The Pennsylvania State University

The Graduate School

Department of Educational and School Psychology and Special Education

CLASS EXTRACTION AND CLASSIFICATION ACCURACY IN LATENT CLASS MODELS

A Dissertation in

Educational Psychology

by

Qiong Wu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2009

The dissertation of Qiong Wu was reviewed and approved* by the following:

Pui-Wa Lei
Associate Professor of Education
Dissertation Advisor
Chair of Committee

Hoi K Suen
Distinguished Professor of Educational Psychology

Jonna Kulikowich
Professor of Education

Aleksandra Slavkovic
Assistant Professor of Statistics

Rayne Sperling
Associate Professor of Education
Professor-in-Charge of Educational Psychology

*Signatures are on file in the Graduate School

ABSTRACT

Despite the increasing popularity of latent class models (LCM) in educational research, methodological studies have not yet accumulated much information on the appropriate application of this modeling technique, especially with regard to requirement on sample size and number of indicators. This dissertation study represented an initial attempt to inform practitioners of desirable sample sizes and number of indicators under various conditions by using simulated data. The performance of LCM was evaluated in terms of both correct class extraction rates (i.e., correct identification of number of latent classes) and classification accuracy for dichotomous and continuous indicators. Manipulated factors included class separation, true number of latent classes, sample size, number of indicators, and class proportion.

For both dichotomous and continuous indicators, class separation and true number of latent classes consistently demonstrated the largest effects on both correct class extraction rates and classification accuracy. The results suggest that LCM may not be recommended for use when class separation is small (operationally defined as .20 difference in indicator means for dichotomous indicators and half a standard deviation difference in indicator means for continuous indicators). When class separation is large (operationally defined as .50 difference in dichotomous indicator means and 2 standard deviation differences in continuous indicator means), models show acceptable correct class extraction rates and classification accuracy when certain conditions on sample size and number of indicators are met. Recommendations on sample size and number of indicators are provided taking into account researchers' level of knowledge of class separation and true number of latent classes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

I would like to thank my academic and dissertation advisor Dr. Pui-Wa Lei for her guidance, support, and friendship during my graduate study. She has been the source and inspiration of my intellectual exploration. She is passionate in sharing her research ideas with me, yet also mindful of training me to be an independent researcher. She holds strong belief in my capacity and has encouraged me to take challenging courses, which proves to be very helpful for my research.

I am indebted to Dr. Hoi Suen for his unwavering support since the first day I knew him. I still remember his unreserved encouragement after my first class presentation in the U.S., which, together with his other inspiring words, has been a great source of motivation for me during my graduate school. I have also benefited tremendously from his excellent analytical ability and amazing presentation skills which add fun to abstract psychometric concepts and models.

I also want to thank Dr. Kulikowich and Dr. Slavkovic for their helpful comments on my dissertation study. Dr. Kulikowich's deep concern for linking research to practice has greatly enhanced the practical implication of the current study. Dr. Slavkovic foresaw a key limitation of the study during the proposal meeting, and prevented me from wasting a lot of time.

My thanks also go to other professors that I have worked closely with during the years, including Drs. James DiPerna, Paul Morgan, and George Farkas. They have provided me with invaluable chances to obtain hands-on research experiences. They treat me with patience and respect. I feel very fortunate to have worked with them.

I also want to thank my husband for his emotional support and his tolerance of my occasional cranky mood. His understanding and love have turned life at the graduate school a sweet memory.

Finally, I owe everything I am today to my parents. Ever since I was a little girl, they have never hesitated to make sacrifices for me. As the only child in my family, I left my hometown for college when I was eighteen, and then went to the other side of the planet for my graduate study. My parents were as supportive as they can be for those decisions, and spared no efforts to make me worry-free. I feel guilty that I cannot spend more time with them.

Chapter 1

Introduction

Motivation of the study

Latent class models (LCM) comprise a subset of the general latent structure models (Clogg, 1995). The family tree of latent structure models has three big branches: the structural equation models (SEM: Bollen, 1989; Kaplan, 2000) dealing with observed variables and continuous latent variables; the latent trait models (also called Item Response theory models or IRT: Lord, 1980) dealing with one or more continuous latent variables and a set of locally independent categorical responses; and latent class models (LCM) dealing with observed variables and categorical latent variables (Clogg; Lazarsfeld & Henry, 1968; Goodman, 1974). Both SEM and IRT have been widely used in the field of educational measurement to validate latent constructs or for scaling purposes, and numerous methodological studies have contributed to improving the quality of their uses. This quality is maintained by practitioners following guidelines on the proper way of using those models (e.g., guidelines on desirable sample sizes, number of indictors or items, and report of fit statistics).

LCM was not as frequently utilized as the other two models, but in recent years we have witnessed a rapid increase in its applications in educational measurement. A rough search of papers published since 2000 from two databases (i.e., Educational Resources Information Center: ERIC and PsycINFO) revealed that LCM has been used for classification (e.g., Reinke, Herman, Petras, & Ialongo, 2008), diagnostic assessment (e.g., Templin & Henson, 2006), detecting aberrant response patterns (e.g., van den Wittenboer, Hox, & de Leeuw, 1997), detecting rules or strategies (e.g., DeCarlo, 2005; Jansen & van der Maas, 2002), differential item functioning (e.g., Webb, Cohen, & Schwanenflugel, 2008), and scaling (e.g., Liao & Tu, 2006). Such a rapid increase in its applications, however, has not been kept up by methodological studies that provide guidelines to inform its proper use in various applications. Unlike SEM or IRT, no "rules of thumb" are available on desirable sample sizes and number of indicators in LCM. As Muthén (2003) has pointed out, LCM has progressed from the phase of "toy applications" to the phase of serious substantive applications; thus scrutiny of this technique is needed to protect against poor applications and promote good uses of it.

LCM has been used both with relatively small samples for theory forming or conforming and with relatively large samples for scaling purposes or policy related studies. For example, Davey and Macready (1990) utilized LCM to test a few hypotheses regarding the structure underlying reading comprehension with a relatively small sample of 170 children, and the authors found their results to be in favor of a holistic view of reading comprehension rather than a subskill view. On the other hand, LCM has also been applied to large samples to address important decision making questions for scaling purposes. For example, Liao and Tu (2006) used LCM to examine the scalability of intimacy permissiveness scale (IPS) with a sample of around 1,800 children. Using the latent distance model (McCutcheon, 1987), the authors found that the IPS were able to form reasonable Guttman-type scale after accounting for errors of measurement.

Despite the growing popularity of LCM in educational research, not much is discussed about whether the model is implemented appropriately. In order to examine current practices with LCM, a search was conducted of peer reviewed empirical studies published since 2000 using LCM. Two databases (i.e., ERIC and PsycINFO) frequently accessed by educational researchers served as sources for the search. Key words included latent class models, latent class analysis and latent profile analysis. These studies were reviewed with regard to their sample sizes, number of indicators, scales of indicators, number of extracted latent classes, estimated class proportions, and reported fit indices. Table 1-1 lists the characteristics of 43 studies from these two databases.

Most of the studies in Table 1-1 focused on constructs related to students' academic development or their behavioral and emotional problems. Other studies focused on social interactions (e.g., marital happiness, adult-child relations) or clinical issues (e.g., Attention Deficit Hyperactivity Disorder). As expected, there is a great variety in the practices. Sample sizes range from as small as 37 to as large as over 8,000, with 10% of the studies having sample sizes smaller than 200 and 40% having sample sizes within 500. Total sample sizes in those studies do not seem very problematic if we consider the recommended sample sizes in SEM, but variation in class sizes seems to be more questionable. Among the 43 studies, the smallest estimated class size is only 5, with 10% of the studies having class sizes smaller than 10; more than half of the studies have class sizes smaller than 50. Since the estimation of class-specific parameters such as mean and variance is based on individuals within a certain class, class size should be given due consideration in addition to total sample size. Similarly, number of indicators from those studies represents a wide range from 2 to 24. Extremely small number of indicators (i.e., 2 or 3) is present in 13% of the studies, and half of the studies have 6 or fewer indicators. Number of extracted latent classes ranges from 2 to 7, with 70% of them having 4 or fewer classes. Figure 1-1 presents histograms of the frequencies for sample sizes, number of indicators and number of extracted latent classes from the 43 studies.

Table 1-1. Summary of Empirical Studies Using LCM from ERIC and PsycINFO since 2000

	First Author	Year	Sample size	Number of indicators	Scale of indicator	Number of classes	Class proportion	Fit indices
1	Reinke	2008	362 & 314	5	Dichotomous	4 & 3	11%-63%	BIC, entropy
2	Tapola	2008	208	3	Continuous	4	10%-37%	BIC, CAIC
3	Dush	2008	1,998	6	Continuous	3	22%-41%	chi-square, BIC
4	Klonsky	2008	205	12	Unclear	4	11%-61%	BIC, entropy
5	Nylund	2007	1,500+	6	Dichotomous	3	6%-69%	BIC & BLRT
6	Chan	2007	2,612	23	Dichotomous	5	9%-36%	AIC, BIC & LMRT
7	Bouwmeester	2007	615	15	Dichotomous	3	25%-39%	BIC
8	Lopez	2007	128	11	Dichotomous	3	13%-61%	BIC & entropy
9	Notenboom	2007	458	7	Dichotomous	5	4%-45%	Likelihood ratio, chi-square
10	van Gaalen	2006	4,990	8	Dichotomous	5	4%-40%	chi-square, BIC
11	Guttentag	2006	241	4	Continuous	4	6%-39%	AIC, moments
12	Liao	2006	1,821	4	Dichotomous	6	8%-27%	chi-square
13	Romano	2006	252	5	Dichotomous	2	21%-79%	chi-square, BIC, residual
14	Geiser	2006	1,724	24	Dichotomous	5	5%-37%	AIC, BIC, chi-square, BLRT, LMR
15	DiStefano	2006	1,228	14	Continuous	3	22%-42%	BIC, entropy
16	Chung	2005	501	10	Dichotomous	3	1%-99%	BIC
17	MacMillan	2005	2,191	4	Dichotomous	2 & 3	9%-89%	Likelihood ratio chi-square, BIC
18	Volk	2005	1,600	19	Dichotomous	7 & 5	3%-53%	BIC
19	DeCarlo	2005	125	8	4-point scale	5	6%-39%	AIC, BIC
20	Yang	2005	3,289	12 & 4	Dichotomous	3 & 2	21%-79%	BIC, chi-square
21	Loken	2004	573	4	3-point scale	3	28%-36%	chi-square
22	Raijmakers	2004	226 & 328	6	Dichotomous	2 & 3	28%-57%	BIC
23	Jansen	2002	805	25	3-point scale	3 & 4	2%-27%	BIC, chi-square
24	DeCarlo	2000	294	10	Unclear	3	24%-44%	BIC
25	Aldridge	2008	354	13	Continuous	3	7%-48%	LMRT, BIC, AIC, entropy
26	Herman	2007	423	7	Continuous	5	10%-32%	BIC, LMRT, BLRT, entropy
27	Bowen	2007	523	8	Continuous	5	>6%	BIC, entropy
28	Macy	2007	415	4	Continuous	4	9%-52%	BIC,LMRT, entropy
29	Pastor	2007	1,868	2 & 3 & 4	Continuous	5 & 5 & 6	2%-44%	BIC, LMRT
30	Hill	2006	383	3	Continuous	4	>11%	BIC
31	Oxford	2005	227	7	Continuous	3	15%-43%	BIC
32	Whiteman	2006	384	4	mixed	4	17%-34%	AIC, BIC
33	Buckley	2005	37	3	Continuous	2	35%-65%	

Table 1-1 Continued.

	First Author	Year	Sample size	Number of indicators	Scale of indicator	Number of classes	Class proportion	Fit indices
34	D'Angiulli	2004	1,108	6	continuous	3	3%-62%	BIC
35	Shaw	2003	284	5	continuous	4	6%-43%	BIC
36	Webb	2008	210	24	dichotomous	2	40%-60%	AIC, BIC
37	Witkiewitz	2008	563	7	continuous	3	6%-82%	ABIC, BLRT, entropy
38	Acosta	2008	107 & 269& 634	18	dichotomous	5 & 6 & 7	7%-29%	chi-square
39	Striegel	2008	8,250	9	Dichotomous	4	14%-62%	BIC,entropy
40	Jackson	2008	3,720	8	dichotomous	5	3%-68%	BIC, entropy
41	Park	2008	2,376	6	Dichotomous	4	2%-64%	AIC, BIC, entropy
42	Reiersen	2008	851	18	continuous	7	1%-27%	BIC
43	Chung	2008	214	21	dichotomous	3	27%-39%	BIC

Note. Studies listed in this table are marked with * in the reference list.

Another source of variation in practices exists among the studies' choices of model fit indices. Similar to SEM, a set of different fit indices is available in LCM. Those fit indices are discussed in details in Chapter 2, but they can be roughly divided into 2 major categories: relative fit index for comparison of alternative models, and absolute fit index focusing on the fit of a particular model to data. Studies often report multiple fit indices, but absolute fit index is rarely reported in the studies reviewed possibly due to two reasons. First, many studies use LCM for exploratory purposes, and the focus is to determine the optimal number of latent classes. Under such circumstance, relative fit indices fit the purpose better. Second, unlike in SEM, variances and covariances are not sufficient statistics for model estimation in LCM when indicators are continuous. Instead, the less frequently discussed higher-order moments (i.e., skewness and kurtosis) have to be examined for absolute goodness-of-fit (Muthén, 2008).

Although the above review of LCM studies is by no means an exhaustive investigation, it has already revealed some interesting patterns of current practices with LCM. First, while users of SEM or IRT are generally aware of the minimal requirement on sample size and number of indicators (items) under different model conditions, users of LCM do not seem to be as concerned perhaps because few research studies have provided an explicit recommendation for LCM. As a result, sample sizes smaller than 50 and class sizes of 5 are used without further consideration. Second, although fit indices are frequently reported in LCM studies, a majority of them only report relative fit indices to support the choice of a model over other worse models; absolute fit indices are rarely used or reported.

Since there is a lack of standards or guidelines on the proper use of LCM, one question would be: does it matter? Does LCM produce poorer estimates or classification if insufficient sample sizes or number of indicators are used under different conditions? Based on existing studies examining fit indices in LCM, the answer is unfortunately yes, and many factors can affect the accuracy of class extraction in LCM.

