Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: http://www.tandfonline.com/loi/hmbr20

Simultaneous Decision on the Number of Latent Clusters and Classes for Multilevel Latent Class Models

Hsiu-Ting Yu & Jungkyu Park

To cite this article: Hsiu-Ting Yu & Jungkyu Park (2014) Simultaneous Decision on the Number of Latent Clusters and Classes for Multilevel Latent Class Models, Multivariate Behavioral Research, 49:3, 232-244, DOI: 10.1080/00273171.2014.900431

To link to this article: http://dx.doi.org/10.1080/00273171.2014.900431

Published online: 02 Jun 2014.

Multivariate Behavioral Research, 49:232–244, 2014

ISSN: 0027-3171 print / 1532-7906 online
DOI: 10.1080/00273171.2014.900431

Routledge
Taylor & Francis Group

Simultaneous Decision on the Number of Latent Clusters and Classes for Multilevel Latent Class Models

Hsiu-Ting Yu and Jungkyu Park
Department of Psychology, McGill University

The Multilevel Latent Class Model (MLCM) proposed by Vermunt (2003) has been shown to be an excellent framework for analyzing nested data with assumed discrete latent constructs. The nonparametric version of MLCM assumes 2 levels of discrete latent components to describe the dependency observed in data. Model selection is an important step in any statistical modeling. The task of model selection for MLCM amounts to the decision on the number of discrete latent components at both higher and lower levels and is more challenging than standard Latent Class Models. In this article, simulation studies were conducted to systematically examine the effects of sample sizes, clusters/classes distinctness, and the number of latent clusters and classes on the performance of various information criteria in recovering the true latent structure. Results of the simulation studies are summarized and presented. The final section presents the remarks and recommendations about the simultaneous decision regarding the number of latent classes and clusters when applying MLCMs to analyze empirical data.

Discrete latent variables are useful tools in psychological research to represent distinct latent components. Specifically, levels of discrete latent variables can be used to capture or symbolize latent categories of theoretical concepts, constructs, entities, or subgroups. Examples of such discrete latent variables include Sternberg's (1998) profile of thinking styles; Ainsworth, Blehar, Waters, and Wall's (1978) three attachment styles (secure, anxious-resistant, and avoidant); Bennett and Jordan's (1975) teaching styles; Fischer and Fischer's (1979) styles of teaching and learning; and Jung's (1971) psychological types. The Latent Class Model (LCM) (Lazarsfeld & Henry, 1968) is the classical analytical approach of using the discrete latent variable to explain dependency in observed categorical responses. The LCM assumes that the dependency between responses is due to population heterogeneity. The different components in a discrete latent variable are called latent classes, and subjects in the same class will respond to items in a similar fashion.

When analyzing data with a nested structure, where the observations are nested within a higher level unit (e.g., school or company), additional considerations should be taken into account because data with such a structure are generally not independent of each other. For example, the scores of an achievement test may correlate more highly for students in the same class than for students from other classes. This dependency may be due to the commonalities they share by receiving the same teaching material and homework, having the same instructor, or interacting with each other. Various multilevel extensions have been proposed within the LCM framework to model such additional dependency as well as dependency due to population heterogeneity (e.g., Asparouhov & Muthén, 2008; Di & Bandeen-Roche, 2011; Henry & Muthén, 2010; Vermunt, 2003, 2004; Vermunt & Magidson, 2008). The Multilevel Latent Class Model (MLCM; Vermunt, 2003) is a proposed model used to incorporate possible dependency due to a nested structure by introducing either a discrete or a continuous random effect to account for systematic variation. The nonparametric version of MLCM assumes discrete latent variables at both higher and lower levels to account for dependency observed in collected data.

Model selection usually refers to selecting a statistical model from a set of potential models for the final interpretation of the results. Ideally, the selected model should be able to describe the data reasonably well without unnecessary complexity. In the context of LCM, the task of model selection is to determine the number of discrete latent components based on observed responses. These latent components characterize different unobserved subgroups that are internally homogenous. Among the many model selection

Correspondence concerning this article should be addressed to Hsiu-Ting Yu, McGill University, Department of Psychology, 1205 Dr. Penfield Avenue, Montreal, Quebec H3A 1B1, Canada. E-mail: ht.yu@mcgill.ca

methods, selecting the model by Information Criteria (IC) is a common and popular method for applied researchers. The relative performance of various IC in selecting the number of latent classes has been extensively studied in the context of standard LCM and mixture models (e.g., Collins, Fidler, Wugalter, & Long, 1993; Dias, 2004; Henson, Reise, & Kim, 2007; Lin & Dayton, 1997; Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2007; Yang, 1998, 2006; Yang & Yang, 2007).

The model selection task in MLCM is to determine the number of latent classes at different levels of the hierarchical structure. Determining the number of latent components at both levels is more challenging than at a single level because decisions (the number of latent components) at different levels are not independent of each other. Two simulation studies conducted by Lukočienė and Vermunt (2010) and Lukočienė, Varriale, and Vermunt (2010) explored the performance of IC for model selection in MLCM under various conditions. These two studies focused mainly on the stepwise approach, whereas the performance of the simultaneous approach has not yet been examined extensively. The present study also studied relatively more complex latent structures than previous studies, which dealt only with relatively simple structures (two or three higher and lower classes).

The purpose of this article is to (a) review the methods and approaches of model selection in MLCM, (b) investigate the factors affecting the performance of IC in recovering the true latent structure, (c) investigate the performance of various IC for model selection simultaneously at higher and lower levels, and (d) provide recommendations and practical guidelines to use IC for determining the number of latent classes simultaneously at both levels.

In the following sections, the statistical specification of the MLCM is introduced. Several methods including IC as methods of model selection are reviewed and compared. Simulation studies were conducted to examine the performance of IC in deciding the number of classes at both higher and lower levels simultaneously. The results of these simulations are presented and discussed. This article concludes with remarks on the usage of IC as a model selection method and general recommendations to applied researchers for using IC for model selection in the MLCM.

MULTILEVEL LATENT CLASS MODELS

The primary purpose of the MLCM (Vermunt, 2003) is to account for possible data dependency with discrete latent variables for the nested data structure. Due to the nature of the nested data structure, the independence assumption of the standard LCM is often violated. To account for such dependency, the MLCM incorporates parametric or nonparametric random parameters that are allowed to be varied across higher level units. Because the goal of this article is to study the performance of IC in recovering the number of latent components simultaneously at higher and lower levels, we focus on the nonparametric MLCM, which assumes discrete latent components at both levels.

MLCM can be formally specified as follows: the latent variable (H_g) is the discrete latent variable at higher level (groups) with L latent clusters, and X_{gi} denotes the discrete latent variables at lower level (individuals) with M latent classes. Each outcome of discrete random variables can be conceptualized as a latent cluster/class consisting of groups/individuals that are homogenous within each cluster/class but distinct between clusters/classes in the response patterns. The terms "clusters" and "classes" are used to differentiate the higher and lower classes. Let Y_{gij} be the response to the jth item of a subject (i) in a group (g), where $g = 1, \ldots, G$, $i = 1, \ldots, n_g$, and $j = 1, \ldots, J$. The vector $\mathbf{Y}_{gi}$ represents J responses for a subject i nested in group g, and $\mathbf{Y}_g$ denotes the full vector of responses for all subjects in group g.

