same statistic for pairs of models in which one or both have equality constraints imposed on their conditional probabilities.

Note that in the former case, the subsuming model contains latent class proportions and conditional probabilities that are not contained in the subsumed model to which it is compared. This is different from the latter case where the pair of compared

models differs only in terms of the number of latent classes. It is, however, unclear why the presence or absence of conditional probability differences found in pairs of models should affect G^2 in such markedly different ways.

The present study did not make the necessary model comparisons to effectively conclude which, if either, of these factors might be important in explaining the divergent effects on G^2 . However, the fact that such different outcomes were obtained for various pairs of models does suggest that, at the very least, extreme caution should be used in identifying a preferred model based on the use of G^2 . This is particularly true when the criteria for selection of a preferred model are based on the use of relatively large critical values (e.g., .10 or .20). Smaller discrepancies between observed and expected proportions of rejection occur when the critical values are more extreme (e.g., .01 or .05). Thus, it may be feasible to establish tentative decision strategies for use when decision criteria are based on these more extreme values.

A reasonable decision strategy for use with more extreme critical values could include: (1) an interval corresponding to clear statistical superiority of the subsuming model, (2) an intermediate interval (which circumscribes the usual critical value) corresponding to no decision regarding statistical superiority, and (3) an interval corresponding to no significant difference in fit provided by the models in the pair. Based on the results in this study, a conservative "no decision" interval might include observed G^2 values with expected probabilities within the interval $.01 \leq [P(G^2) > C] \leq .10$, given $\alpha = .05$. Alternatively, when the models in a specific pair of models of interest were among those considered in this study and the number of manifest variables is five, it is possible to use the empirical proportions of rejection reported here to select an appropriate critical value for G^2 .

References

- Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modeling of data on teaching styles. *Journal of the Royal Statistical Society, A*, 144, 419–461.
- Aitkin, M., & Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, B*, 47, 67–75.
- Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proceedings Cambridge Philosophical Society*, 34, 33–40.
- Bartlett, M. S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society Supplement, B*, 9, 176–197.
- Bergan, J. R. (1982). Latent-class models in education research. *Review of Research in Education, Chapter 10*, 305–360.
- Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika*, 65, 31–38.
- Birch, M. W. (1964). A new proof of the Pearson-Fisher theorem. *Annals of Mathematical Statistics*, 35, 818–824.
- Bowman, K. O., & Shenton, L. R. (1973). Space of solutions for a normal mixture. *Biometrika*, 60, 629–636.
- Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36, 317–346.
- Clogg, C. C. (1977). *Unrestricted and restricted maximum likelihood latent structure analysis*. University Park PA: Population Issues Research Office, Pennsylvania State University.
- Clogg, C. C. (1981). New developments in latent structure analysis. In D. M. Jackson & E. F. Borgatta (Eds.), *Factor analysis and measurement in sociological research*. Beverly Hills CA: Sage.
- Clogg, C. C., & Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79, 762–771.
- Clogg, C. C., & Goodman, L. A. (1985). Simultaneous latent structure analysis on several groups. In N. B. Tuma (Ed.), *Sociological methodology*. San Francisco: Jossey-Bass.
- Clogg, C. C., & Goodman, L. A. (1986). On scaling models applied to data from several groups. *Psychometrika*, 51, 123–135.
- Dayton, C. M., & Macready, G. B. (1976). A probabilistic model for validation of behavioral hierarchies. *Psychometrika*, 41, 189–204.
- Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association*, 83, 173–178.
- Duncan, O. D., & Sloane, D. M. (1982). Latent classes inferred from response-consistency effect. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation, part I*. Amsterdam: North-Holland.
- Everitt, B. S. (1981). A Monte Carlo investigation of the likelihood ratio test for the number of components in a mixture of normal distributions. *Multivariate Behavioral Research*, 16, 171–180.

- Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.
- Formann, A. K. (1982). Linear logistic latent class analysis. *Biometrical Journal*, *24*, 171–190.
- Formann, A. K. (1984). *Die Latent-Class-Analyse*. Weinheim: Beltz.
- Formann, A. K. (1985). Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology*, *38*, 87–111.
- Goodman, L. A. (1974a). The analysis of systems of quantitative variables when some of the variables are unobservable. Part I-A: Modified latent structure approach. *American Journal of Sociology*, *79*, 1179–1259.
- Goodman, L. A. (1974b). Exploratory latent structure analysis using both identified and unidentified models. *Biometrika*, *61*, 215–231.
- Goodman, L. A. (1979). On the estimation of parameters in latent structure analysis. *Psychometrika*, *44*, 123–128.
- Hartigan, J. A. (1975). *Clustering algorithms*. New York: Wiley.
- Hayek, L. C. (1978). *Properties of maximum likelihood estimators for a class of probabilistic models*. Unpublished doctoral dissertation, University of Maryland.
- Hogg, R. V. (1956). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics*, *27*, 529–532.
- IMSL, Inc. (1980). *International Mathematical and Statistical Libraries reference manual*. Houston: Author.
- Macready, G. B. (1982). The use of latent class models for assessing prerequisite relations and transference among traits. *Psychometrika*, *47*, 477–488.
- Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, *2*, 99–120.
- Macready, G. B., & Dayton, C. M. (1980). The nature and use of state mastery models. *Applied Psychological Measurement*, *4*, 493–516.
- Macready, G. B., & Dayton, C. M. (1986). *The use of concomitant variables in latent class acquisition modeling*. Unpublished manuscript, University of Maryland.
- McLachlan, G. J., & Basford, K. E. (1987). *Mixture models: Inference and applications to clustering*. New York: Dekker.
- Quinn, B. G., McLachlan, G. J., & Hjort, N. L. (1987). A note on the Aitkin-Rubin approach to hypothesis testing in mixture models. *Journal of the Royal Statistical Society, B*, *49*, 311–314.
- Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.
- Sedlacek, D. S. (1983). *A Monte Carlo investigation into the robustness of latent class models to violations of the local independence and equal rates assumptions*. Unpublished doctoral dissertation, University of Maryland.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, *50*, 253–263.
- Titterton, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester: Wiley.
- Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, *5*, 329–350.
- Wolfe, J. H. (1971). *A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions* (Technical Bulletin STB 72-2). San Diego: Naval Personnel & Training Research Laboratory.

Acknowledgments

The authors thank the anonymous reviewers for their helpful comments on this paper. The authors are also indebted to the University of Maryland Computer Science Center for providing the computer time necessary for implementing this study.

Author's Address

Send requests for reprints or further information to Judith A. Holt, Gallaudet Research Institute, Center for Assessment and Demographic Studies, Gallaudet University, Washington DC 20002-3625, U.S.A., or George B. Macready, Department of Measurement, Statistics & Evaluation, University of Maryland, College Park MD 20742, U.S.A.