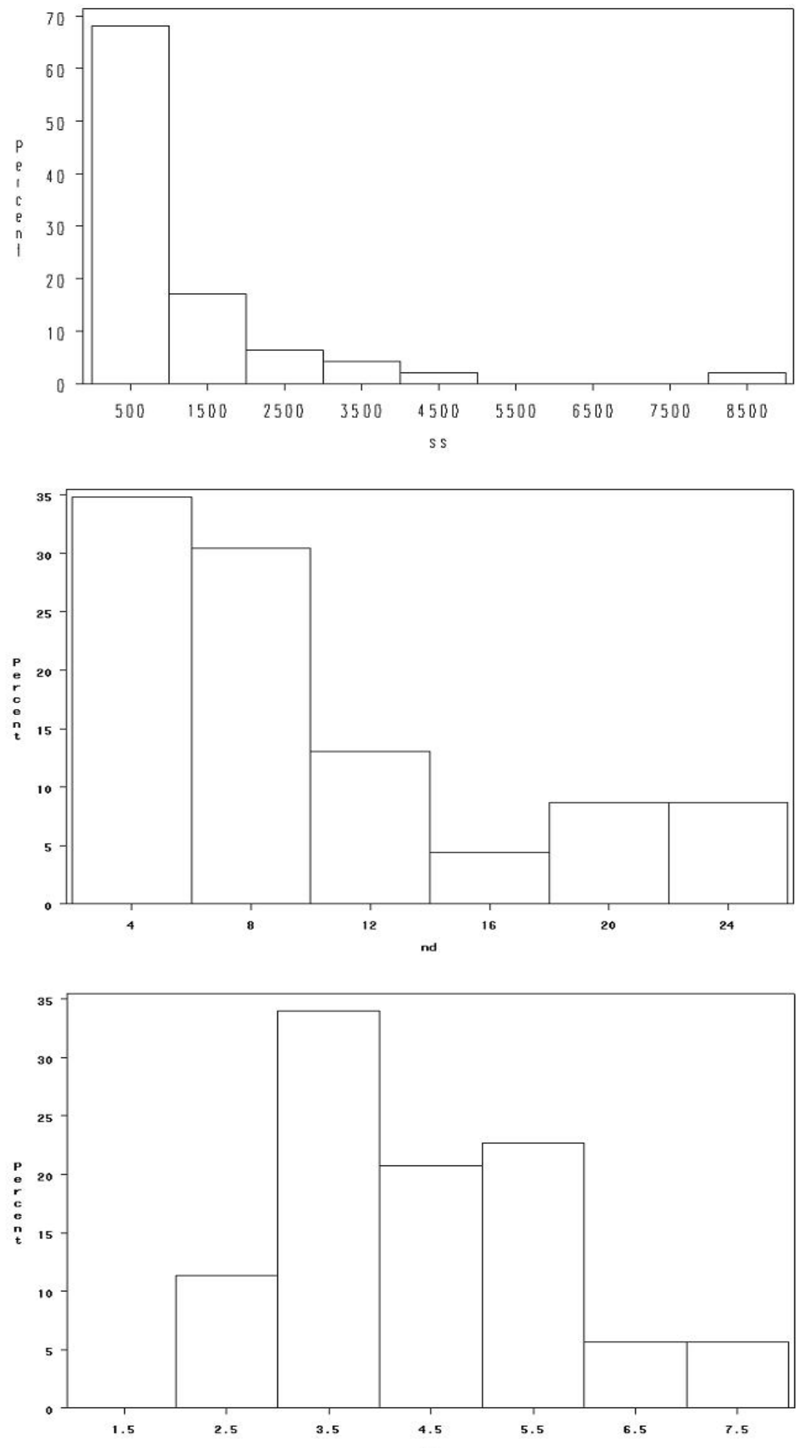

Figure 1-1. Histogram of sample size (ss), number of indicators (nd), and number of extracted latent classes (nc)

Factors affecting the performance of LCM

Although existing studies have so far focused on evaluating the performance of LCM in terms of its ability to correctly identify the number of latent classes, performance of LCM defined in this dissertation adds another component: the accuracy of assigning each individual into the appropriate class (i.e., classification accuracy). This second aspect is also considered in this study because classification is a very important application of LCM (Clogg, 1995), and its accuracy is not necessarily predictable from the correct identification of number of latent classes.

Similar to SEM and IRT or any other statistical techniques, the performance of LCM depends on the information afforded by data. In general, the more information the data (in terms of both the number of subjects and number of variables) can provide, the less complicated the true model structure is, the more accurate the model estimation tends to be. This section introduces various factors that are known to affect model estimation and hence model performance. Possible effects of those factors on the performance of different model fit indices are also discussed.

Model complexity

In parametric models, model complexity can be indicated by number of free parameters in a specified model; given a set of observed variables, a model is more complicated if there are more parameters to be estimated. Model complexity has been found in both SEM and IRT to affect model estimation when sample size is not very large (Cudeck & Henley, 1991; Stone & Yumoto, 2004). It is expected to have a similar effect on model estimation in LCM.

For LCM with dichotomous indicators, free parameters include class proportions and class-specific indicator means. For a C-class model with p indicators, there are C-1 class proportions to estimate since all class proportions should sum to 1. Within each class, there are *p* indicator means (or item probabilities) when indicators are independent of one another within each class. The total number of free parameters *l* is simply a sum of the above two terms, and is expressed as,

$$l = C - 1 + C \times p \tag{1.1}$$

For LCM with continuous indicators, free parameters include class proportion, class-specific means, variances and covariances. For a *C*-class model with *p* indicators, there are *C-1* class proportions to estimate. Within each latent class, there are *p* means and *p* variances to estimate. Furthermore, if within each class, indicators are allowed to correlate with one another (i.e., local independence is not assumed), there are *p*(*p*-1)/2 covariance terms to estimate. Therefore, the most general model of LCM with continuous indicators without covariates has *l* free parameters, and *l* is expressed as,

$$l = C - 1 + C \times [2p + p(p-1)/2] \tag{1.2}$$

Number of free parameters can be significantly reduced by imposing certain restrictions on the model (e.g., fixing covariance terms to be zero, or setting equality constraints on the covariance structures across latent classes), and it can also be increased by adding other components into the model (e.g.., covariates). The most frequently adopted restriction in LCM is setting the within-class covariances to be zero. In other words, observed variables are assumed to be independent of one another within latent classes (Vermunt & Magidson, 2002). This restriction is also termed as local independence in latent variable literature. Based on the above equations, the following characteristics directly affect model complexity.

Number of latent classes. Based on Equations 1.1 and 1.2, number of latent classes (C) has a multiplicative effect on the number of free parameters, which suggests a possibly increasing challenge in model estimation with more latent classes given the specific amount of data. Existing studies also noted more difficulty in class extraction even when resources (e.g., number of indicators) were proportionally increased with more latent classes (Yang, 2006). Yang studied LCM with dichotomous responses, and found that holding other factors constant, models with more latent classes were relatively more difficult to identify correctly than models with fewer latent classes based on most fit indices, and this was true even when the number of indicators was increased proportionally to the number of true latent classes. Yang suggested that the decreasing performance may be due to the fact that given the same sample sizes, fewer subjects were available for each latent class in a larger number of latent classes than in a smaller number of latent classes. Wu and Lei (2008) studied LCM for continuous indicators, and have found that some indices (e.g., BIC, Raftery, 1995, sample size adjusted BIC, Sclove, 1987, and bootstrapped likelihood ratio test, BLRT, McLachlan & Peel, 2000) appeared to be more robust to the increase in number of latent classes than other indices (i.e., AIC, Akaike, 1987; and Vuong-Lo-Mendell-Rubin test: LMR, Vuong, 1989). Those fit indices are introduced in details in Chapter 2.

Number of indicators. Increase in number of indicators introduces two competing forces in model estimation: on the one hand, more indicators bring in more information and it should contribute to an increased likelihood of getting the true model; on the other hand, more indicators also mean more free parameters to estimate because number of free parameters is positively correlated with number of indicators (p) based on Equations 1.1 and 1.2. Furthermore, adding indicators does not necessarily bring in extra useful information to distinguish among classes because additional indicators do not necessarily differ significantly among classes. Therefore, it is not clear how model performance may be affected by adding indicators. Number of indicators was explicitly studied in Wu and Lei (2008) with continuous responses, and they reported a notable increase in correct class extraction rates with the increase of number of indicators from 3 to 6, especially at sample sizes of 500 or below.

Sample size, class proportion and class size

Sample size is directly related to the amount of information available to the user. This factor has been studied in LCM related methodological studies, and larger sample size has been found to yield higher correct class extraction rates (Nylund, et al., 2007; Wu & Lei, 2008; Yang, 2006). Holding other factors constant, Yang found sample sizes to have a significant effect on BIC, but not sample size adjusted BIC (ABIC); this pattern, however, is not observed in Nylund et al. Wu & Lei also found a more significant effect of sample size on ABIC than on BIC.

Another frequently studied factor in existing studies is class proportion. Both Nylund et al. (2007) and Wu and Lei (2008) reported higher correct class extraction rates with equal class proportions than with unequal class proportions. However, it is unclear so far whether the decreasing performance of unequal class proportions was attributable to the unequal class proportions per se or to the resulting smaller class size.

Class separation

Class separation represents differences in class-specific indicator means across different latent classes. It is expected that the more separated or more distant classes are from one another, the easier it is to correctly extract them. This factor was explicitly studied in Tofighi and Enders (2007), who evaluated different fit indices in extracting latent growth trajectories using growth mixture models (GMM: Muthén, 2001). GMM is a special form of LCM, where each latent class is described by a distinct growth trajectory represented by distinct intercept and slope. The authors noted higher correct class extraction rates with larger class separation.

Limitations of existing studies on the performance of LCM

Although previous studies have generated some consistent messages on the effects of different factors on model performance in LCM, we could not draw from those studies explicit guidelines similar to those provided in SEM or IRT. The following section discusses three limitations in the previous research that this study intends to address.

First of all, evaluation of model performance tends to focus more on identifying the correct number of latent classes, and less on other issues such as classification accuracy. It is without dispute that deciding on the number of latent classes is one of the most important questions in LCM. However, other important issues remain largely unattended, such as correct interpretation of latent classes and classification accuracy. Considering that an important potential application of LCM is for classification, its classification accuracy should be examined closely. Clogg (1995) explicitly stated that the quality of the prediction of individuals’ class membership provided by LCM is a very important issue. The study of classification accuracy in LCM may also help promote the use of LCM as a useful tool for classification because correctly assigning individuals from smaller groups has been proven to be a very challenging task (e.g., Fan & Wang, 1999; Finch & Schneider, 2006). This is the case even for the so-called supervised classification method (i.e., where group membership is observed) such as discriminant analysis and logistic regression (Vermunt & Magidson, 2002). Previous studies on supervised classification methods reported very high error rates for the smaller group: for example, at sample sizes of 300, with 3 indicators where indicator mean differences across groups are .5 standard deviation, Finch and

Schneider reported group error rates from logistic regression for the smaller group to be higher than .95 with class proportions of 10% and 90%; in contrast, group error rates for the larger group were lower than .01. Vermunt and Magidson found in their simulation study that LCM showed slightly higher classification accuracy rates than logistic regression, however, their study only involved equal class proportion conditions. It would be of great interest to understand how LCM performs under unequal class sizes.

Second, although existing studies (e.g., Nylund et al., 2007; Wu & Lei, 2008; Yang, 2006) have found that larger sample sizes, and more indicators contribute to enhanced performance of LCM, this information is not detailed enough to inform choices of sample size and number of indicators. As a result, users of LCM lack references to guide their experiment or analysis. Such guidelines have been established, although not without controversy, in SEM and IRT (e.g., Kline, 2005), where users are more vigilant towards their uses of the latter two methods. It is important that we study more detailed levels of sample sizes and number of indicators that are often observed in empirical studies to accumulate information for best practices with LCM. In addition, the issue of class size was briefly mentioned in Yang (2006) but has never been examined closely; unequal class proportion seems to have adverse effects on correct class extraction rates controlling for total sample sizes, but it is not clear whether this decreasing performance is due to the reduced class size of the smaller class or the unequal proportion per se, and studies involving detailed levels of total sample sizes will facilitate the understanding of this issue. These two possible causes of the decreasing performance have different implications for LCM: if the decreasing performance is due to the reduced class size, researchers may increase their total sample sizes to make sure that the smaller class has sufficient cases; but if the cause is the unequal proportion per se, that means LCM is not capable of detecting classes of small proportions.

The third point concerns model fit indices studied in existing literature. Although a variety of fit indices have been studied in LCM, and there has been some consensus on which indices are better, the choice has been among models with different numbers of latent classes and only relative model fit indices have been studied. Absolute fit indices were not assessed in those studies. This is a reflection of what has been done in empirical studies: rarely did studies check the absolute goodness-of-fit of the model to data, but only relied on relative fit indices to identify the best from the specified models. However, in practice, it is possible that the best model among a set of candidate models still does not fit the data well, in which case relative fit indices are not capable of identifying such model misfit. This concern was noted earlier by Bauer and Curran (2003), and in a comment to that concern, Muthén (2003) proposed that absolute fit index such as multivariate skewness test (MST) and multivariate kurtosis test (MKT) should be used as a complement to the relative fit index for continuous indicators. A poor fit based on the absolute fit indices may motivate researchers to keep looking for alternative structures for the basic model.

Statement of the problem

This dissertation study intends to address a few limitations in existing studies and focuses on the effects of sample sizes and number of indicators on the performance of LCM, not only in terms of the identification of correct number of latent classes (i.e., correct class extraction rates), but also classification accuracy when the correct model is identified. Absolute fit indices are studied in terms of their ability to identify the correct model. The systematic study involving detailed levels of sample sizes and number of indicators is expected to yield important implications for practice.

Sample sizes and number of indicators are not the only factors that affect the performance of LCM. Other factors include class proportions, number of latent classes, and class separation (e.g., Nylund, et al., 2007; Tofighi & Enders, 2007). This study has a special interest in sample size and number of indicators because researchers most likely have some control over these two factors at the design stage of experiments. Other factors such as number of latent classes, class separation, and class proportion are also manipulated in the simulation study to serve as control variables.