The standard LCM could be defined by a latent variable at individual level X_i without consideration of group level; therefore, the density of the response of subject i on item j is

$$P(Y_{ij}) = \sum_{m=1}^{M} P(X_i = m) \prod_{j=1}^{J} P(Y_{ij} | X_i = m). \quad (1)$$

An MLCM is defined by two separate equations for higher and lower levels. For the higher level the subscript i in Equation (1) is replaced by g. So the probability of observing a certain response pattern for all subjects nested in group g is

$$P(\mathbf{Y}_g) = \sum_{l=1}^{L} P(H_g = l) \prod_{i=1}^{n_g} P(\mathbf{Y}_{gi} | H_g = l). \quad (2)$$

Equation (2) assumes that each group belongs to only one l(latent cluster), and conditional densities for each of the n_g (individuals) within the g (group) are independent of each other given the latent cluster membership. The first term, $P(H_g = l)$, is represented by a vector with each element representing the probabilities of g being in the cluster l ($l = 1, \ldots, L$). Because the clusters are assumed to be exhaustive and mutually exclusive, elements of this vector can be conceptualized as cluster sizes and thus the sum of this vector equals 1.

At the individual level, the probability of obtaining a certain response pattern for each subject is

$$P(\mathbf{Y}_{gi} | H_g = l) = \sum_{m=1}^{M} P(X_{gi} = m | H_g = l) \times \prod_{j=1}^{J} f(Y_{gij} | X_{gi} = m, H_g = l). \quad (3)$$

The term $P(X_{ig} = m | H_g = l)$ denoted as π_{ml} is the conditional latent class probabilities. It represents the distribution of latent class probabilities within a particular latent cluster.

The conditional response density, $f(Y_{gij}|X_{gi} = m, H_g = l)$, is the probability of observing a particular response on variable j of individual i in group g given the latent cluster membership (l) and latent class membership (m). In most multilevel extensions of LCM (e.g., Asparouhov & Muthén, 2008; Vermunt, 2003, 2004; Vermunt & Magidson, 2008), a restricted model is preferred by imposing a constraint on the conditional response density.

This constraint implies that the conditional response density is affected only by the latent class memberships but not the higher level latent cluster membership. This assumption not only reduces the number of parameters and simplifies the model but also makes interpretation of the results more intuitive: only the latent class membership has effects on the observed responses. In addition, the local independence assumption, which is typically made in the latent variable modeling, is also assumed. By combining Equations (2) and (3) with the assumption of no effects of latent cluster membership to response probabilities and the assumption of local independence, the MLCM is

$$P(\mathbf{Y}_g) = \sum_{l=1}^{L} P(H_g = l) \times \left(\prod_{i=1}^{n_g} \sum_{m=1}^{M} P(X_{gi} = m|H_g = l) \prod_{j=1}^{J} f(Y_{gij}|X_{gi} = m) \right). \quad (4)$$

The density $f(Y_{gij}|X_{gi} = m)$ depends on the assumed distributions of responses. Suppose the response vector $\mathbf{Y}_{gi} = (Y_{gi1}, Y_{gi2}, \ldots, Y_{giJ})^T$ consists of J binary variables. Then, the mth latent class density is given by $f(Y_{gij}|X_{gi} = m) = \rho_{mj}^{Y_{gij}}(1-\rho_{mj})^{1-Y_{gij}}$, where ρ_{mj} denotes the probability of endorsing item j for an individual belonging to latent class m. The conditional response probabilities can take the form of multinomial or other more general distributions such as Poisson. The estimations of model parameters in an MLCM can be obtained through a modified version of the exception-maximization algorithm (Vermunt, 2003, 2004), which is available in Latent GOLD 4.5 (Vermunt & Magidson, 2008). Other specifications of MLCM are discussed by other authors: for example, Henry and Muthén (2010) presented three different versions of MLCMs having normally distributed random effects (i.e., the parametric approach) at different levels, whereas Di and Bandeen-Roche (2011) introduced an MLCM with a Dirichlet distributed random effect.

MODEL SELECTION METHODS

Statistical tests and fit indices have been proposed as the model selection methods in determining the number of latent classes. The application of the likelihood ratio test (LRT) has been questioned because the parameters tested are at the boundary of the parameter space. In such a case, the test statistic does not follow the chi-square distribution (see Aitkin & Rubin, 1985; Clogg, 1995; Everitt, 1981, 1988; McLachlan & Peel, 2000). Moreover, Read and Cressie (1988) indicated that the likelihood ratio statistic might also not follow the chi-square distribution for sparse data. The usage of LRT for model selection is usually not recommended in LCM, although there have been several attempts to determine a proper distribution of the test statistic (e.g., Lo, Mendell, & Rubin, 2001; Vuong, 1989).

Several alternative methods have been proposed for determining the number of latent classes. For example, the bootstrapped likelihood ratio test (BLRT), which is based on resampling techniques, was proposed to obtain the approximate p values through empirical sampling distributions (Dias & Vermunt, 2008; McLachlan, 1987; McLachlan & Peel, 2000). The BLRT is reported to be the most accurate and consistent tool for determining a number of classes among adjusted likelihood ratio tests (Nylund et al., 2007). However, these methods are computationally intensive, especially when there are many candidate models to choose from. Moreover, the performance of these alternative methods has not been systematically studied when applying to multilevel models, thus there are no standards or guidelines available to assist researchers in the task of model selection.

Information Criteria

Information criteria (IC) were suggested as an alternative to overcome the problem of LRT. IC are measures of the relative goodness of fit of a statistical model and can be formulated in this general form:

$$IC = -2 \log L + C. \quad (5)$$

The first term, $-2 \log L$, is the negative logarithm of the maximum likelihood, which decreases when model complexity increases. The second term, C, penalizes the complexity of the model. Therefore, IC aim to find a good balance between model fitness (trying to maximize the likelihood function) and parsimony (penalizing additional complexity).

Akaike information criterion (AIC; Akaike, 1973) was one of the earlier propositions of information criteria. AIC chooses a model that minimizes Equation (5) with $C = 2p$, where p is the number of parameters in the model. Several studies reported the tendencies to overestimate the number of latent classes (e.g., Celeux & Soromenho, 1996; Dias, 2004; Soromenho, 1993). Nevertheless, Gonzalo and Pitarakis (1996) showed that AIC outperforms other IC in models with high-dimensional parameter spaces. Two modified versions of AIC were suggested by Bozdogan (1987, 1993) for multivariate normal mixture models: AIC3 and consistent Akaike information criterion (CAIC). AIC3 uses the value of 3 as the cost of fitting an additional parameter instead of 2 in the regular AIC (i.e., $C = 3p$). CAIC includes the sample size as part of the penalty term and has been shown to consistently outperform AIC (Nylund et al., 2007;

TABLE 1
Summary of Information Criteria

Criterion	Definition	Reference
AIC	$-2LL + 2P$	Akaike, 1973
AIC3	$-2LL + 3P$	Bozdogan, 1993
CAIC	$-2LL + (1 + \log(n))P$	Bozdogan, 1987
BIC	$-2LL + \log(n)P$	Schwarz, 1978
ABIC	$-2LL + \log((n + 2)/24)P$	Sclove, 1987

Note. CAIC = consistent AIC; ABIC = adjusted BIC.

Yang, 2006). From a different theoretical background, the Bayesian information criterion (BIC; Schwarz, 1978) aims to find an appropriate modification of maximum likelihood by studying the asymptotic behavior of Bayes's estimators under a class of proper priors, which assigns a positive probability on the same lower dimensional spaces of the parameter vector. The penalty term for the BIC is $C = p \log N$. Sclove (1987) suggested the adjusted BIC (ABIC) using $C = p \log((N + 2)/24)$ as a penalty term for the models with limited sample sizes.

The discussed IC are summarized in Table 1. Note that AIC and AIC3 depend only on the number of parameters, whereas CAIC, BIC, and ABIC involve both the number of parameters and the sample size in the penalty term. Considering the penalty term, CAIC usually imposes the greatest penalty among discussed IC when sample size is more than 100, and BIC has a slightly smaller penalty than CAIC. This property makes CAIC and BIC better for smaller models. On the other hand, AIC and ABIC have a relatively smaller weight on the number of parameters, so they favor larger models. The amount of penalty of ABIC will exceed AIC when the sample size becomes more than 176.

MODEL SELECTION METHODS FOR MLCM

Previous studies of model selection for standard LCMs have addressed three important components that determine the overall performance of IC in selecting the true number of latent classes: sample sizes, model complexity, and conditional response probabilities (e.g., Lukočienė et al., 2010; Nylund et al., 2007; Yang & Yang, 2007).

In general, larger sample sizes yield better accuracy of IC (e.g., Lin & Dayton, 1997). There are two definitions of sample sizes in MLCM: the number of higher level units (G) and the total number of lower level units (N). The performance of IC using either definition has not been studied extensively but is systematically investigated in the present study. Model complexity concerns the number of parameters to be estimated in a model. The performance of IC decreases as the number of latent classes increases in LCM. However, the total number of parameters in a MLCM is determined by the number of latent clusters, the number of latent classes, and the number of items. How these three factors affect the performance of IC in a MLCM is also examined in the simulation study of this article.

The conditional response probabilities directly reflect the degree of separation between classes in an LCM (Yang & Yang, 2007), and the performance of IC is better when classes are well separated. In the present study, we use the term "clusters/classes distinctness" to characterize the clusters and classes defined by the conditional response probabilities (ρ_{mj}) and the conditional latent class probability (π_{ml}) in an MLCM and expect the IC to perform better when clusters/classes are more distinct from each other. However, we also note that in addition to the clusters/classes distinctness, the separation of classes and clusters in the MLCM also relates to the previously mentioned factors of sample size and model complexity. Lukočienė et al. (2010) described how those three factors in higher and lower levels influence the degree of separation among clusters and classes.

The prior studies have different conclusions in the performance of individual information criteria in recovering the number of latent classes. For the single-level LCM, researchers have suggested using BIC as a standard measure to determine the number of latent classes because it consistently outperforms other IC (e.g., Hagenaars & McCutcheon, 2002; Magidson & Vermunt, 2004). Fonseca and Cardoso (2007) also reported that BIC works well when the responses are continuous; however, other studies (e.g., Yang, 1998) reported that BIC sometimes underestimates the number of latent classes especially when the sample size is small or classes are not well defined. The simulation conducted by Andrews and Currim (2003) and Dias (2004) showed that AIC3 was the best criterion for the model selection in the single-level LCM with categorical responses, whereas other authors advocated ABIC for the same type of model (e.g., Nylund et al., 2007; Yang, 2006).

For model selection in other types of single-level mixture models such as Factor Mixture Models (Arminger, Stein, & Wittenberg, 1999; Dolan & Van Der Maas, 1998) or Growth Mixture Models (Muthén & Shedden, 1999), ABIC has been reported as the best index in recovering the number of mixture components (Henson et al., 2007; Tofighi & Enders, 2007), whereas Nylund et al. (2007) concluded that BIC outperforms ABIC.

Two Model Selection Approaches

One recommended model selection strategy in MLCM is the simultaneous approach. Specifically, a set of candidate models with combinations of L and M are fitted, and then the IC obtained from the models are compared to determine the optimal combination. For example, Bijmolt, Paas, and Vermunt (2004) determined the country-level latent classes and consumer-level latent classes by comparing the models with particular ranges in L and M(1 to 15 for M and 1 to 8 for L). The simultaneous approach usually requires more models to be fitted than the iterative approach; however, this approach is more straightforward and can be easily applied in empirical applications.

The decisions at higher and lower levels are related because the decisions are made based on the data with a nested structure. Specifically, the clusters specified in Equation (4) are defined by marginalizing over the distribution of classes. This specification implies that cluster membership is defined through the class membership of individuals in the same group, not directly through their responses. Thus, the latent structure at the lower level (classes; e.g., the number of classes and the distinctness among classes) has an impact on the decision of the number of latent clusters at the higher level. Although less prominent, the higher level structure, such as the number of clusters and cluster distinctness, also affects the decision at the lower level. For example, well-separated clusters will provide additional information on individuals' class membership when responses within the same group are similar. Therefore, the model selection at a given level is influenced not only by the factors in the same level but also by the latent structure at the other level. One advantage of the simultaneous approach is to take into account the between-level dependency in the same step.

An alternative strategy to determine the number of discrete latent components in an MLCM is the stepwise approach originally proposed by Vermunt (2003). Lukočienė et al. (2010) proposed a three-step approach to determine the optimal number of latent clusters and classes in an iterative fashion. The first step of this method is to determine the optimal number of classes (M) when assuming there is only one latent cluster (i.e., $L = 1$). The next step is to decide the number of clusters (L) by fixing the number of latent classes (M) to the value obtained in the first step. The third step is to redetermine M by fixing L to the value chosen from the previous step. The second and third steps are then iterated until there is no change in the number of the latent clusters and classes.

When discrete latent variables are assumed at both levels, the iterative procedure has the advantage of greater efficiency, as the decision of the current step is "guided" by the previous step and thus fewer candidate models are needed to be fitted in the process. However, the main disadvantage of the stepwise approach is that the dependency between the two decisions is only partially accounted for in each step. Moreover, the task of model selection needs to be conducted separately at each level because different definitions of "sample sizes" are used at higher and lower levels for IC with sample size in the penalty term.

SIMULATION STUDY

The simulation study has three goals: (a) to explore the effects of factors determining the performance of IC, (b) to compare the performance of different IC for choosing the correct number of latent classes, and (c) to compare the simultaneous and stepwise approaches in recovering true multilevel latent class structures under a variety of conditions.

Seven factors were included in the design to create various levels of cluster/class distinctness: (a) number of clusters, (b) number of classes, (c) conditional latent class probabilities, (d) conditional response probabilities, (e) higher level sample sizes, (f) lower level sample sizes, and (g) number of items (binary). The probabilities of assigning groups to clusters $P(H_g = l)$ were assumed to be equal; in other words, the clusters had the same sizes.

In this study, the numbers of clusters and classes were specified as varying between two and five, respectively. In theory, this specification results in a total of 16 models. However, our experience in estimating the parameters of MLCM with complex higher level structures (i.e., having more clusters than classes) has suggested a much poorer fit. Therefore, only models with an equal or smaller number of clusters than classes were considered in the simulation study. This constraint resulted in 10 different MLCMs: models with the same number of latent classes and clusters (denoted as L2M2, L3M3, L4M4, and L5M5) and models with fewer clusters than classes (L2M3, L2M4, L2M5, L3M4, L3M5, and L4M5). These 10 models covered a wide range of possible latent structures in empirical applications of the MLCM.