The purpose of this dissertation is three-fold: 1) extend existing studies on correct class extraction rates by investigating more fine-grained levels of sample sizes and number of indicators; 2) examine classification accuracy when number of latent classes is correctly identified and the usefulness of entropy as a measure of classification accuracy; 3) examine the usefulness of less frequently studied fit indices (entropy, absolute fit indices).

Chapter 2

Literature Review

This chapter reviews some fundamental concepts in LCM from its basic mathematical model to parameter estimation and model evaluation. Various fit indices are introduced in details in this chapter. Descriptions of study designs and findings are provided for three studies that concentrate on the correct identification of number of latent classes. Factors that have shown to affect model performance in those three studies serve as the foundation of design for this dissertation study.

Fundamentals of LCM

LCM is also known as latent class analysis, latent profile analysis, latent class cluster analysis, mixture models, model-based cluster analysis, and structural equation mixture analysis (Yang et al., 2005). These names are either used interchangeably or referring to a particular subset, or a broader set of models. “Mixture models” and “structural equation mixture analysis” refer to the broadest set of models where the observed probability distribution is a mixture of class-specific probabilities, and no further constraints are imposed on how within-class variability is accounted for (Bauer & Curran, 2003). For example, within-class variability can be further explained by a factor analytical model (i.e., factor mixture model; Lubke & Muthén, 2007) or a growth curve model (i.e., growth mixture model; Muthén, 2002). On the other hand, “latent class analysis,” “latent profile analysis,” “latent class cluster analysis,” and “model-based cluster analysis” often refer to mixture models where the within-class variation is considered random. Some researchers distinguish between latent class analysis and latent profile analysis based on the scale of the indicator variables. The former refers to models with categorical indicators, and the latter refers to models with continuous indicators (Muthén). This dissertation focuses on a subset of the general mixture models where the within-class variability is considered random, and latent class models (LCM) are used to refer to all such models regardless of the scales of their indicator variables.

Two lines of development contribute to the current state of LCM. The first line has been motivated by the efforts to explain relationship or the dependency among the row and column variables of contingency tables (Lazarsfeld & Henry, 1968; Goodman, 1974; Haberman, 1974). After introducing the latent class variable, the apparent relationship between row and column variables is completely explained by that latent variable, so that the row and column variables are independent of each other within the class.

The second line of development focuses on continuous indicators instead, and is often under another label called finite mixture models (McLachlan & Peel, 2000). Finite mixture models can be dated back to 100 years ago when Karl Pearson used it to fit a mixture of two normal probability density functions to detect subspecies among 1000 sampled crabs (McLachlan & Peel). Karl Pearson used the moments-based estimation to analyze the data. Estimation of mixture models was a daunting task before modern computers were available, so it was not until the middle 1970s did this technique begin to play an important role in data analysis (Goodman, 2002).

The following sections present fundamental concepts in LCM including its mathematical formulation, assumptions, model estimation and model evaluation.

Mathematical models and assumptions

Suppose $\mathbf{Y}_{1,\,\ldots,}\,\mathbf{Y}_n$ denote a random sample of size *n*, where $\mathbf{Y}_i$ is a *p*-dimensional random vector with probability density function $f(\mathbf{y}_i)$ on R^p. LCM states that the density $f(\mathbf{y}_\mathbf{i})$ of $\mathbf{Y}_i$ can be written in the form of

$$f(\mathbf{y}_\mathbf{i}) = \sum_{c=1}^{C} \pi_{ic} f_c(\mathbf{y}_i)\,, \tag{2.1}$$

where $\mathbf{y}_i$ denotes the response vector for individual *i* on *p* indicators, and $f_c(\mathbf{y}_i)$ is the density function for individual *i* in class *c*, and π_{ic} are the mixing weights that sum to one across different latent classes (McLachlan & Peel, 2000). $f_c(\mathbf{y}_i)$ is called the component density of the mixture.

Equation 2.1 is the basic LCM model, and it expresses the observed distribution of **Y** as a mixture of class-specific densities. Extensions of Equation 2.1 include adding covariates ($\mathbf{x}_i$) to predict class membership or/and to allow the covariates to have direct impacts on the indicators ($\mathbf{y}_i$). The extended model with covariates is expressed in Equation 2.2,

$$f(\mathbf{y}_i \mid \mathbf{x}_i) = \sum_{c=1}^{C} \pi_{ic|\mathbf{x}_i} f_c(\mathbf{y}_i \mid \mathbf{x}_i)\,, \tag{2.2}$$

where $\mathbf{x}_i$ is the vector of covariates for subject *i* (Vermunt & Magidson, 2002).

Scales of **Y** can be categorical, continuous, or both (i.e., mixed format). Different forms of probability density functions can be assumed depending on the scales of **Y**. For example, if **Y** are continuous, class-specific probability function is often assumed to be multivariate normal, possibly after applying an appropriate non-linear transformation (Banfield & Raftery, 1993). The most general multivariate normal distribution of which all the restricted versions are special cases is the one with mean vectors μ_c and covariance matrices $\sum_c$. If no further restrictions are imposed, LCM analysis involves estimating a set of means, variances and covariances for each latent class *c*. In most applications, the main objective is to find classes that differ with respect to means. Other possible choices of distributions for continuous **Y** are student, gamma, and log-normal distributions (Vermunt & Magidson, 2002), but these options are much less utilized than normal distributions. If **Y** are categorical (binary or ordinal), the probability density function is often assumed to be multinomial distribution; if **Y** are counts, the probability density can be assumed to follow Poisson, binomial, or negative binomial distribution (Vermunt & Magidson).

As with any other statistical modeling, LCM is constructed based on a set of assumptions. The most basic assumption of LCM is that latent classes are internally homogeneous, and relationships among observed variables can be explained by the latent categorical variable *C* (Clogg, 1995). In other words, data is generated by a mixture of underlying probability distributions, and this is the foundation for Equation 2.1. The above assumption underlies every LCM study, but it is rarely discussed or tested. Instead, the most often discussed assumption in LCM is the local independence assumption, which states that indicators are independent of one another within each latent class (McCutcheon, 2002). For continuous indicators, this means that within-class covariance structure is a diagonal matrix; for categorical indicators, this means that the within-class joint probability is a multiplication of marginal probabilities. This assumption is very appealing in LCM because it means that all the observed correlations among indicator variables are completely accounted for by the categorical latent variable, and by further relating the categorical latent variable to possible covariates, one may be able to explain the source of heterogeneity in the population.

Model estimation

Two main methods for parameter estimation in LCM are maximum likelihood (ML) and maximum a posteriori (MAP). The log-likelihood function required in ML and MAP can be derived from the probability density function presented in Equation 2.1 or Equation 2.2. The following section describes ML estimation with Expectation-Maximization (EM) algorithm from Mplus (Muthén & Muthén, 2007).

The observed data log likelihood is,

$$\log L = \sum_{i=1}^{n} \log f(\mathbf{y}_i \mid \mathbf{x}_i)\,, \tag{2.2}$$

where $f(\mathbf{y}_i \mid \mathbf{x}_i)$ is a mixture distribution for individual *i*, defined as

$$\sum_{c=1}^{C} P(c_{ic} = 1 \mid \mathbf{x}_i)\, f(\mathbf{y}_i \mid c_{ic} = 1, \mathbf{x}_i)\,. \tag{2.3}$$

ML estimation uses the EM algorithm, which begins with a set of starting values for estimated model parameters, while data on **c** are considered missing. The complete-data log is

$$\sum_{i=1}^{n} (\log P(\mathbf{c}_i \mid \mathbf{x}_i) + \log f(\mathbf{y}_i \mid \mathbf{c}_i, \mathbf{x}_i))\,. \tag{2.5}$$

In the E step, the conditional expectation of the complete-data log likelihood is expressed using the conditional probability of individual *i* belonging to class *c*. Such conditional probability is computed based on the observed data and the current parameter estimates,

$$p_{ic} = P(c_{ic} = 1 \mid \mathbf{y}_i, \mathbf{x}_i) = P(c_{ic} = 1 \mid \mathbf{x}_i)\, f[\mathbf{y}_i \mid c_{ic}, \mathbf{x}_i] / f[\mathbf{y}_i \mid \mathbf{x}_i]\,. \tag{2.6}$$

From a Bayesian point of view, this is the posterior probability of group membership. In the M step, the conditional expectation of the complete-data log likelihood is maximized in order to give updated values of the parameter estimates. These new estimates of the parameters replace the initial estimates and the algorithm returns to the E step. Maximizing the expected complete-data log likelihood leads to a separate M step for each model parts: **c** related to **x**; and **y** related to **c** and **x**.

The maximization for **c** related to **x** leads to a multinomial regression optimization,

$$\sum_{i=1}^{n}\sum_{c=1}^{C} p_{ic} \log P(c_{ic} = 1 \mid \mathbf{x}_i) . \qquad (2.7)$$

If **Y** are categorical (binary or ordinal), the maximization of y related to c and x leads to logistic optimization,

$$\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{c=1}^{C} p_{ic} \log P(y_{ij} \mid c_{ic}, \mathbf{x}_i) . \qquad (2.8)$$

If **Y** are continuous,

$$f(\mathbf{y}_i \mid \mathbf{c}_i, \mathbf{x}_i) = f(\mathbf{y}_i \mid \mathbf{x}_i)_1^{c_{i1}} f(\mathbf{y}_i \mid \mathbf{x}_i)_2^{c_{i2}} \ldots f(\mathbf{y}_i \mid \mathbf{x}_i)_C^{c_{iC}} , \qquad (2.9)$$

where $f(\mathbf{y}_i \mid \mathbf{x}_i)_c$ is multivariate normal density function for class *c* conditional on x. So

$$\sum_{i=1}^{n} \log f(\mathbf{y}_i \mid \mathbf{c}_i, \mathbf{x}_i) = \sum_{i=1}^{n}\sum_{c=1}^{C} c_{ic} \log f(\mathbf{y}_i \mid \mathbf{x}_i)_c . \qquad (2.10)$$

The maximization considers an optimization of

$$E[\sum_{i=1}^{n} \log f(\mathbf{y}_i \mid \mathbf{c}_i, \mathbf{x}_i) \mid \mathbf{y}_i, \mathbf{x}_i] = \sum_{i=1}^{n}\sum_{c=1}^{C} p_{ic} \log f(\mathbf{y}_i \mid \mathbf{x}_i)_c .$$

A potential problem with ML approach is that they can have several local maxima (McCutcheon, 2002) and LCM frequently encounters this problem. For this reason, it is usually recommended that researchers repeat the procedure with a number of different starting values and choose the solution with the highest likelihood (Muthén, 2002). Another problem with ML approach is that it cannot deal with perfect responses, which leads to the development of the alternative estimation method—MAP.

Bayesian MAP estimation involves maximizing the log-posterior distribution, which is the sum of the log-likelihood function and the logs of the priors for the parameters. The latter prevents the occurrence of boundary solutions. MAP estimation is available from the LatentGold program (Vermunt & Magidson, 2005). Suppose that we denote the vector of all free parameters as $\boldsymbol{\theta}$, and the prior for $\boldsymbol{\theta}$ as p($\boldsymbol{\theta}$), and the posterior as P, then the log-posterior function is expressed as,

$$\log P = \log L + \log p(\boldsymbol{\theta}), \tag{2.11}$$

where log*L* is the log-likelihood function for ML estimation, and p($\boldsymbol{\theta}$) serves as a function that penalizes solutions that are too close to the boundary of the parameter space and, therefore, smoothing the estimates away from the boundary (Vermunt & Magidson).

Model identification

Model estimation is closely related to model identification, which concerns whether there is sufficient information from the data for the estimation of free parameters. A model is identified if there are sufficient observed data to obtain unique estimates of the free parameters (van de Pol, et al., 1996, as cited in Jansen et al., 2002). In other words, degrees of freedom cannot be smaller than zero. Take LCM with dichotomous indicators as an example, degrees of freedom is calculated as the difference between number of free parameters in the saturated model (i.e., number of observed response patterns minus 1) and number of free parameters in the model of interest. Suppose there are *p* dichotomous indicators, the maximal number of response patterns is equal to 2^p, and number of free parameters in the model specified in Equation 1.1, is

$l = C - 1 + C \times p$, so a necessary condition for model identification is that $2^p - 1 \geq l$. This places a constraint on the maximal number of latent classes to be extracted. For example, for 3 dichotomous indicators, number of latent classes cannot exceed 2; for 4 dichotomous indicators, number of latent classes cannot exceed 3.

However, the above rule is only a necessary but not sufficient condition for identification. For example, with 4 dichotomous indicators, models with 3 latent classes are not identified although degrees of freedom are positive (Goodman, 1974). As a matter of fact, it is often difficult to verify that a mixture model is identified (Muthén & Muthén, 2007). Several practical approaches are proposed to check whether a model is identified. The researcher may begin with quite different starting values for each of the model parameters and estimate the same model several times; if the final estimates of some or all of the model parameters are quite different but the estimated frequencies and the chi squares are the same, the model is not identified (McCutcheon, 2002). The invertibility of the Fisher information matrix used to produce the SE provides another empirical criterion to evaluate whether a model is identified (Muthén & Muthén).

Model evaluation

There are two issues in model evaluation for LCM: the first one concerns the decision on the number of latent classes, the second one concerns evaluating the form of the within-class covariance structure (e.g., whether local independence holds or not) given the number of latent classes (Vermunt & Magidson, 2002). Assumptions with respect to the within-class covariance structure given the number of latent classes (i.e., the second issue above) can be tested using standard likelihood-ratio (LR) tests between nested models. Although the first issue above also involves nested models (i.e., comparison of models with different numbers of latent classes), it has been well documented that the standard LR test is not applicable because the necessary regularity conditions are not met in mixture model problems since, under the null hypothesis, the general mixing proportions are on the boundary of the parameter space (i.e., zero) and the parameters are not identifiable under the null model (Lo, Mendell, & Rubin, 2001). Most research with regard to model evaluation in LCM has been focusing on the first issue, and the decision on the number of latent classes often relies on data-model fit. Frequently used fit indices for LCM fall into either of the following two categories: information criteria (IC) and LR based tests.