The values of the conditional latent class probabilities (π_{ml}) and conditional response probabilities (ρ_{mj}) were systematically manipulated. These two factors are referred to as *cluster distinctness* and *class distinctness* following the definition in Yang and Yang (2007). Although the distinctness among clusters/classes can be defined in many ways in practice, the construct of cluster/class distinctness is very close to the construct of uniqueness among clusters/classes. Two sets of values were chosen to generate different degrees of distinctness among clusters and classes, respectively. The values of π_{ml} were set to differ greatly among the clusters in the conditions designed to have more distinctive clusters. Meanwhile, the values of π_{ml} were designed to be more evenly distributed in less distinctive conditions.

Likewise, the values of ρ_{mj} were designed to differ to a greater extent across classes in the more distinctive conditions among classes, whereas the values of ρ_{mj} were more evenly distributed in the less distinctive conditions. These specifications resulted in four conditions of class and cluster distinctness; they are referred to as H-H (more distinctive clusters and classes), H-L (more distinctive clusters and less distinctive classes), L-H (less distinctive clusters and more distinctive classes), and L-L (less distinctive clusters and classes). The exact values used in the simulation study are presented in the Appendix.

The sample sizes were manipulated by specifying the number of groups (G) and number of individuals per group (ng). The number of groups was set to 50 and 100, with 10, 25, and 50 individuals per group to represent small, medium, and large groups. These specifications resulted in five levels of total sample size (N): 500, 750, 1,000, 2,500, and 5,000. The specified number of individuals per group was similar to the design in Lukočienė et al. (2010; $ng = 5$, 10, 20, and 50), but the number of groups in their study ($G = 30$, 100, 1,000) was generally larger than the values chosen in this study. The number of items in the simulation was set as either 8 or 12.

The specifications described earlier resulted in 480 ($10 \times 2 \times 2 \times 3 \times 2 \times 2$) conditions. For each condition, 50 data sets were generated according to the parameter specifications of that particular condition. Each data set was then fitted to the 10 possible models with different numbers of clusters and classes. Among the 10 possible models, one of them was the "true" model with a latent structure identical to the structure that the data were generated from. The other 9 models were "wrong" models and had latent structures different from the true model. Data generation was conducted using MATLAB V7.13 (R2011b), and Latent Gold 4.5 with a syntax module (Vermunt & Magidson, 2008) was used to estimate the model parameters. The log-likelihood values of each data set fitted to each model were recorded to compute the eight IC (including ABIC, AIC, AIC3, BIC, and CAIC with two different N values) for further analyses.

SIMULATION RESULTS

The specifications of the seven factors described earlier resulted in different degrees of difficulty in simultaneously recovering the true numbers of the latent classes and clusters. The probability of correctly recovering the true model by chance was only 0.1 because there were only 10 possible latent structures to choose from. To gain a better sense of the overall difficulty level of the design, the results show that about 15% of the conditions had an average recovery rate greater than 80% when combining all the levels of the seven controlled factors, and about 50% of the conditions created in the simulation study had an average recovery rate lower than 20%. Therefore, because the majority of the conditions designed in our simulation are considered "difficult" in terms of recovering the true number of latent clusters and classes, the analyses and reports focus on comparing the relative performance between the levels of the controlled factors rather than the absolute recovery rate of the true latent structure.

A linear mixed-effect model was fitted to the data; the fixed effects included the seven design factors, the interactions of each of the factors with the IC, and the pairs among the seven factors interacting with the IC. Among the seven factors, the cluster distinctness and class distinctness were combined and regarded as one factor with four levels to investigate the effect in a more integrative manner. Five IC, including ABIC, AIC, AIC3, BIC, and CAIC, were considered within factors. For IC with the sample size (N) in the equation (i.e., ABIC, BIC, and CAIC), two different versions were computed for using the total sample size ($N = G \times ng$) and the number of groups (G) as the penalty term. The IC with different "sample sizes" as penalty terms were denoted by their corresponding subscript (i.e., IC_N, and IC_G). The dependent variable was the percentage of correctly recovering the latent structure simultaneously at both the higher and lower levels. In the following sections, the results are presented separately for each of the controlled factors.

Distinctness Among Latent Clusters and Classes

The overall recovery rates between the four levels of cluster/class distinctness are significantly different ($F_G(3, 224) = 4.98$, $p < 0.01$). As shown in Figure 1, all IC

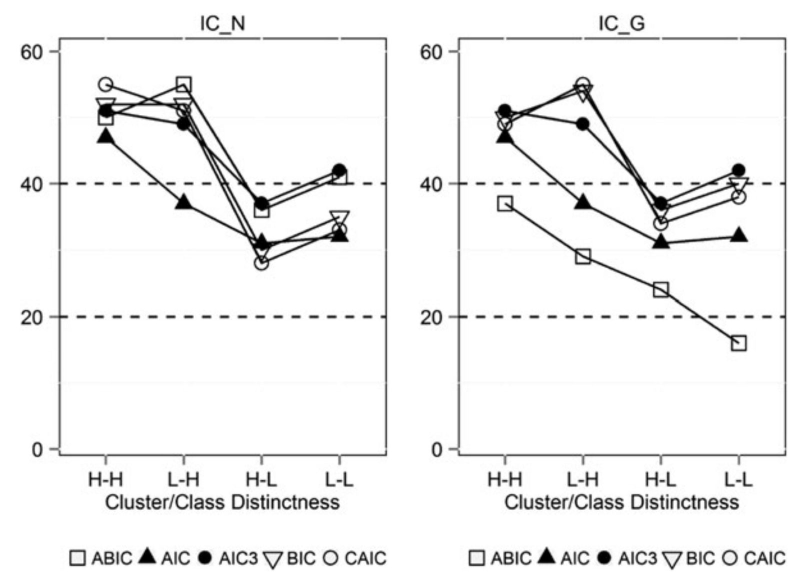

FIGURE 1 The mean recovery rate of the two sets of IC_N, IC_G under the four scenarios of cluster/class distinctness.