Both IC and LR based tests are relative fit indices assessing the relative merit of a model when compared with an alternative model. IC has been widely used in comparing the fit of non-nested models, and in the case of LCM, it has been used to compare the fit of nested models which do not meet the regularity assumption of the chi square difference tests. IC takes into account both log likelihood estimates and model complexity, and favors the less complicated models with comparable log likelihood estimates. Frequently used ICs include AIC (Akaike, 1974), BIC (Raftery, 1995) and adjusted versions of them. The basic AIC and BIC are calculated as follows,

$$AIC = \text{-2logL} - 2df \ , \tag{2.12}$$

$$BIC = \text{-2logL} - df \times \ln n \, , \tag{2.13}$$

where *df* stands for degrees of freedom and *n* is the total sample size. IC is usually scaled in such a way that better goodness-of-fit is indicated by lower values in IC. Based on the above equations, AIC penalizes more complex models by the total number of free parameters (the larger the number of free parameters, the smaller the degrees of freedom), and BIC penalizes more complex models by both the total number of free parameters and the total sample size. Note that a different computing equation of AIC and BIC statistics is available by replacing *df* with the number of free parameters and the minus sign with a plus sign (see Yang, 2006). Since degrees of freedom and number of free parameters are perfectly and linearly related, the two computing equations are equivalent to each other.

Adjusted versions of the above two ICs are available to improve on their performance. For example, Bozdogan (1987) derived a consistent version of AIC (CAIC) by adding sample size adjustment to the original calculation of AIC,

$$CAIC = \text{-2logL} - df\,(\ln n + 1)\,. \tag{2.14}$$

CAIC executes more severe penalty on over-parameterization than BIC or AIC, so it tends to favor more parsimonious models. Sclove (1987) adjusted the sample size element in the BIC computation by replacing n^* with n, where,

$$n^* = (n + 2) / 24\,. \tag{2.15}$$

This new BIC is called sample size adjusted BIC (ABIC) and it somewhat reduces the penalties of over-parameterization from the original BIC.

Although the standard LR test is not applicable to testing models with different numbers of latent classes, adjusted versions of the LR test are available to approximate the distribution of the LR differences. Two frequently used LR based tests are the parametric bootstrap likelihood ratio test (BLRT; MaLachlan & Peel, 2000) and the Vuong-Lo-Mendell-Rubin (LMR) likelihood ratio test (Vuong, 1989). Both tests examine whether the *c-1*-class model should be rejected in favor of the *c*-class model.

BLRT uses parametric re-sampling technique to approximate true distribution of the log likelihood difference between the *c-1*- and the *c*-class models. More specifically, a bootstrap sample is generated from Equation 2.1 or 2.2 based on the *c-1*-class model using model parameter estimates; the value of -2log*L* is computed for the bootstrap sample after fixing mixture models for *c-1*-class model and *c*-class model in turn and the difference in -2logL is then computed. This process is repeated independently for a number of *K* times, and the replicated values of the difference between -2log*L* form the empirical distribution which can be used as a reference to assess the difference in -2log*L* from the observed data. A significant *p* value from the test provides evidence to reject the c-*1*-class model in favor of the *c*-class model.

LMR was built based on the theorem that under certain regularity conditions (which are met in identified latent class models), the limiting condition of the LR statistic is a weighted sum of $a + b$ independent chi-square random variables when the competing models are nested (for more information, refer to Theorem 1 from Lo, Mendell, & Rubin, 2001). Here, *a* denotes the number of free parameters in one model, and *b* in the other. A significant test result indicates that the *c*-class model is preferred over the *c-1*-class model.

Note from Table 1-1 that another frequently used model selection criteria is entropy. Entropy indicates the quality of classification or prediction of class membership based on the current classification (Celeux & Soromenho, 1996). LCM is sometimes called a “fuzzy” clustering technique because each individual has a certain probability of belonging to each latent class, and an individual is assigned to a certain class based on their most likely membership. In other words, LCM allows for uncertainty in their classification. Entropy is a measure of this uncertainty (or certainty). Entropy (denoted as *E* in the following equation) is calculated using the posterior probabilities, the overall sample size *n* and number of latent classes *C*,

$$E = 1 - \frac{\sum_{i=1}^{n}\sum_{c=1}^{C}(-\pi_{ic}\ln\pi_{ic})}{n\ln C}\,. \tag{2.16}$$

Entropy ranges between 0 and 1, with higher values indicating better classification quality. Entropy approaches 1 when the posterior probability (i.e., π_{ic}) is close to 1 for one class, and close to 0 for all other classes.

So far, there is no widely accepted cutoff for “good” entropy values. Existing research has implied a link between a high entropy value (e.g., > .80) and high correct assignment (Lubke & Muthén, 2007). Some researchers have also conveniently used above .60 to indicate acceptable classification quality (e.g., Murphy, Shevlin, & Adamson, 2007). Due to the lack of consensus on what is a good value of entropy, Pastor et al. (2007) suggested that the statistic is best used to compare the classification utility of different models fit to the same sample or of the same model fit to different samples. But this proposal has not been examined. To a certain extent, entropy does not directly address the question of how the model fits the data; it only addresses the easiness of individual assignment under the current model. It is studied together with all other fit indices in this dissertation because it is sometimes used for model selection and the “true” model should have high classification quality.

All of the above mentioned fit statistics assess the relative merit of a model as compared to an alternative model. The major pitfall of all those relative fit indices is that they cannot address the need for checking how well the model fits the data by merely comparing models with different numbers of latent classes (Bauer and Curran, 2003). Absolute goodness-of-fit indices address the comparison of the specified model with an unrestricted model. For mixture models, when the indicators are categorical, one can consider the observed frequencies in the contingency table as the unrestricted model, and compare the estimated frequencies with observed frequencies using either Pearson chi square statistic (χ^2) or the likelihood ratio chi square (G^2) test statistic.

For each possible response pattern, the expected number of occurrences based on estimated model parameters is compared with the actual number of occurrences. Both χ^2 and G^2 statistics are computed across all possible response patterns. χ^2 is computed using the following formula,

$$\chi^2 = \sum_r \frac{(F_r - f_r)^2}{f_r}, \tag{2.17}$$

where F_r and f_r denote the observed and expected counts for cell r from a multi-way contingency table. G^2 is computed using the following formula,

$$G^2 = 2\sum_r F_r \ln(F_r / f_r). \tag{2.18}$$

Both χ^2 and G^2 have asymptotic chi-square distributions with respect to degrees of freedom; thus the probability of rejecting the null hypothesis in favor of the alternative hypothesis can be determined. However, a widely known limitation with χ^2 and G^2 is that in the case of sparse data, the means of the computed statistics tend to deviate from the expectation of chi square distribution, but in a variety of situations, χ^2 is distributed more like a chi-squared than G^2 in the case of sparse data (Collins, Fidler, Wugalter, & Long, 1993). Muthén (2008) has pointed out that those statistics are not very useful when number of indicators exceeds eight since too many response patterns are involved.

Although unrestricted model is conveniently defined as the observed contingency tables for categorical indicators, for mixture models with continuous indicators, no unrestricted models exist. Variances and covariances are not sufficient statistics of model estimates in LCM; therefore, absolute goodness-of-fit has to resort to higher-order moments such as skewness and kurtosis (Muthén, 2008). In response to the lack of absolute fit index, Muthén and Asparouhov (2002) proposed the multivariate skewness and kurtosis tests (MST and MKT) for testing the fit of a *k*-class mixture model for continuous outcomes. It relies on testing whether the estimated multivariate skewness and kurtosis fit the corresponding sample quantities. The sampling distributions of the multivariate skewness and kurtosis are approximated by parametric bootstrap techniques. Significant *p* values from those tests indicate poor model-data fit. Muthén (2003) suggests that MST and MKT might be a useful complement to the more often used relative fit indices.

So far, simulation studies on fit indices used in LCM have focused on ICs and LR based tests (Nylund, et al., 2006; Wu & Lei, 2008; Yang, 2006). Further research on MST and MKT is needed (Muthén, 2003). The next section reviewed designs and findings of three simulation studies focusing on identifying the correct number of latent classes in LCM.

Existing studies on correct class extraction rates in LCM

As stated earlier, studies on the performance of LCM have been focusing on latent class extraction. This section reviewed three studies (i.e., Nylund et al., 2007; Wu & Lei, 2008; Yang, 2006) that examined the performance of different fit indices in identifying the correct number of latent classes with either categorical or continuous indicators using LCM.

The Yang study

The first study was conducted by Yang (2006) to inform phenotype research. This simulation study manipulated sample size, number of latent classes, and number of indicators, and evaluated information criteria including AIC, BIC, CAIC, and ABIC. Guided by empirical research in the phenotypic field, the study's latent classes ranged from 4 to 6, with number of binary indicators of 12, 15 and 18. Number of binary indicators was nested within number of latent classes so that number of indicators per class was fixed at 3 in this study. Each latent class was defined by three binary variables with their within-class indicator means equal to 0.6 and for the rest of the indicators, their within-class indicator means are equal to 0.1. Six levels of sample sizes were included (i.e., 100, 200, 300, 500, 700, and 1000). Class proportions were fixed to be equal across latent classes. A thousand replications were implemented.

Yang (2006) observed a significant sample size effect on the correct identification of latent classes. At the smallest sample sizes (i.e., n = 100), average correct identification rates of different ICs ranged from just above 0 to around .60 with AIC and ABIC being the best and other ICs with correct rates under .40. Using correct identification rates of .80 or above as cutoff for satisfactory correct class extraction rates, only ABIC and AIC demonstrated satisfactory correct identification rates at sample sizes of 200. While ABIC's performance continued to improve as sample sizes increased, AIC's performance began to decrease as sample sizes increased from 200. BIC did not achieve satisfactory correct class extraction rates until sample sizes were around 700. Number of true latent classes also showed a significant impact on most fit indices except for ABIC: average correct identification rates ranged from .75 for 4-class models to .30 for 6-class models for BIC and from .65 to .25 for CAIC.

Based on the simulation results, Yang (2006) concluded that ABIC was a great improvement over BIC; effects of sample sizes and number of latent classes seemed to be critical in LCM. Yang recommended that ABIC needed at least 50 subjects per latent class. The author also discarded the use of AIC in LCM because of its inconsistency property. Yang also noted the limitation of this sample size recommendation by highlighting the balanced proportions of latent classes in the simulation as well as the distinct structure of the conditional probabilities for data generation.

The Nylund et al. study

Nylund et al. (2007) conducted a more comprehensive study on the performance of a variety of fit indices for mixture models. They examined the performance of both LR-based tests (traditional chi-square difference test, LMR & BLRT) and ICs (AIC, BIC, ABIC & CAIC) in determining the number of latent classes in LCM with categorical (dichotomous) or continuous indicators. For each set of models (i.e., dichotomous and continuous), Nylund et al. studied several conditions which differed by number of indicators (8, 10 and 15), number of latent classes (3 and 4), sample sizes (200, 500, and 1000), distribution of class sizes (equal and unequal with 5% being the smallest class) and the structure of the generating models. The structure of the models was either simple, in which the means of certain indicators were particularly high or low for a given class so that these indicators discriminated among the classes (i.e., the same pattern as the structure used in Yang, 2006), or complex, in which no single indicator mean was particularly high or low for a specific class. For the complex structure, they referred to the area of Attention Deficit Hyperactive Disorder (ADHD), where one class had high means on all indicators, while other classes had high means on a certain subset of indicators but low means on other indicators.

Nylund et al. found that BIC performed the best of all ICs except when class sizes were unequal and data structure was complex with categorical indicators; under those circumstances, ABIC outperformed BIC. Out of the three LR-based tests, BLRT consistently outperformed the other two across all models considered. Sample size showed a significant effect, especially for unequal class proportion conditions. When classes were equally distributed, the lowest sample sizes of 200 seemed to be sufficient for all models studied with almost perfect identification rates. When classes were unequal, ABIC seemed to be superior to BIC for categorical variables, and a total of 500 cases were needed for the correct identification rates to be above .80. If BLRT was used, sample sizes of 200 were sufficient for equal class proportions for categorical indicators, and all conditions for continuous indicators. In the case of unequal class proportions and categorical indicators, BLRT would need 500 cases to yield satisfactory correct identification rates.

Nylund et al. (2007) seemed to suggest that sample sizes of 50 per class is sufficient for equal class proportions, which is consistent with Yang (2006)'s recommendation. But since they did not examine sample sizes smaller than 200, there was little information on how low the class sizes could go while still maintaining satisfactory identification rates. When class proportions were unequal, the study seemed to suggest that a class size of 25 was needed (total sample sizes of 500 times the smallest class proportion of 5%) for categorical indicators and a class size of 10 was needed for continuous indicators (total sample sizes of 200 times the smallest class proportion of 5%). This seems to suggest a different requirement of class sizes for the categorical indicators and continuous indicators. But a major limitation with this study is that some manipulated factors were confounded with one another, e.g., the only model with unequal class proportions for the continuous model was the one with the complex data generating structure, so it was not clear whether the unequal class proportion or the complex data generating structure contributed to different performance.

The Wu & Lei study

In order to address the possible confounding effects of some factors from the Nylund et al. (2007) study as well as the lack of attention to models with continuous indicators, more recently, Wu & Lei (2008) focused their attention on continuous indicators, and manipulated sample sizes (n = 102, 204, 504, and 1,002), number of true latent classes (2 and 3), number of indicators (3 and 6), class proportions (equal and unequal) and profile patterns (quantitative vs. qualitative difference). Quantitatively different classes occur when one class has higher means on every indicator than another class, and qualitatively different classes occur when one class has higher means on some indicators than another class, but lower means on other indicators. All manipulated factors were fully crossed in this study.

The authors found that sample sizes contributed the most to the variation in correct identification rates, followed by types of fit indices. In general, BLRT outperformed all other fit indices, followed by BIC, and then ABIC. Using .80 as the cutoff for satisfactory correct class extraction rates, BLRT would need sample sizes of 200 or above to achieve an average identification rates that was above .80 for 6 indicators; BIC and ABIC did not achieve that level until when sample sizes were around 500. In terms of class sizes, the smallest class size that showed satisfactory performance for 6-indicator models was 40 (a total sample of 200 times the smallest class proportion of .20). Three indicators posed much bigger challenge for correct class extraction, and even BLRT would need sample sizes of 1,000 to have identification rates that just reached .80.