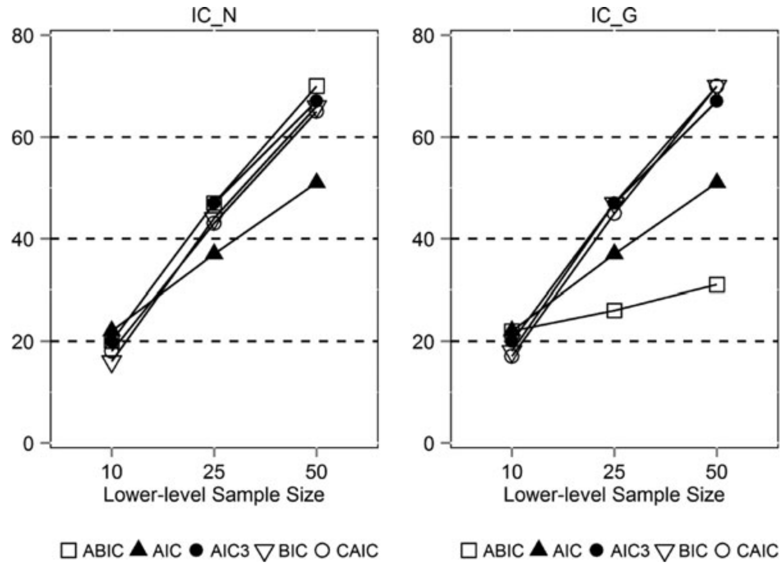

FIGURE 2 The mean recovery rate of IC_N and IC_G under different numbers of lower level sample sizes.

clearly performed better in recovering the true latent structure under the conditions that classes were more distinctive (H-H and L-H conditions) than the conditions of less distinctive classes (H-L and L-L conditions). This result suggests that a greater degree of distinctness among classes gives better recovery rates for all IC regardless of the clusters being more or less distinctive; however, having more distinctive clusters does not improve recovery rates, especially in the conditions that classes are less distinctive. In other words, the distinctness among classes is pivotal in correctly recovering the true latent structure compared with the distinctness among clusters. This result is not surprising because the cluster structure in an MLCM is built based on the information of class membership. Therefore, when classes are not well separated at the lower level, the higher level clusters are not well defined.

There is a significant interaction between the IC and cluster/class distinctness ($F_G(12, 224) = 5.78$, $p < 0.01$, $F_N(12, 224) = 4.30$, $p < 0.01$). AIC is inferior to other IC in most conditions regardless of cluster or class distinctness. The IC imposing a relatively larger penalty, such as CAIC and BIC, perform better under the condition with more distinctive classes (H-H and L-H conditions), but AIC3 works best in the condition with less distinctive classes (H-L and L-L conditions). In general, the two sets of IC (IC_N and IC_G) have similar patterns of recovery rates, except for the $ABIC_G$, which is poorer than $ABIC_N$ due to insufficient penalty.

Sample Sizes

For the effects of the number of higher level sample sizes, the two levels of group size (50 and 100) did not differ in terms of the recovery rates among IC when using N in the penalty term. However, there was a significant interaction between the number of groups and IC_G ($F_G(4, 224) = 6.68$, $p < 0.05$). Further examinations found that $ABIC_G$ performs much worse than IC with a smaller group size ($G = 50$), which may result from the fact that the $ABIC_G$ does not impose sufficient penalties when using G in the penalty term.

The effects of lower level sample sizes are presented in Figure 2. The results suggest that recovery rates sharply increase as the number of subjects in a group (ng) increases ($F_N(2, 224) = 81.55$, $p < 0.01$, $F_G(2, 224) = 31.61$, $p < 0.01$). The recovery rates were only around 20% with the small sample sizes ($ng = 10$), but increased to almost 70% when the sample size was large ($ng = 50$). This pattern suggests that having sufficient number of subjects in a group is a major factor in improving the accuracy of IC. This pattern is consistent in IC_N and IC_G.

A significant interaction between IC and lower level sample sizes was observed ($F_G(8, 224) = 27.55$, $p < 0.01$, $F_N(8, 224) = 3.07$, $p < 0.01$). As shown in Figure 2, CAIC, BIC, and AIC3 maintained high recovery rates when the sample sizes were medium ($ng = 25$) and large ($ng = 50$). On the other hand, AIC showed a slightly higher recovery rate than CAIC, BIC, and AIC3 when the sample size was small ($ng = 10$), but it performed worse than IC with larger sample sizes ($ng = 25$ and 50). In addition, ABIC was the worst criterion with a number of groups (G) as the "sample size," especially in the conditions of medium and large sample sizes, but ABIC outperformed other IC when using N in the penalty term.

Figure 3 presents the recovery rates of IC_N under different levels of sample size and four conditions of class/cluster distinctness. The recovery rates for small sample size ($ng = 10$) were considerably lower than other sample sizes across all four conditions of class/cluster distinctness. The overall recovery rate of IC was about 30%, even in more

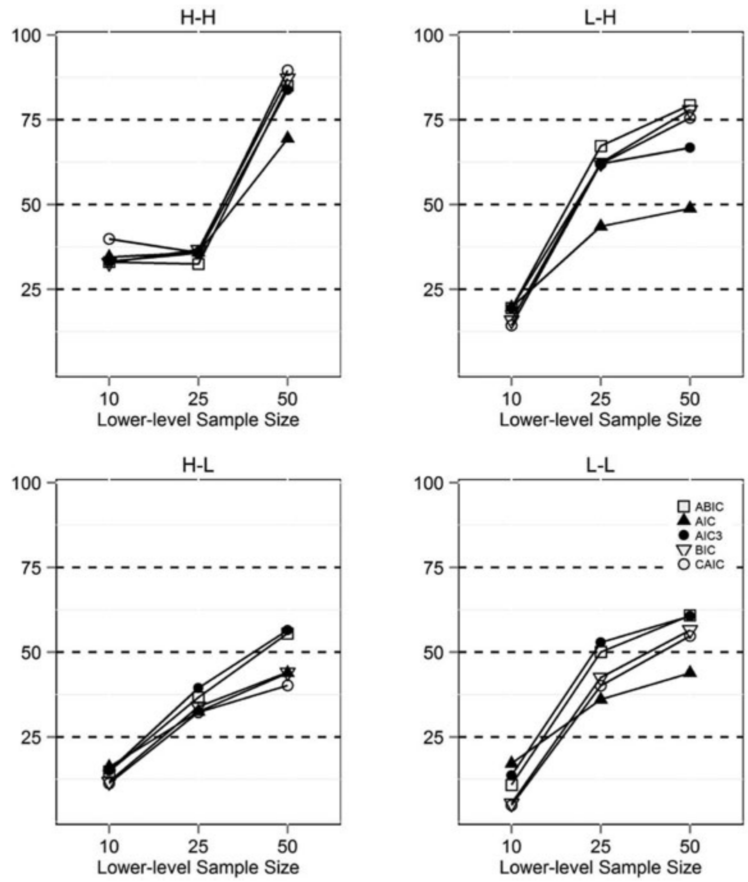

FIGURE 3 The effects of lower level sample size for IC_N in recovering the true latent structure under four levels of cluster/class distinctness.

distinctive cluster and class conditions (H-H). The rates of recovery rapidly increased as the sample size increased in all four conditions for all IC, although this rate of improvement was not as strong for AIC. There was an interesting pattern observed in the H-H condition: the largely improved recovery rate was observed only in the large sample size ($ng = 50$). This pattern suggests that even in more distinctive cluster/class conditions, if the number of subjects is fewer than 25, it will not warrant a satisfactory recovery rate. Further examinations also found that having a small sample size leads to poor performance ($< 40\%$) when interacting with other controlled factors in this study, suggesting that a sufficient number of subjects per group is crucial to achieve better recovery rates.

Numbers of Latent Classes and Clusters

The patterns of recovery rates of IC_N and IC_G, with respect to number of clusters and classes, were similar, except that $ABIC_G$ performed much worse than $ABIC_N$; thus, the mean recovery rates of IC_N of different numbers of latent clusters and classes were plotted in Figure 4. Unlike a decreasing pattern of recovery rate under the more complex model typically found in related studies (e.g., Yang & Yang, 2007), there was no clear pattern in terms of the overall performance related to the number of latent classes and clusters in the present simulation study. However, the recovery rate of individual IC significantly interacted with different numbers of classes ($F_N(12, 224) = 4.31$, $p < 0.01$).