Wu and Lei (2008)'s findings were consistent with Nylund et al. (2008) in that BIC and BLRT performed better in general than other indices. Wu and Lei also seemed to suggest that when number of indicators was 6, class sizes of 40 appeared to be sufficient to yield satisfactory class identification rates when BLRT was used; but larger class sizes were needed when number of indictors was 3.

Summaries of findings from the three studies

All three studies found that larger sample sizes and more indicators contributed to better identification rates, and some fit indices (e.g., BLRT, ABIC, and BIC) were better than others (e.g., AIC, LMR). Furthermore, the scales of indicators may affect the requirement on sample size for the purpose of correctly extracting latent classes. Information regarding sample sizes or class sizes was still very sporadic though: sample sizes of 200 seemed to be sufficient for equal class proportions when number of indicators was 6 or above for either categorical and continuous indicators (Nylund et al., 2007; Wu & Lei, 2008; Yang, 2006); lowest class sizes observed from those studies that had shown satisfactory correct identification rates were 10 for the 4-class model with continuous indicators, and 25 for the 4-class model with categorical indicators from Nylund et al. A systematic study is needed to examine fine-grained levels of sample sizes and number of indicators. By studying more levels of sample sizes, a closer look into the variation in class sizes may be conducted that can help us decide whether total sample sizes or class sizes play a more important role in affecting performance of LCM.

Chapter 3

Method

Since one never knows the true model with real data, simulation seems to be the appropriate approach for the purpose of evaluating model performance. The advantage of simulated data is that the truth is known and thus allows the evaluation of LCM's performance under different conditions. This chapter provides detailed description of the simulation design and data analytical strategies. Manipulated factors in this study intend to include a complete list of factors that have been shown to affect model performance. Some of those factors (e.g., sample size, scales of indicators, and number of indicators) are directly observable by the researcher and can even be manipulated by researchers during designs of the experiment, so findings with respect to those factors have important implications for practitioners. Other factors are not observable by researchers from raw data (i.e., number of latent classes, class proportions, class separation), not to mention to manipulate them, but researchers may have some hypotheses about the level of those factors based on educated guess. Since the purpose of this study is to inform research practice, special interest is placed on the first set of "experimental" factors (i.e., sample size, number of indicators, and indicator scales) while controlling for other factors that are not easily controlled by researchers in applied research.

Manipulated factors

This study manipulated factors that have been identified in previous chapters as important in affecting model performance. The levels of those factors were selected based on the review of empirical studies reported in Table 1-1. Table 3-1 lists a summary of the simulation design.

Table 3-1. Manipulated factors in the simulation study

Manipulated factors	
Score scale	Dichotomous (D), Continuous (C)
Number of indicators	5, 6, 9, 12, 15
Sample size	50, 100, 200, 300, 400, 500, 1000, 2000, 5000
Number of true classes	2, 3, 4
Class proportion	Equal, unequal with 10% being the smallest group
Class separation	Small (.2 for D, .5 SD for C) large (.5 for D, 2 SD for C)

Note. All manipulated factors were fully crossed.

Score scale. Dichotomous responses are very common in LCM, and almost half of the studies from Table 1-1 deal with dichotomous responses. It is also common in those studies that the originally continuous scale was converted to dichotomous scale before LCM analysis. Continuous scales are becoming more common, and they constitute most of the remaining half. As a result, this study included score scales that are either dichotomous or continuous.

Sample size. Sample sizes range from less than 50 to over 8,000 in studies from the literature search, with around 85% of the studies having sample sizes within 2,000 (top panel of Figure 1-1). The smallest size was set just above the smallest sample from Table 1-1. To represent reasonably detailed levels of sample sizes, this study included nine levels (i.e., n = 50, 100, 200, 300, 400, 500, 1000, 2000, and 5000). Smaller sample sizes (e.g., 50 to 200) were intended to address the needs of researchers who conduct smaller scale studies for theory forming or investigation of individual differences; on the other hand, larger sample sizes (e.g., n > 1,000) were intended to address the needs of researchers who conduct large scale studies or use existing database to answer policy related questions.

Number of indicators. Number of indicators in empirical studies from Table 1-1 ranges from as few as 2 to as many as 24, and around 80% of them have 18 or fewer indicators (the middle panel of Figure 1-1). To represent reasonably detailed levels of number of indicators, this study included five levels (i.e., p = 5, 6, 9, 12, and 15). The smallest number of indicators was chosen to be 5 because as discussed in Chapter 2, with 4 or fewer indicators, LCM with 4 classes was already under-identified for dichotomous indicators.

Number of latent classes. As the bottom panel of Figure 1-1 shows, number of estimated latent classes ranges from 2 to 7, with 70% of those studies having 4 or fewer classes. Therefore, this study examined models with 2, 3, and 4 latent classes.

Class proportions. Both equal and unequal class proportions were studied. For unequal class proportions, mixing proportions of 10% and 90% were used for 2-class models, 10%, 10% and 80% were used for 3-class models, and 10%, 15%, 15%, and 60% were used for 4-class models. All those levels were selected based on empirical studies in Table 1-1.

Class separation. In Nylund et al., class mean differences was fixed at 2 standard deviations for continuous indicators. Class mean differences at similar levels are also observed in a number of empirical studies listed in Table 1-1 (e.g., Aldridge & Roesch, 2008; DiStefano & Kamphaus, 2006; Klonsky & Olino, 2008; Macy, Nurius, & Norris, 2007; Oxford, Gilchrist, Lohr, Gillmore, Morrison, & Spieker, 2005). This was used to represent large class separation for continuous indicators in the current study; to represent small class separation, class mean difference was fixed at .5 standard deviation. For dichotomous indicators, Yang fixed the indicator probability difference at .5, and Nylund et al. fixed the difference at .70. This study used .5 (Yang's difference) to represent large class separation and .2 to represent small class difference. Similar levels of indicator mean differences have been observed in empirical studies (e.g., Bouwmeester & Sijtsma, 2007; Chan, Leu, & Chen, 2007; Chung, Martin, Cornelius & Clark, 2008; Chung & Martin, 2005; Jackson & Sher, 2008; Morgan-Lopez & Fals-Stewart, 2007; Notenboom & Reitsma, 2007).

Data generation

All data were generated by Mplus 5.0 (Muthén & Muthén, 2007) using its Monte Carlo facility. The Mplus Monte Carlo facility provides a mixture model procedure, through which users can specify sample size, number of indicators, scale of indicators, number of true latent classes, and number of replications. For continuous indicators, users provide class proportions and class specific means and variances and covariances; multivariate normal distribution is assumed within each latent class with class-specific means and covariances. This study has specified a diagonal covariance matrix for continuous indicators so that local independence holds in data generating models. For dichotomous indicators, users provide class proportions and class-specific indicator probabilities (i.e., indicator mean); within-class indicator probability can be translated into threshold values on logit scales for each indicator using the following equation,

$$\tau_{pc} = \log \frac{p_{pc}}{1 - p_{pc}}, \tag{2.1}$$

where τ_{pc} denotes the threshold value of indicator *p* for class *c*, and p_{pc} indicates probability for indicator *p* in class *c*. Note that when indicator probabilities approaches 0 or 1, the threshold values approaches negative or positive infinity, respectively. Mplus deals with perfect response (those with indicator probabilities of 1 or 0) by fixing the threshold values to -15 and 15, respectively. Appendix shows an example of using this procedure to generate 100 replications of a 2-class model with 5 dichotomous indicators with equal class proportions.

Table 3-2 shows class-specific indicator means used for data generation. For continuous indicators, variances of all indicators were set to 1. In terms of profile patterns or data structure, the simple structure (i.e., each class has higher means on some indicators but low means on other indicators) from Nylund et al. (2007) was utilized because a) Wu and Lei (2008) found that the effect of structure difference was small compared with that of sample size and fit indices and b) simple structures were used more often in previous studies. Therefore, each class was represented by a number of indicators that identify that class (i.e., high means on those indicators, but low means on other indicators). The class means of all 15 indicators are listed in Table 3-2. Conditions involving *p* indicators ($p \leq 15$) are formed by selecting the first *p* indicators from Table 3-2. Table 3-2 is arranged in such a way that class-specific means of an indicator (represented by a single line in Table 3-2) repeat after a number of lines so that the added indicators for models with more indicators have the same means with indicators for models with fewer indicators. This is done to ensure that the number of indicators does not confound with indicator's ability to distinguish classes. The fully crossed design of the manipulated factors resulted in a total of 1080 simulation conditions (9 sample sizes × 5 number of indicators × 3 number of true latent classes × 2 class proportions × 2 class separation x 2 scales of indicators). One hundred replications were conducted for each condition.

Table 3-2. True within-class indicator means for data simulation.

	Continuous indicators, Small Class Separation								
	2-class		3-class			4-class			
	1	2	1	2	3	1	2	3	4
I1	.5	0	.5	0	0	.5	0	0	0
I2	0	.5	0	.5	0	0	.5	0	0
I3	.5	0	0	0	.5	0	0	.5	0
I4	0	.5	.5	0	0	0	0	0	.5
I5	.5	0	0	.5	0	.5	0	0	0
I6	0	.5	0	0	.5	0	.5	0	0
I7	.5	0	.5	0	0	0	0	.5	0
I8	0	.5	0	.5	0	0	0	0	.5
I9	.5	0	0	0	.5	.5	0	0	0
I10	0	.5	.5	0	0	0	.5	0	0
I11	.5	0	0	.5	0	0	0	.5	0
I12	0	.5	0	0	.5	0	0	0	.5
I13	.5	0	.5	0	0	.5	0	0	0
I14	0	.5	0	.5	0	0	.5	0	0
I15	.5	0	0	0	.5	0	0	.5	0
	Continuous indicators, large class separation								
	2-class		3-class			4-class			
	1	2	1	2	3	1	2	3	4
I1	2	0	2	0	0	2	0	0	0
I2	0	2	0	2	0	0	2	0	0
I3	2	0	0	0	2	0	0	2	0
I4	0	2	2	0	0	0	0	0	2
I5	2	0	0	2	0	2	0	0	0
I6	0	2	0	0	2	0	2	0	0
I7	2	0	2	0	0	0	0	2	0
I8	0	2	0	2	0	0	0	0	2
I9	2	0	0	0	2	2	0	0	0
I10	0	2	2	0	0	0	2	0	0
I11	2	0	0	2	0	0	0	2	0
I12	0	2	0	0	2	0	0	0	2
I13	2	0	2	0	0	2	0	0	0
I14	0	2	0	2	0	0	2	0	0
I15	2	0	0	0	2	0	0	2	0

Table 3-2. Continued

Dichotomous indicators, Small class separation									
	2-class		3-class			4-class			
	1	2	1	2	3	1	2	3	4
I1	.7	.5	.7	.5	.5	.7	.5	.5	.5
I2	.5	.7	.5	.7	.5	.5	.7	.5	.5
I3	.7	.5	.5	.5	.7	.5	.5	.7	.5
I4	.5	.7	.7	.5	.5	.5	.5	.5	.7
I5	.7	.5	.5	.7	.5	.7	.5	.5	.5
I6	.5	.7	.5	.5	.7	.5	.7	.5	.5
I7	.7	.5	.7	.5	.5	.5	.5	.7	.5
I8	.5	.7	.5	.7	.5	.5	.5	.5	.7
I9	.7	.5	.5	.5	.7	.7	.5	.5	.5
I10	.5	.7	.7	.5	.5	.5	.7	.5	.5
I11	.7	.5	.5	.7	.5	.5	.5	.7	.5
I12	.5	.7	.5	.5	.7	.5	.5	.5	.7
I13	.7	.5	.7	.5	.5	.7	.5	.5	.5
I14	.5	.7	.5	.7	.5	.5	.7	.5	.5
I15	.7	.5	.5	.5	.7	.5	.5	.7	.5

Dichotomous indicators, Large class separation									
	2-class		3-class			4-class			
	1	2	1	2	3	1	2	3	4
I1	.8	.3	.8	.3	.3	.8	.3	.3	.3
I2	.3	.8	.3	.8	.3	.3	.8	.3	.3
I3	.8	.3	.3	.3	.8	.3	.3	.8	.3
I4	.3	.8	.8	.3	.3	.3	.3	.3	.8
I5	.8	.3	.3	.8	.3	.8	.3	.3	.3
I6	.3	.8	.3	.3	.8	.3	.8	.3	.3
I7	.8	.3	.8	.3	.3	.3	.3	.8	.3
I8	.3	.8	.3	.8	.3	.3	.3	.3	.8
I9	.8	.3	.3	.3	.8	.8	.3	.3	.3
I10	.3	.8	.8	.3	.3	.3	.8	.3	.3
I11	.8	.3	.3	.8	.3	.3	.3	.8	.3
I12	.3	.8	.3	.3	.8	.3	.3	.3	.8
I13	.8	.3	.8	.3	.3	.8	.3	.3	.3
I14	.3	.8	.3	.8	.3	.3	.8	.3	.3
I15	.8	.3	.3	.3	.8	.3	.3	.8	.3

Note. Variance is fixed to 1 for the continuous indicators.

Analytical strategies

Model estimation was carried out in Mplus 5.0 with ML. In order to minimize the possibility of getting local maximum estimates, Mplus generates a random set of starting values, and chooses solutions that produce the maximum likelihood. Evaluation of model performance in this study focused on two aspects: correct class extraction rates and classification accuracy.

The first aspect extended existing studies on the correct identification of number of latent classes in LCM. The complete simulation design involving all levels of factors in Table 3-1 was included for this purpose. For true models with *c* latent classes, models with *c*-1, *c*, and *c*+1 latent classes were specified. Those three models were compared and then a best fitting model was identified. Inclusion of the three models in model comparisons was implemented because either under-estimation or over-estimation of number of latent classes can be detected.

In order to decide on the number of latent classes, fit indices produced from each run of models were compared. This study chose commonly used fit indices in empirical studies (i.e., AIC, BIC, ABIC, and BLRT) as well as the less frequently used χ^2, G^2 and entropy statistic. In addition, MST and MKT were evaluated with regard to their Type I error rates with the correctly specified model. For IC related fit indices (i.e., AIC, BIC, and ABIC), models with the smallest value of the particular fit index among the three alternative models was selected. For χ^2 and G^2, the first time a non-significant *p* value (i.e., $p \geq .05$) occurred in the *c*-class model was an indication of choosing the *c*-class model. For BLRT, the first time a non-significant *p* value (i.e., $p \geq .05$) occurred in the *c*-class model was an indication of choosing the *c*-1 model. For MST and MKT, models were rejected when *p* value was smaller than .05.