BIC and CAIC performed well under the model with smaller numbers of classes (simpler structure), but the recovery rates gradually decreased as the latent structures became more complex by imposing more penalties than necessary. The mean recovery rates of AIC, AIC3, and ABIC exhibited U-shaped patterns, signifying that performance was better in the most simple and complex conditions. One possible explanation for this pattern is that the number of classes affects both class and cluster separation but in an opposite direction (see Lukočienė et al., 2010, p. 261). The classes became less separated as the number of classes increased, but having a greater numbers of classes provided additional information among the clusters and led to better separated clusters.

The performance of AIC3 was between AIC and CAIC/BIC because of the balanced penalty term imposed. Further examinations found that AIC performed relatively worse than other IC in most conditions but that it outperformed other IC in the most complex conditions ($L = 5$ and $M = 5$). Moreover, the performance of $ABIC_N$ was relatively

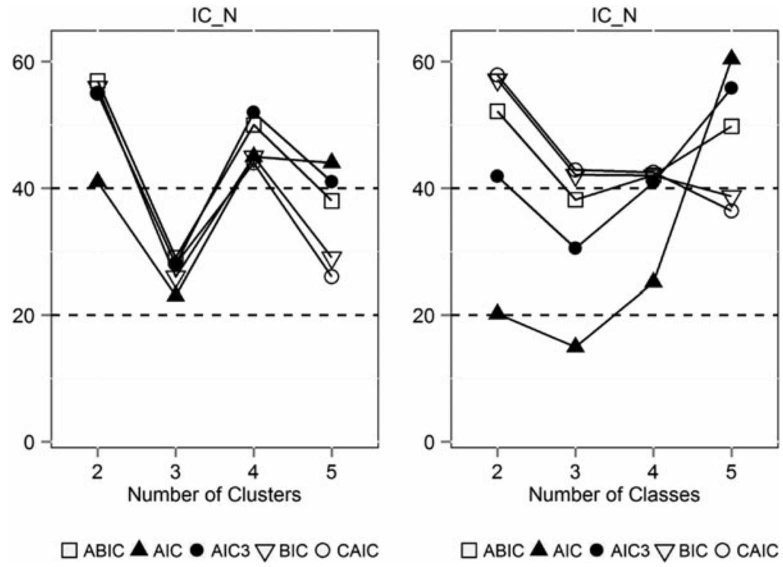

FIGURE 4 The mean recovery rate of IC_N under different numbers of latent clusters and classes.

stable under different numbers of classes. Specifically, the recovery rate of ABIC was higher/lower than AIC and AIC3 in simpler/more complex structures. Similarly, the recovery rate of ABIC was lower/higher than BIC and CAIC in simpler/more complex structures.

Comparisons Between the Simultaneous Approach and the Stepwise Approach

We compared the three-step approach proposed by Lukočienė et al. (2010) and the simultaneous approach investigated in this article in terms of the performance of IC in recovering the true latent structure. The overall recovery rate of each IC using the three-step procedure, simultaneous approach using N in the penalty term, and simultaneous approach using G in the penalty term are plotted in Figure 5. As shown in Figure 5, the mean recovery rates of AIC3, BIC, and CAIC performed better under the three-step approach than the simultaneous approach, whereas the $ABIC_N$ outperforms the ABIC in the three-step procedure, and the performance of AIC is similar in both approaches. However, we note that the two approaches are not exactly comparable. First, the design in the present simulation concerned a restricted set of 10 models (i.e., models with more numbers of latent classes than clusters), but the stepwise procedure considered 5 models with a fixed number of classes or clusters at each step.

DISCUSSION AND CONCLUSIONS

Model selection is an important step of any statistical analysis. The goal is to find a model that has minimal discrepancy between the predicted data and the empirically observed data. Employing IC has commonly been proposed as the model selection method for LCM; however, the performance of IC as the model selection method for MLCM has not been extensively studied. The two most relevant studies, conducted by Lukočienė and Vermunt (2010) and Lukočienė et al. (2010), recommended determining the number of latent classes at the lower and higher levels with a stepwise fitting strategy to make the two dependent decisions more efficient.

This study took a simultaneous approach to model selection in MLCM. The results suggest that the performance of IC largely depends on factors that affect class and cluster separations. The performance of IC increased significantly as the classes (lower level) became more distinct, whereas the effect of cluster distinctness was not clearly observed in the study.

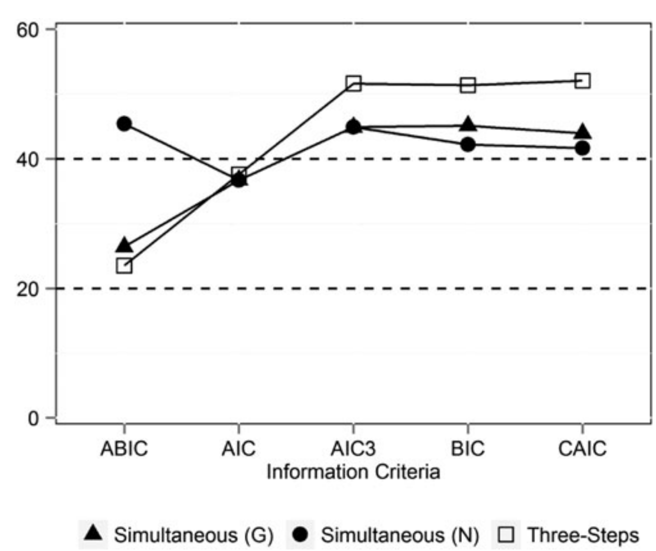

FIGURE 5 The comparisons of mean recovery rate of IC between the simultaneous approach and three-step approach.

The results related to the effect of sample size suggested that there was no significant difference between the two levels of the number of groups ($G = 50$ and 100). This result may be due to the lack of sufficient differences among the specified values to exhibit the difference in the performance of IC. However, the larger group sizes (ng), in general, led to better performance in recovering the latent structure for all compared IC. The improvement in the recovery rate was very clear when the group size was large. The accuracy rate reached 90% under the condition of more distinct clusters and classes. Therefore, we recommend that researchers take this into consideration when designing studies. CAIC, BIC, and AIC3 performed better than AIC and ABIC in most conditions, whereas AIC worked well only when the sample was small ($ng = 10$).

The results of the simulation study also showed that the complexity of the latent structure affected the performance of IC. Generally speaking, the simplest latent structure (e.g., H2L2) can easily be recovered by most IC. In addition, we observed some unique patterns in the recovery rates, that is, the recovery rates for BIC and CAIC dropped as the number of classes increased, but AIC, AIC3, and ABIC performed relatively better even in the condition of five latent classes.

One main characteristic of nested data is that lower level units are grouped into higher level units, and this nested structure may lead to several possible choices of N as the penalty term in IC. The results of our simulation study suggested that using G (number of groups) led to a slightly better performance than N (total sample size) in BIC and CAIC. Specifically, the penalty that using N in the formula is too harsh for BIC and CAIC and led to poorer recovery rates than IC_G. One exception was ABIC, which performed much better when total sample size (N) was used in the penalty term. This result is in agreement with the recommendation of Lukočienė et al. (2010).

IC are simple and feasible methods for model selection in MLCM. Further, IC are generally part of the standard output in most statistical software. It is important to note that the default setting of N may differ in different software, and the calculation of IC may require additional steps to ensure that the proper N is used. As most software reports the log-likelihood value in the output, IC can be easily calculated by hand before proceeding to the step of selecting the best model.

The performance of an individual information criterion largely varied depending on the level of separation between classes. In general, BIC_G and $CAIC_G$ could be great tools under the conditions of well-separated classes, such as having sufficient number of lower level samples and well-defined class structures. On the other hand, AIC is a more favorable criterion than others when sample sizes at the lower level are limited. Lukočienė and Vermunt (2010) recommended using AIC3 and BIC_G for model selection in MLCM using the stepwise approach, and our simulation study supported this recommendation. We also recommend $ABIC_N$ because it reached a recovery rate similar to that of AIC3 and BIC_G in our simulation study.

Another key factor in the model selection of MLCM relates to the model's latent structure. The distinctness among classes plays an important role in correctly recovering the true latent structure. Although the true latent structure of the data is "unknown," the literature review, substantive knowledge of the topics, or the patterns of estimated conditional response probabilities can provide clues about the true latent structure. This information can guide the selection of the appropriate IC to make a more confident decision.

One limitation of this study is the performance between the simultaneous and stepwise approach cannot be directly compared because of the simulation design of this study. In addition, the recommendations of model selection using IC in this study are limited to the MLCM with discrete latent variables at both levels. Future research can explore related models such as the Multilevel Mixture Factor Model (Varriale & Vermunt, 2012) or Multilevel Growth Mixture Model (Palardy & Vermunt, 2010).

In summary, this study investigated the degree of clusters/class distinctness, sample sizes, and model complexity in relation to the performance of IC in recovering the true latent structure of MLCM. Issues such as the iterative versus the simultaneous approach and the choice of N as part of the penalty term in IC were also reviewed and discussed. As MLCMs have been shown to have a wide range of possible applications, this article provides practical guidelines and recommendations to researchers regarding model selection when analyzing empirical data.

ACKNOWLEDGMENTS

We thank Kris Onishi and Ming-Mei Wang for helpful discussion and comments, and we thank Yinan Yu for editorial assistance.

FUNDING

This research was supported by National Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants PGPIN 391334.

REFERENCES

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.

Aitkin, M., & Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, Series B*, *47*, 67–75.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Andrews, R. L., & Currim, I. S. (2003). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing*, *20*, 315–322.

Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance- structure models. *Psychometrika*, *64*, 475–494.

Asparouhov, T., & Muthén, B. (2008). Multilevel mixture models. In G. R. Hancock & K. M. Samuelson (Eds.), *Advances in latent variable mixture models* (pp. 27–51). Charlotte, NC: Information Age.

Bennett, N., & Jordan, J. (1975). A typology of teaching styles in primary schools. *British Journal of Educational Psychology*, *45*, 20–28.

Bijmolt, T. H. A., Paas, L. J., & Vermunt, J. K. (2004). Country and consumer segmentation: Multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing*, *21*, 323–340.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370.

Bozdogan, H. (1993). Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-fisher in-formation matrix. In O. Opitz, B. Lausen, & R. Klar (Eds.), *Information and classification* (pp. 218–234). Heidelberg, Germany: Springer.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*, 195–212.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences*. New York, NY: Plenum Press.

Collins, L. M., Fidler, P. L., Wugalter, S. E., & Long, J. D. (1993). Goodness-of-fit testing for latent class models. *Multivariate Behavioral Research*, *28*, 375–389.

Di, C. Z., & Bandeen-Roche, K. (2011). Multilevel latent class models with dirichlet mixing distribution. *Biometrics*, *67*, 86–96.

Dias, J. M. G. (2004). *Finite mixture models: Review, applications, and computer-intensive methods*. Groningen, The Netherlands: University of Groningen, Research School Systems, Organization and Management.

Dias, J. G., & Vermunt, J. K. (2008). A bootstrap-based aggregate classifier for model-based clustering. *Computational Statistics*, *23*, 643–659.

Dolan, C. V., & Van Der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*, 227–253.

Everitt, B. S. (1981). A Monte Carlo investigation of the likelihood ratio test for number of components in a mixture of normal distributions. *Multivariate Behavioral Research*, *16*, 171–180.

Everitt, B. S. (1988). A Monte Carlo investigation of the likelihood ratio test for number of classes in latent classes analysis. *Multivariate Behavioral Research*, *23*, 531–538.

Fischer, B. B., & Fischer, L. (1979). Styles in teaching and learning. *Educational Leaderships*, *36*, 245–254.

Fonseca, J. R. S., & Cardoso, M. G. M. S. (2007). Mixture-model cluster analysis using information theoretical criteria. *Intelligent Data Analysis*, *11*, 155–173.

Gonzalo, J., & Pitarakis, J. Y. (1996). *Lag length estimation in large dimensional systems* (Tech. Rep.). Reading, UK: University of Reading, Department of Economics.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis models*. Cambridge, UK: Cambridge University Press.

Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling*, *17*, 193–215.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling*, *14*, 202–226.

Jung, C. G. (1971). *Psychological types (Collected works of C. G. Jung)* (Vol. 6). R. F. Hull (Ed.): London, UK: Routledge and Kegan Paul.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.

Lin, T. H., & Dayton, C. M. (1997). Model selection information criteria for nonnested latent class models. *Journal of Educational and Behavioral Statistics*, *22*, 249–264.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767–778.

Lukočienė, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology*, *40*, 247–283.

Lukočiené, O., & Vermunt, J. K. (2010). Determining the number of components in mixture models for hierarchical data. In A. Fink, L. Berthold, W. Seidel, & A. Ultsch (Eds.), *Advances in data analysis, data handling and business intelligence* (pp. 241–249). Berlin-Heidelberg, Germany: Springer.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 175–198). Thousand Oaks, CA: Sage.

McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics, 36*, 318–324.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463–469.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535–569.

Palardy, G., & Vermunt, J. K. (2010). Multilevel growth mixture models for classifying groups. *Journal of Educational and Behavioral Statistics*, *35*, 532–565.

Read, T. R. C., & Cressie, N. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York, NY: Springer-Verlag.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333–343.

Soromenho, G. (1993). Comparing approaches for testing the number of components in a finite mixture model. *Computational Statistics*, *9*, 65–78.

Sternberg, R. J. (1998). Mental self-government: A theory of intellectual styles and their development. *Human Development*, *31*, 197–224.

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in a growth mixture model. In G. R. Hancock (Ed.), *Mixture models in latent variable research* (pp. 317–341). Greenwich, CT: Information Age.

Varriale, R., & Vermunt, J. K. (2012). Multilevel mixture factor models. *Multivariate Behavioral Research*, *47*, 247–275.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, *33*, 213–239.

Vermunt, J. K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, *58*, 220–233.

Vermunt, J. K., & Magidson, J. (2008). *LG-Syntax user's guide: Manual for Latent GOLD 4.5 syntax module*. Belmont, MA: Statistical Innovations.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica*, *57*, 307–333.