Correct class extraction rates for each condition were calculated by dividing the number of times the true number of latent classes was identified by number of replications (i.e., 100). Correct class extraction based on the best performing fit indices would then be further analyzed using ANOVA to examine the effects of manipulated factors. Since correct class extraction rates (the dependent variable) were proportions, logit transformation was performed on the proportions before ANOVA was conducted. For extreme values (i.e., 0 or 1), a .005 adjustment was made to bring extreme values away from the boundary (i.e., a proportion of 0 was replaced with .005, and a proportion of 1 was replaced with .995). The adjustment of .005 was chosen so that the adjusted values for 0 and 1 still maintained their relative magnitudes of being the lowest and largest values. That is, the adjusted value of 0 was still lower than the next lowest value of .01, and the adjusted values of 1was still higher than the next highest value of .99.

The second set of analysis focused on classification accuracy, which was computed by comparing the estimated latent class membership with the true latent class membership. Both group classification accuracy and overall classification were examined because high overall classification accuracy did not necessarily mean high group classification accuracy. Consider a situation where one group is very large (say 90%), the overall classification can be very high (around 90%) even if one group classification accuracy rate is close to zero. Group classification accuracy was defined by the percentage of subjects within a certain true latent class being correctly assigned to the same latent class; overall classification accuracy was defined by the percentage of subjects being assigned to their corresponding true latent classes over the entire sample. Mathematically, overall classification accuracy rate is a weighted average of group classification accuracy rates, where class sizes are used as weights. Note that for each latent class, there would be a group classification accuracy measure; in order to simplify the analysis, group classification accuracy was operationally defined as the lowest group classification accuracy among all latent classes.

ANOVA was conducted on the logit transformed values of the overall and group classification accuracy similar to that for correct class extraction rates. For overall classification accuracy, .80 was chosen as the cutoff for acceptable overall classification accuracy. For group classification accuracy, a uniform standard for 2-, 3- and 4-class models would be too stringent, especially for the 4-class models because under random classification, 2-class models would have an average group classification rate of .50 while 4-class models would have an average group classification rate of only .25; therefore, the same value in group classification for the 2-class model and the 4-class model does not mean comparable classification quality. Alternatively, a group classification accuracy rate was considered acceptable if it was statistically significantly higher than what would be observed under random classification, at an alpha level of .01. For example, for 2-class models, random classification would lead to a group classification rate of .50. Using the exact Binomial test with 100 replications, the upper 99% confidence bound was .63, .46, and .37 for the 2-, 3-, and 4-class models respectively.

Because entropy is a frequently used indicator of classification quality in practice, the consistency of entropy values with overall and group classification accuracy measures was examined to evaluate the usefulness of entropy used in such manner. ANOVA was conducted on the logit transformed values of entropy to demonstrate whether or not entropy was mostly affected by the same set of factors that affected the overall and group classification accuracy measures. In addition, conditions with high ($\geq .80$) and low (< .60) values of entropy were identified; overall and group classification accuracy rates in those conditions were examined to see whether the pattern of high and low entropy values corresponded with those of classification accuracy measures.

Chapter 4

Results

This section summarized findings from the simulation study. Results are presented separately for dichotomous and continuous indicators. For each score scale, firstly, findings on correct class extraction rates are reported in the following sequence: the performance of different fit indices was compared in an attempt to provide information on the best fit index/indices; effects of manipulated factors on correct class extraction rates were further discussed. Secondly, effects of manipulated factors on overall and group classification accuracy are presented. Conditions with high ($\geq$ .80) and low (< .60) entropy values were examined for its consistency with high and low classification accuracy. Recommendation for practitioners were then proposed taking into account both correct class extraction and classification accuracy.

Dichotomous indicators

Correct class extraction

Tables 4-1 to 4-7 present correct class extraction rates for each of the studied fit indices under all studied conditions. In order to assist interpretations of the multitude of results, cells with values above .80 were bolded, indicating an at least 80% correct class extraction rate for a simulation condition.

Comparing fit indices

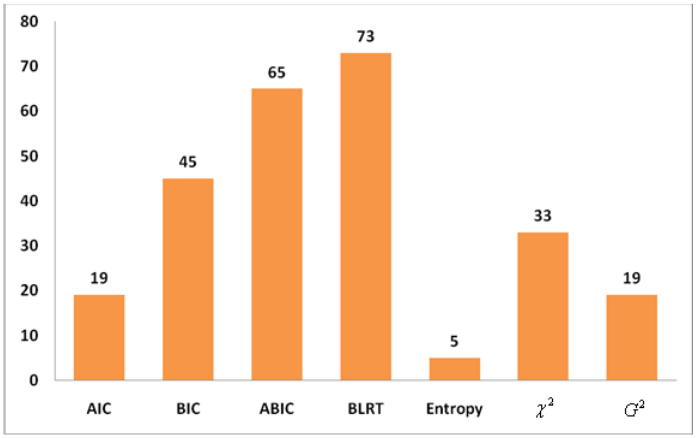

Figure 4-1. Number of cells with acceptable correct class extraction rates ($\geq 80\%$) by fit indices for dichotomous indicators

Figure 4-1 presents for each fit index, number of cells that show acceptable correct class extraction rates. Looking across all fit indices, entropy had the smallest number of cells (5 out of 270 cells) with acceptable correct class extraction rates. Bolded cells for entropy were all 2-class models with a large number of indicators (e.g., 15 indicators). AIC, χ^2 and G^2 performed better than entropy, but not as well as BIC, ABIC, or BLRT. AIC had 19 cells with acceptable correct class extraction rates, mostly for 2- and 3-class models with large class separation. The most distinctive pattern with AIC is its property of being not statistically consistent, which was demonstrated in Table 4-1 as frequent occurrences of decreasing correct class extraction rates with increasing sample sizes. Consistent with Collins et al. (1993), χ^2 from Table 4-6 appeared to have better performance than G^2 from Table 4-7, and number of cells with acceptable correct class extraction rates was 33 and 19 respectively, all for large class separation and mostly for conditions when number of indicators was not very large (i.e., < 12). Both χ^2 and G^2 presented the same pattern in that correct class extraction rates were mostly decreasing as number of indicators increased from 9. This was probably because as number of indicators increased, there would be many more response patterns with frequencies less than 5. Note that a conventional recommendation with the use of χ^2 is that at least 80% of the cells should have cell counts no smaller than 5 and no cell counts should be smaller than 1. Based on this recommendation, sample sizes of at least 135 would be needed for 5 indicators (i.e., $5\times 2^5 \times 0.8 + 2^5 \times 0.2$), 269 for 6 indicators, 2151 for 9 indicators, and over 10 thousand for 12 indicators. This was partially consistent with the pattern presented in Table 4-6: once the number of indicators exceeded 9, correct class extraction rates were largely decreasing.

BIC, ABIC, and BLRT showed better overall performance than the above mentioned fit indices. Among the three fit indices, BLRT, with 73 bolded cells, showed a slight advantage over ABIC which had 65 bolded cells, which in turn outperformed BIC, whose bolded cells were a subset of those for BLRT and ABIC. In general, those three fit indices had higher correct class extraction rates with larger class separation, smaller number of classes, larger sample sizes, and more indicators. Thus, cells with acceptable correct class extraction rates were clustered in the lower right corner for each combination of class separation and true number of latent classes.

Table 4-1. Correct class extraction rates based on AIC for dichotomous indicators

	Number of indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	18-78	62-86	38-74	69-74	38-80	11-16	18-20	21-25	18-35	11-37
100	23-89	77-83	57-79	**81-81**	68-80	19-17	19-22	19-45	27-48	27-48
200	49-90	95-77	**84-80**	80-78	67-74	18-32	13-31	23-52	29-54	29-49
300	65-93	**93-76**	72-82	79-78	72-74	23-45	22-48	28-58	29-54	31-44
400	68-87	**96-77**	73-80	73-76	69-73	26-55	21-52	31-63	32-48	25-38
500	66-93	**95-79**	**81-82**	71-78	65-67	18-54	29-67	32-63	42-50	32-36
1000	**88-95**	**88-83**	79-76	67-74	63-74	24-92	37-86	37-57	41-41	51-34
2000	**88-89**	**90-87**	78-71	69-75	51-69	33-92	41-81	58-64	42-56	49-42
5000	**90-91**	**88-80**	72-77	59-68	54-68	69-90	**84-81**	70-61	47-45	38-35
	3-class									
50	14-33	14-34	33-69	29-73	31-74	6-5	10-5	13-11	14-17	32-21
100	23-42	18-68	46-87	58-71	38-77	4-1	7-3	16-9	27-21	16-22
200	25-69	30-87	76-86	74-61	46-73	2-2	4-6	12-10	25-22	30-21
300	46-80	53-93	80-85	72-71	38-72	1-4	4-4	13-6	18-29	27-37
400	54-87	66-84	89-86	67-77	32-74	1-4	5-4	19-19	22-45	28-22
500	61-86	77-88	13-86	69-70	39-67	4-6	8-7	13-19	32-37	26-32
1000	**87-95**	**91-86**	71-75	64-76	28-66	2-11	6-12	19-34	25-43	37-33
2000	**94-92**	**91-87**	70-85	59-66	66-60	3-17	8-29	28-55	36-40	33-38
5000	**93-93**	**90-89**	73-84	62-60	50-58	13-59	21-83	49-58	46-44	28-31
	4-class									
50	4-8	6-4	3-10	9-0	0	0-1	0-1	9-8	27-11	21-9
100	10-23	8-8	8-18	5-6	2-1	0-2	0	32-8	21-12	19-12
200	8-14	10-20	16-37	6-7	16-6	0	0-2	0-7	20-13	16-26
300	11-11	9-35	31-43	6-13	20-5	0-2	0-2	0-14	18-19	42-17
400	13-10	16-39	42-51	7-12	15-11	0-2	0-10	12-15	31-17	24-27
500	10-12	27-52	38-45	13-20	11-7	0	0-4	0-13	9-19	12-28
1000	16-10	40-76	45-45	30-24	21-8	0	0-6	17-13	46-21	27-39
2000	15-7	70-83	63-48	16-20	22-10	0	0-3	12-16	31-51	81-33
5000	9-8	79-80	55-44	20-17	9-29	0-2	10-11	14-48	58-39	37-21

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-2. Correct class extraction rates based on BIC for dichotomous indicators

	Number of indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	4-68	10-90	17-100	34-99	29-99	2-0	0-1	0	0	0
100	7-93	29-100	45-100	67-100	**80-100**	0-1	1-1	0	0	2-0
200	24-100	60-99	**90-100**	**99-100**	**96-99**	0	0	0	1-0	3-1
300	47-100	**85-100**	**97-100**	**100**	**100**	0-1	0-1	0	1-3	5-5
400	63-100	**99-100**	**99-100**	**100**	**100**	0-2	0	0-2	2-7	1-13
500	70-100	**98-100**	**100**	**100**	**100**	0-1	0-1	1-3	0-11	1-30
1000	**91-100**	**100**	**100**	**100**	**100**	0-5	0-10	0-36	0-72	0-96
2000	**96-100**	**100**	**100**	**100**	**100**	0-35	0-53	0-93	0-100	0-100
5000	**96-100**	**100-80**	**100**	**100**	**99-100**	0-98	0-100	4-100	32-100	74-100
	3-class									
50	1-2	1-1	2-14	2-30	2-33	5-1	2-0	2-0	2-0	7-0
100	0-3	0-7	1-43	3-73	3-88	1-1	1-0	1-0	2-0	1-0
200	0-11	0-41	7-90	9-99	19-98	0-1	1-0	1-0	2-0	1-0
300	0-25	0-84	9-99	20-100	40-99	0	1-0	0	0-1	4-0
400	1-46	4-93	15-100	39-99	62-100	0	0	1-0	1-1	3-10
500	0-70	2-98	28-100	56-100	**81-99**	2-0	0	0	3-10	1-0
1000	15-97	16-100	**84-100**	**100**	**100**	0	0	0	1-0	2-0
2000	56-100	79-100	**99-100**	**100**	**94-100**	0	0	1-0	0	1-0
5000	**100-99**	**100-99**	**99-100**	**100**	**99-100**	2-0	1-0	0-2	0-28	0-93
	4-class									
50	0	0-1	1-0	0	0	0	0	0	0	0
100	1-3	0	0	0-2	0-3	0	0	9-0	0	0
200	0	1-0	0	3-9	0-16	0	0	0	0	0
300	0-1	0	0-3	2-25	3-57	0	.5	0	0	0
400	0-3	0	1-12	1-69	0-76	0-1	9-0	0	0	0
500	0-2	0-1	1-71	1-80	8-74	0	0	0	0	8-0
1000	1-1	0-1	9-98	3-90	73-84	0	0	0	0	0
2000	1-0	0-18	62-100	50-95	100-70	0	0	0	0	0
5000	0-1	25-79	**100**	**98-100**	**100-79**	0	7-0	0	0	0

Note. Values to the left and right sides of “-” represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-3. Correct class extraction rates based on ABIC for dichotomous indicators

	Number of indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	23-31	37-4	9-3	12-2	5-20	25-18	13-12	8-1	0	0
100	23-81	73-41	34-43	33-36	18-24	33-35	33-39	23-28	18-15	8-4
200	50-91	**96-83**	**90-85**	**90-84**	77-78	15-32	13-33	23-55	25-61	29-57
300	69-98	**98-95**	**91-95**	**97-94**	**95-95**	10-29	8-33	13-59	17-74	13-87
400	73-99	**100-98**	**97-98**	**99-100**	**98-98**	5-29	5-33	6-68	10-73	11-93
500	70-99	**98-100**	**100-99**	**100**	**100**	5-36	4-40	4-77	8-84	11-99
1000	**90-100**	**99-100**	**100**	**100**	**100**	3-51	2-69	2-96	1-100	8-100
2000	**96-100**	**100**	**100**	**100**	**100**	0-83	1-95	6-100	28-100	61-100
5000	**96-100**	**100**	**100**	**100**	**99-100**	9-100	14-100	64-100	**89-100**	**97-100**
	3-class									
50	30-32	47-39	41-44	16-11	32-33	29-28	15-16	2-6	4-0	8-0
100	37-48	45-68	59-66	38-23	35-30	14-17	17-17	35-27	22-24	13-18
200	25-65	27-90	72-91	77-80	65-82	2-0	4-3	9-7	18-13	19-18
300	27-84	28-98	**82-97**	**92-93**	**93-96**	1-1	1-1	1-6	5-5	8-11
400	34-87	37-95	**89-98**	**97-96**	**97-99**	0-1	0-2	1-1	3-5	4-3
500	33-91	38-99	**91-99**	**100**	**93-99**	2-0	0	0-1	2-3	3-3
1000	70-97	**82-100**	**99-100**	**100**	**100**	0	0-1	0-1	1-3	2-7
2000	**93-100**	**98-100**	**99-100**	**100**	**94-100**	0	0-1	1-4	0-18	1-49
5000	**100-99**	**99-100**	**100-99**	**100**	**99-100**	2-1	4 1-7	0-60	0-98	1-100
	4-class									
50	19-32	2-1	0	0	0	33-0	28-30	0-7	0-2	14-2
100	18-22	18-19	4-33	1-1	1-0	6-10	19-13	30-32	26-32	19-19
200	73-12	7-16	17-43	6-14	25-12	0	0-1	0-6	22-9	0-15
300	66-7	1-17	32-75	11-46	57-50	0	0-2	0-2	11-1	0
400	5-5	3-19	35-93	8-67	75-71	0-1	8-0	0	0	0-1
500	42-7	9-15	38-96	15-74	88-70	0	0	0	0	8-1
1000	2-2	10-37	**81-100**	**79-89**	**99-84**	0	0	0	0	0
2000	4-2	25-74	**100**	**79-95**	**100-70**	0	0	0	0	0
5000	1-1	**83-100**	**100**	**98-100**	**100-76**	0	5-0	0	0	0

Note. Values to the left and right sides of “-” represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-4. Correct class extraction rates based on BLRT for dichotomous indicators

		Number of indicators									
		5	6	9	12	15	5	6	9	12	15
		Large class separation					Small class separation				
N	N_c	2-class									
50	5	19-82	41-96	40-99	63-96	47-95	5-7	2-5	5-15	6-7	6-12
100	10	28-99	73-99	70-98	**94-94**	**85-98**	7-13	6-17	7-16	2-19	6-38
200	20	66-94	**98-89**	**96-97**	**99-98**	**97-98**	11-26	3-19	6-42	10-47	9-67
300	30	**84-98**	**99-95**	**91-96**	**97-98**	**95-98**	15-32	7-25	11-57	17-77	12-89
400	40	**89-97**	**99-98**	**94-96**	**97-100**	**93-97**	12-43	6-38	7-78	16-79	15-94
500	50	**96-98**	**99-94**	**97-99**	**99-100**	**97-96**	13-45	15-55	7-84	21-89	25-92
1000	100	**99-100**	**96-93**	**91-97**	**98-99**	**95-99**	15-80	34-97	31-88	36-96	52-97
2000	200	**99-94**	**95-95**	**98-98**	**92-100**	**94-96**	38-94	34-97	55-93	74-95	**97-97**
5000	500	**93-93**	**96-91**	**97-99**	**90-98**	**97-96**	60-95	**82-95**	**92-90**	**94-95**	**92-93**
		3-class									
50	5	0-3	1-7	3-26	4-43	2-64	5-0	2-0	2-0	2-0	4-0
100	10	1-9	2-43	15-84	24-91	18-96	1-0	1-0	1-1	2-0	1-0
200	20	4-47	9-89	48-100	75-99	**87-97**	1-0	1-0	1-0	2-2	2-2
300	30	10-79	19-98	74-98	**92-98**	**94-99**	0	1-1	0-2	2-1	4-3
400	40	18-87	35-92	**89-98**	**95-96**	**94-98**	1-0	0-2	1-1	1-2	3-2
500	50	26-91	46-95	**95-99**	**99-98**	**89-96**	2-0	0-1	0-1	4-6	1-4
1000	100	74-98	**95-99**	**97-99**	**96-98**	**91-95**	0-2	0-3	1-9	1-29	3-44
2000	200	**97-95**	**98-98**	**96-99**	**98-96**	**90-96**	2-9	0-10	5-50	4-88	11-93
5000	500	**97-96**	**96-96**	**96-99**	**98-98**	**97-94**	7-49	7-81	14-95	18-94	39-93
		4-class									
50	5	3-1	0-1	0	0	0	0	0-1	0	0	0
100	10	1-3	1-0	0-2	0-2	0-7	0	0	7-0	0	0
200	20	1-1	1-3	0-19	1-35	12-56	0	0	0	0	0
300	30	1-1	0-4	6-58	5-63	46-80	0	0	0	0	0
400	40	2-1	1-13	11-86	10-73	66-83	0	5-0	0	0	0
500	50	2-3	6-19	24-92	16-83	**85-82**	0	0	0	0	0
1000	100	2-3	14-69	68-89	**83-90**	**91-89**	0	0	0	0	0
2000	200	2-2	46-93	**97-97**	**83-93**	**91-90**	0	0	0	0	0-5
5000	500	3-6	**90-93**	**94-95**	**89-91**	**88-93**	0	7-1	0-8	0-23	0-73

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively. N_c denotes the smallest class size for unequal class proportion conditions.

Table 4-5. Correct class extraction rates based on entropy for dichotomous indicators

	Number of indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	24-26	51-14	46-24	59-36	51-50	44-38	42-43	32-36	47-39	42-39
100	34-15	50-18	56-41	75-57	84-79	51-40	44-43	37-40	50-35	55-36
200	30-16	56-24	82-60	90-74	92-66	48-32	45-40	53-44	48-34	55-19
300	44-16	55-35	80-58	78-70	**92-89**	52-31	42-41	39-22	41-16	50-12
400	42-12	45-40	74-61	88-78	**94-84**	56-37	54-30	45-22	46-7	45-5
500	32-16	41-40	77-61	89-77	**95-88**	56-33	47-25	41-24	54-7	44-6
1000	43-19	42-36	76-64	81-78	**89-87**	44-13	45-15	44-5	58-2	44-6
2000	47-23	31-38	78-45	85-75	**91-81**	46-7	52-2	49-1	47-2	43-3
5000	34-18	36-27	74-56	72-70	87-73	44-2	40-2	42-6	28-2	27-9
	3-class									
50	37-20	41-23	20-23	58-26	15-23	27-28	30-24	27-17	22-17	31-16
100	20-34	44-20	23-8	60-5	19-9	28-40	28-27	29-31	27-25	19-17
200	24-21	35-5	37-0	68-10	56-17	24-29	30-25	23-33	21-21	31-19
300	28-18	32-2	43-3	82-19	48-18	29-27	31-33	26-28	24-23	28-24
400	18-12	41-2	47-3	85-31	47-13	27-26	26-19	23-32	19-16	28-19
500	26-10	34-1	49-5	84-40	46-24	30-26	29-28	22-22	20-25	23-21
1000	25-5	56-3	58-15	85-43	47-30	29-25	31-22	29-25	29-25	29-10
2000	21-2	57-3	49-23	79-42	81-35	28-31	29-23	21-6	31-7	29-5
5000	38-1	62-0	58-23	75-35	76-22	26-14	27-21	26-1	20-1	24-1
	4-class									
50	26-37	28-17	25-20	14-12	29-15	32-27	11-21	12-37	17-18	14-40
100	32-33	31-26	16-20	10-18	38-11	15-36	10-19	18-29	20-17	11-41
200	41-32	33-20	19-32	7-15	42-5	40-25	17-24	35-31	20-25	25-38
300	48-28	29-12	8-33	14-18	35-3	22-34	36-29	38-29	23-31	33-40
400	42-31	37-19	5-26	11-13	21-3	46-23	40-30	23-32	38-38	53-40
500	35-31	31-14	23-3	10-9	33-3	13-27	41-24	18-30	44-33	50-34
1000	32-29	32-11	39-0	26-11	28-15	0-27	44-17	11-17	73-30	20-26
2000	30-34	26-3	42-2	26-13	22-22	36-26	22-29	25-30	31-27	55-28
5000	32-42	26-1	37-1	28-14	25-12	43-33	27-31	30-23	45-14	26-2

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-6. Correct class extraction rates based on χ^2 for dichotomous indicators

	Number of indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	27-72	74-85	3-84	0	0	2-7	7-5	5-6	0	0
100	38-93	**84-95**	30-93	0-52	0	5-10	5-6	4-5	0	0
200	47-95	**94-96**	79-88	0-100	0	5-17	6-8	2-5	0-2	0
300	70-95	**95-99**	**90-89**	0-100	0	8-20	6-13	5-5	1-7	0
400	70-96	**95-99**	**92-91**	0-100	0	5-22	5-24	4-10	7-13	0
500	67-93	**96-97**	**95-90**	0-95	0-27	9-27	4-27	4-10	17-6	0
1000	**87-94**	**91-96**	**95-94**	17-88	0-100	11-57	7-61	6-48	29-13	0
2000	**92-94**	**93-94**	**95-98**	**83-93**	0-100	71-92	16-92	12-89	52-47	0
5000	**88-95**	**93-96**	**93-93**	**89-92**	0-100	44-95	50-95	56-95	80-92	87-54
	3-class									
50	12-11	4-11	10-17	0-1	0-1	1-2	3-0	2-3	0	0
100	12-21	9-20	31-31	0	0	0-1	2-2	1-2	0	0
200	17-48	16-76	45-76	2-59	1-0	0	1-2	3-1	4-0	2-0
300	27-75	17-94	43-94	6-83	1-1	1-0	0	0-4	8-4	2-0
400	35-83	30-90	39-95	34-87	0	1-3	1-2	0-2	5-2	1-0
500	41-87	37-94	61-95	52-87	3-0	0-2	1-2	1-2	6-3	0
1000	76-94	78-95	**85-93**	67-90	0-18	1-4	0-3	3-4	4-3	0
2000	**92-92**	**88-96**	**92-94**	**89-89**	4-100	1-6	0-7	3-7	4-5	2-1
5000	**96-90**	**96-91**	**91-91**	**95-92**	**99-87**	2-24	2-36	0-44	3-19	3-6
	4-class									
50	2-4	5-1	1-3	0	2-56	0	0-1	0	0	0
100	4-13	1-4	4-3	5-0	4-3	0	0-1	8-1	0	0
200	5-9	4-3	10-10	0-2	1-4	0-1	0-1	0-1	13-2	0
300	8-8	1-8	9-13	5-10	2-0	0-2	0-1	0	0-1	0
400	7-8	4-11	9-23	14-23	0	0	0-1	0-1	0-1	0
500	5-8	11-17	13-35	23-34	0	0-1	0	0-2	24-2	0
1000	13-10	21-35	40-92	68-79	0	0	0	0-2	0	0-2
2000	10-4	47-85	73-100	78-86	0-11	0	0-1	0-2	0-1	0-3
5000	2-3	**92-96**	**94-97**	**92-92**	**99-94**	0	0-4	12-0	0-3	0-1

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-7. Correct class extraction rates based on G^2 for dichotomous indicators

	Number of indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	10-73	7-90	0	0	0	4-12	3-3	0	0	0
100	24-85	52-91	0	0	0	5-11	9-11	0	0	0
200	33-90	**80-83**	0-100	0	0	7-16	10-20	0	0	0
300	68-92	**87-84**	0-100	0	0	6-21	6-22	0-1	0	0
400	70-94	**94-92**	26-100	0	0	6-25	8-24	0-13	0	0
500	64-90	**93-86**	74-100	0	0	8-28	7-29	7-40	0	0
1000	**86-92**	**92-90**	**97-80**	0	0	9-58	9-59	18-46	0	0
2000	**92-94**	**91-94**	64-43	0-100	0	16-90	18-90	26-61	0	0
5000	**87-95**	**90-95**	58-65	**100**	0	44-95	50-95	55-83	40-13	0
	3-class									
50	12-17	0-6	1-1	0-1	0-1	4-3	4-1	0	0	0
100	11-24	7-35	1-1	0	0	2-3	1-4	0	0	0
200	14-44	18-75	0	0-1	1-0	0-1	3-4	0	1-0	2-0
300	22-74	21-85	0-59	0	1-1	1-1	2-1	0	0	2-0
400	31-78	30-82	0-99	0-1	0	1-2	4-2	1-4	0	1-0
500	36-86	30-93	5-93	0	3-0	0-2	2-2	4-13	1-0	0
1000	72-93	79-94	70-39	0	0	1-4	0-5	8-10	0	0
2000	**90-92**	**87-96**	47-47	0-100	0	2-5	2-5	9-27	0	1-0
5000	**96-90**	**96-89**	82-72	100-76	1-0	2-24	1-32	4-52	1-0	0
	4-class									
50	1-7	0	0-1	0	2-67	0	0	0	0	0
100	5-14	2-5	0-1	5-0	4-3	0-1	0-3	0	0	0
200	9-8	8-15	0-1	0-2	1-4	0-1	0-1	0	0-1	0
300	8-9	7-20	1-1	1-0	2-0	0-2	0-1	0	0	0
400	10-7	7-18	0-9	0	0	0-1	0	5-5	0	0
500	7-8	8-25	2-56	0-1	0	0-1	0-1	0-8	14-0	0
1000	12-10	21-36	39-48	0	0	0	0	0-5	0	0
2000	4-11	46-85	46-61	0	0	0	0-1	18-5	0	0
5000	6-2	**92-96**	78-89	82-54	0	0-1	0-3	28-6	0	0

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively.

Effects of manipulated factors on correct class extraction rates

In order to examine the effects of the five manipulated factors on correct class extractions, ANOVA was conducted on the logit transformed correct class extraction rates based on the three best performing fit indices (i.e., BLRT, ABIC, and BIC). In order to keep interpretations within a manageable scope in the case of a total of 30 effect terms (5 main effects and 25 interaction effects excluding the five-way interaction), effect sizes based on eta squared were examined instead of statistical significance tests. Eta squared measures the proportion of variance in the dependent variable attributable to a specific term. Using Cohen's (1988) criteria, terms of at least moderate effects (with eta squared larger than .06) were further interpreted.