Yang, C. C. (1998). *Finite mixture model selection with psychometrics applications* (Doctoral dissertation). University of California, Los Angeles.

Yang, C. C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis*, *50*, 1090–1104.

Yang, C. C., & Yang, C. C. (2007). Separating latent classes by information criteria. *Journal of Classification*, *24*, 183–203.

APPENDIX

Specifications of the Conditional Latent Class Probabilities (π_{ml}) for More and Less Distinct Clusters.
More distinct clusters

	L = 2	L = 3	L = 4	L = 5
M = 2	$\begin{bmatrix} .9 & .1 \\ .1 & .9 \end{bmatrix}$			
M = 3	$\begin{bmatrix} .8 & .1 \\ .1 & .1 \\ .1 & .8 \end{bmatrix}$	$\begin{bmatrix} .8 & .1 & .1 \\ .1 & .1 & .8 \\ .1 & .1 & .8 \end{bmatrix}$		
M = 4	$\begin{bmatrix} .7 & .1 \\ .1 & .1 \\ .1 & .1 \\ .1 & .7 \end{bmatrix}$	$\begin{bmatrix} .7 & .1 & .1 \\ .1 & .7 & .1 \\ .1 & .1 & .1 \\ .1 & .1 & .7 \end{bmatrix}$	$\begin{bmatrix} .7 & .1 & .1 & .1 \\ .1 & .7 & .1 & .1 \\ .1 & .1 & .7 & .1 \\ .1 & .1 & .1 & .7 \end{bmatrix}$	
M = 5	$\begin{bmatrix} .6 & .1 \\ .1 & .1 \\ .1 & .1 \\ .1 & .1 \\ .1 & .6 \end{bmatrix}$	$\begin{bmatrix} .6 & .1 & .1 \\ .1 & .1 & .1 \\ .1 & .6 & .1 \\ .1 & .1 & .1 \\ .1 & .1 & .6 \end{bmatrix}$	$\begin{bmatrix} .6 & .1 & .1 & .1 \\ .1 & .6 & .1 & .1 \\ .1 & .1 & .1 & .1 \\ .1 & .1 & .1 & .6 \\ .1 & .1 & .1 & .6 \end{bmatrix}$	$\begin{bmatrix} .6 & .1 & .1 & .1 & .1 \\ .1 & .6 & .1 & .1 & .1 \\ .1 & .1 & .6 & .1 & .1 \\ .1 & .1 & .1 & .6 & .1 \\ .1 & .1 & .1 & .1 & .6 \end{bmatrix}$

Less distinct clusters

	L = 2	L = 3	L = 4	L = 5
M = 2	$\begin{bmatrix} .85 & .15 \\ .15 & .85 \end{bmatrix}$			
M = 3	$\begin{bmatrix} .7 & .15 \\ .15 & .15 \\ .15 & .7 \end{bmatrix}$	$\begin{bmatrix} .7 & .15 & .15 \\ .15 & .7 & .15 \\ .15 & .15 & .7 \end{bmatrix}$		
M = 4	$\begin{bmatrix} .55 & .15 \\ .15 & .15 \\ .15 & .55 \\ .15 & .55 \end{bmatrix}$	$\begin{bmatrix} .55 & .15 & .15 \\ .15 & .55 & .15 \\ .15 & .15 & .15 \\ .15 & .15 & .55 \end{bmatrix}$	$\begin{bmatrix} .55 & .15 & .15 & .15 \\ .15 & .55 & .15 & .15 \\ .15 & .15 & .55 & .15 \\ .15 & .15 & .15 & .55 \end{bmatrix}$	
M = 5	$\begin{bmatrix} .40 & .15 \\ .15 & .15 \\ .15 & .15 \\ .15 & .15 \\ .15 & .40 \end{bmatrix}$	$\begin{bmatrix} .40 & .15 & .15 \\ .15 & .15 & .15 \\ .15 & .40 & .15 \\ .15 & .15 & .15 \\ .15 & .15 & .40 \end{bmatrix}$	$\begin{bmatrix} .40 & .15 & .15 & .15 \\ .15 & .40 & .15 & .15 \\ .15 & .15 & .15 & .15 \\ .15 & .15 & .40 & .15 \\ .15 & .15 & .15 & .40 \end{bmatrix}$	$\begin{bmatrix} .40 & .15 & .15 & .15 & .15 \\ .15 & .40 & .15 & .15 & .15 \\ .15 & .15 & .40 & .15 & .15 \\ .15 & .15 & .15 & .40 & .15 \\ .15 & .15 & .15 & .15 & .40 \end{bmatrix}$

Specifications of Conditional Response Probabilities (ρ_{mj}) for More and Less Distinct Classes.

	More distinct classes	Less distinct classes
M = 2	$\begin{bmatrix} .8 & .8 & .8 & .8 & .8 & .8 & .8 & .8 \\ .2 & .2 & .2 & .2 & .2 & .2 & .2 & .2 \end{bmatrix}$	$\begin{bmatrix} .7 & .7 & .7 & .7 & .7 & .7 & .7 & .7 \\ .3 & .3 & .3 & .3 & .3 & .3 & .3 & .3 \end{bmatrix}$
M = 3	$\begin{bmatrix} .8 & .8 & .8 & .8 & .8 & .8 & .8 & .8 \\ .2 & .2 & .2 & .2 & .8 & .8 & .8 & .8 \\ .2 & .2 & .2 & .2 & .2 & .2 & .2 & .2 \end{bmatrix}$	$\begin{bmatrix} .7 & .7 & .7 & .7 & .7 & .7 & .7 & .7 \\ .3 & .3 & .3 & .3 & .7 & .7 & .7 & .7 \\ .3 & .3 & .3 & .3 & .3 & .3 & .3 & .3 \end{bmatrix}$
M = 4	$\begin{bmatrix} .8 & .8 & .8 & .8 & .8 & .8 & .8 & .8 \\ .2 & .2 & .2 & .2 & .8 & .8 & .8 & .8 \\ .8 & .8 & .8 & .8 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 & .2 & .2 & .2 \end{bmatrix}$	$\begin{bmatrix} .7 & .7 & .7 & .7 & .7 & .7 & .7 & .7 \\ .3 & .3 & .3 & .3 & .7 & .7 & .7 & .7 \\ .7 & .7 & .7 & .7 & .3 & .3 & .3 & .3 \\ .3 & .3 & .3 & .3 & .3 & .3 & .3 & .3 \end{bmatrix}$
M = 5	$\begin{bmatrix} .8 & .8 & .8 & .8 & .8 & .8 & .8 & .8 \\ .2 & .2 & .2 & .2 & .8 & .8 & .8 & .8 \\ .2 & .2 & .2 & .8 & .8 & .2 & .2 & .2 \\ .8 & .8 & .8 & .8 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 & .2 & .2 & .2 \end{bmatrix}$	$\begin{bmatrix} .7 & .7 & .7 & .7 & .7 & .7 & .7 & .7 \\ .3 & .3 & .3 & .3 & .7 & .7 & .7 & .7 \\ .3 & .3 & .3 & .7 & .7 & .3 & .3 & .3 \\ .7 & .7 & .7 & .7 & .3 & .3 & .3 & .3 \\ .3 & .3 & .3 & .3 & .3 & .3 & .3 & .3 \end{bmatrix}$

Note. The same pattern is doubled when item = 12.