Table 4-8 reported ANOVA results for BLRT-based correct extraction rates. Class separation showed the largest effect on correct class extraction rates, and accounted for about a third of the total variance in the transformed correct class extraction rates. On average, large class separation conditions had correct class extraction rates of .66, in contrast to only .18 for low class separation conditions (see left panel of Figure 4-2). Focusing on number of cells with acceptable class extraction rates, large class separation conditions had 68 such cells, compared to only 5 cells for small class separation conditions.

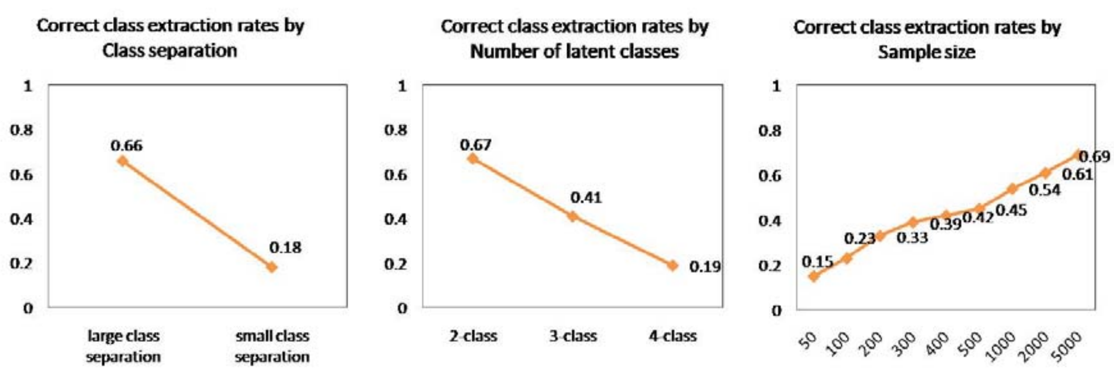

Figure 4-2. Main effects of class separation, number of latent classes and sample sizes on correct class extraction rates for dichotomous indicators

Table 4-8. ANOVA table for logit transformed BLRT-based correct class extraction rates: dichotomous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	**1963.34**	**2**	**1091.00**	**.00**	**.29**
SS(sample size)	**890.23**	**8**	**123.67**	**.00**	**.13**
ND(#indicators)	203.61	4	56.57	.00	.03
CS(class separation)	**2180.96**	**1**	**2423.85**	**.00**	**.32**
CP(class proportion)	204.95	1	227.78	.00	.03
NC×SS	39.63	16	2.75	.00	.01
NC×ND	30.07	8	4.18	.00	.00
NC×CS	87.79	2	48.79	.00	.01
NC×CP	19.48	2	10.83	.00	.00
SS×ND	30.99	32	1.08	.39	.00
SS×CS	140.49	8	19.52	.00	.02
SS×CP	3.20	8	0.44	.89	.00
ND×CS	44.51	4	12.37	.00	.01
ND×CP	6.29	4	1.75	.15	.00
CS×CP	4.81	1	5.35	.02	.00
NC×SS×ND	64.63	64	1.12	.32	.01
NC×SS×CS	240.71	16	16.72	.00	.04
NC×SS×CP	24.47	16	1.70	.07	.00
SS×ND×CS	36.05	32	1.25	.22	.01
SS×ND×CP	11.52	32	0.40	.99	.00
ND×CS×CP	5.28	4	1.47	.22	.00
NC×ND×CS	56.78	8	7.89	.00	.01
NC×ND×CP	7.90	8	1.10	.38	.00
NC×CS×CP	37.32	2	20.74	.00	.01
SS×CS×CP	81.03	8	11.26	.00	.01
NC×SS×ND×CS	116.00	64	2.01	.00	.02
NC×SS×ND×CP	42.30	64	0.73	.89	.01
NC×ND×CS×CP	14.61	8	2.03	.06	.00
SS×ND×CS×CP	9.66	32	0.34	.99	.00
NC×SS×CS×CP	77.98	16	5.42	.00	.01
Error	57.59	64			
Correct Total	6734.20				

True number of latent classes showed the second largest effect on correct class extraction rates, explaining 29% of the total variance in the transformed correct class extraction rates. This effect was evident from Table 4-4: in general, correct class extraction rates and number of bolded cells decreased with increasing number of true latent classes. Average correct class rates went from .67 for 2-class models, to .41 for 3-class models, and then further down to .18 for 4-class models (shown in the middle panel of Figure 4-2). Number of bolded cells decreased from 41 for 2-class models to only 9 for 4-class models.

In addition to class separation and true number of latent classes, sample size also showed a moderate effect, accounting for 13% of the total variance. As shown in the right panel of Figure 4-2, average class extraction rates at the smallest (i.e., 50) and largest sample size levels (i.e., 5000) were .15 and .69 respectively.

Table 4-9. ANOVA table for logit transformed ABIC-based correct class extraction rates: dichotomous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	**1232.41**	**2**	**652.29**	**0.00**	**0.17**
SS(sample size)	**664.21**	**8**	**87.89**	**0.00**	**0.09**
ND(#indicators)	139.26	4	36.85	0.00	0.02
CS(class separation)	**2307.83**	**1**	**2442.99**	**0.00**	**0.33**
CP(class proportion)	139.36	1	147.52	0.00	0.02
NC×SS	240.54	16	15.91	0.00	0.03
NC×ND	23.92	8	3.17	0.00	0.00
NC×CS	152.07	2	80.49	0.00	0.02
NC×CP	38.54	2	20.40	0.00	0.01
SS×ND	263.82	32	8.73	0.00	0.04
SS×CS	**720.35**	**8**	**95.32**	**0.00**	**0.10**
SS×CP	63.07	8	8.35	0.00	0.01
ND×CS	29.67	4	7.85	0.00	0.00
ND×CP	6.57	4	1.74	0.15	0.00
CS×CP	16.91	1	17.90	0.00	0.00
NC×SS×ND	66.67	64	1.10	0.35	0.01
NC×SS×CS	125.55	16	8.31	0.00	0.02
NC×SS×CP	76.82	16	5.08	0.00	0.01
SS×ND×CS	63.53	32	2.10	0.01	0.01
SS×ND×CP	14.41	32	0.48	0.99	0.00
ND×CS×CP	25.98	4	6.87	0.00	0.00
NC×ND×CS	60.27	8	7.98	0.00	0.01
NC×ND×CP	29.33	8	3.88	0.00	0.00
NC×CS×CP	82.71	2	43.78	0.00	0.01
SS×CS×CP	48.39	8	6.40	0.00	0.01
NC×SS×ND×CS	163.08	64	2.70	0.00	0.02
NC×SS×ND×CP	69.77	64	1.15	0.28	0.01
NC×ND×CS×CP	38.69	8	5.12	0.00	0.01
SS×ND×CS×CP	20.30	32	0.67	0.89	0.00
NC×SS×CS×CP	71.04	16	4.70	0.00	0.01
Error	60.46	64			
Correct Total	7055.56	539			

Table 4-10. ANOVA table for logit transformed BIC-based correct class extraction rates: dichotomous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	**1536.49**	**2**	**478.34**	**.00**	**.16**
SS(sample size)	**930.01**	**8**	**72.38**	**.00**	**.10**
ND(#indicators)	310.00	4	48.25	.00	.03
CS(class separation)	**3540.56**	**1**	**2204.48**	**.00**	**.38**
CP(class proportion)	331.24	1	206.24	.00	.04
NC×SS	26.40	16	1.03	.44	.00
NC×ND	5.76	8	0.45	.89	.00
NC×CS	535.28	2	166.64	.00	.06
NC×CP	63.58	2	19.79	.00	.01
SS×ND	43.18	32	0.84	.70	.00
SS×CS	284.07	8	22.11	.00	.03
SS×CP	13.20	8	1.03	.43	.00
ND×CS	139.43	4	21.70	.00	.01
ND×CP	2.80	4	0.44	.78	.00
CS×CP	108.65	1	67.65	.00	.01
NC×SS×ND	91.31	64	0.89	.68	.01
NC×SS×CS	383.42	16	14.92	.00	.04
NC×SS×CP	76.83	16	2.99	.00	.01
SS×ND×CS	23.11	32	0.45	.99	.00
SS×ND×CP	13.00	32	0.25	.00	.00
ND×CS×CP	4.67	4	0.73	.58	.00
NC×ND×CS	54.63	8	4.25	.00	.01
NC×ND×CP	15.82	8	1.23	.30	.00
NC×CS×CP	119.21	2	37.11	.00	.01
SS×CS×CP	179.71	8	13.99	.00	.02
NC×SS×ND×CS	157.44	64	1.53	.05	.02
NC×SS×ND×CP	118.30	64	1.15	.29	.01
NC×ND×CS×CP	30.78	8	2.40	.03	.00
SS×ND×CS×CP	27.96	32	0.54	.97	.00
NC×SS×CS×CP	171.17	16	6.66	.00	.02
Error	102.79	64			
Correct Total	9440.79	539			

ANOVA results for ABIC and BIC displayed largely similar patterns to those for BLRT, however, ABIC had one additional term that also showed a moderate effect size. For ABIC, in addition to the number of true latent classes, sample size, and class separation, the interaction between sample size and class separation also showed a moderate effect with eta squared of .10. For large class separation, correct class extraction rates based on ABIC steadily increased with larger sample sizes; for small class separation, the effect of sample size was much less consistent, although the general trend still seemed to be increasing with increasing sample size.

In sum, correct class extraction rates were higher when classes were more distant from one another, number of true latent classes was smaller, and sample sizes were larger. This pattern appeared to hold for all three best performing fit indices (BLRT, ABIC, and BIC), although the sample size effect was less consistent in the small class separation conditions for ABIC. The two remaining main effects (i.e., number of indicators and class proportion) demonstrated only small effects on correct class extraction rates, each accounting for 2 to 4% of the total variance. Correct class extraction rates were larger for conditions with more indicators and equal class proportion.

Classification accuracy

This section reports the overall classification accuracy and group classification accuracy under different studied conditions when the correct number of latent classes is specified. Tables 4-11 and 4-13 list overall classification accuracy and group classification accuracy for dichotomous indicators.

Overall classification accuracy

Results from ANOVA (Table 4-12) indicated that class separation was the most influential factor on overall classification accuracy, accounting for 43% of the total variance in the logit transformed overall classification accuracy. Average overall classification accuracy was .81 for large class separation and .54 for low class separation. This effect is notable from Table 4-11: for large class separation shown in the left panel, there were 69 bolded cells with overall classification accuracy rates at or above .80 out of a total of 135 cells; in contrast, no cells achieved the same level of overall classification accuracy for small class separation conditions.

Table 4-11. Overall classification accuracy for dichotomous indicators

		Number of indicators									
		5	6	9	12	15	5	6	9	12	15
		Large class separation					Small class separation				
N	N_c	2-class									
50	5	.67-.83	**.85-.88**	.72-.92	.83-.97	.75-.97	.63-.59	.51-.59	.47-.63	.57-.63	.49-.65
100	10	.71-.86	**.87-.89**	**.84-.94**	**.96-.97**	**.91-.98**	.53-.60	.65-.60	.53-.62	.60-.65	.59-.65
200	20	.73-.88	**.94-.90**	**.96-.95**	**.98-.97**	**.97-.98**	.60-.60	.51-.60	.59-.64	.53-.65	.57-.71
300	30	**.83-.89**	**.95-.90**	**.97-.95**	**.98-.97**	**.99-.98**	.56-.60	.53-.61	.49-.64	.55-.68	.56-.73
400	40	**.84-.89**	**.95-.90**	**.97-.95**	**.98-.97**	**.99-.98**	.60-.62	.60-.65	.58-.65	.54-.70	.78-.75
500	50	**.86-.89**	**.95-.90**	**.97-.95**	**.98-.97**	**.99-.98**	.44-.60	.75-.62	.60-.66	.52-.71	.71-.75
1000	100	**.90-.93**	**.95-.90**	**.97-.95**	**.98-.97**	**.99-.98**	.88-.61	.61-.62	.73-.69	.74-.74	.83-.78
2000	200	**.90-.94**	**.95-.90**	**.97-.95**	**.98-.97**	**.99-.98**	.52-.62	.87-.65	.77-.71	.90-.75	.91-.78
5000	500	**.90-.94**	**.95-.90**	**.98-.95**	**.98-.97**	**.99-.98**	.71-.65	.78-.67	.90-.73	.91-.76	.91-.79
		3-class									
50	5	.75-.61	.80-.66	.84-.75	**.87-.80**	**.87-.81**	.59-.43	.59-.45	.63-.45	.68-.44	.70-.47
100	10	.76-.62	.80-.68	.85-.79	**.89-.85**	**.89-.89**	.57-.42	.60-.43	.65-.43	.69-.44	.73-.46
200	20	.77-.66	.72-.81	**.87-.87**	**.87-.91**	**.93-.91**	.59-.41	.59-.42	.68-.43	.71-.44	.73-.45
300	30	.77-.68	.83-.78	**.89-.83**	**.92-.88**	**.94-.91**	.57-.42	.58-.41	.68-.42	.73-.43	.73-.45
400	40	.79-.69	.84-.75	**.90-.83**	**.93-.88**	**.95-.92**	.58-.42	.62-.41	.68-.43	.73-.43	.75-.46
500	50	.80-.70	.84-.75	**.90-.83**	**.93-.88**	**.95-.92**	.56-.41	.63-.41	.68-.42	.75-.45	.76-.47
1000	100	.83-.73	.86-.75	**.90-.83**	**.93-.88**	**.95-.91**	.56-.41	.63-.42	.73-.43	.75-.46	.77-.50
2000	200	.85-.74	.87-.75	**.91-.83**	**.95-.91**	**.95-.91**	.52-.41	.61-.42	.73-.45	.78-.51	.77-.55
5000	500	.87-.75	.87-.75	**.91-.83**	**.93-.88**	**.95-.92**	.47-.43	.62-.42	.72-.49	.76-.55	.80-.59
		4-class									
50	5	.61-.46	.58-.52	.67-.56	.67-.59	.74-.61	.37-.37	.35-.34	.43-.35	.56-.37	.47-.37
100	10	.61-.45	.60-.53	.67-.58	.72-.61	.72-.67	.43-.33	.37-.32	.43-.34	.54-.34	.51-.35
200	20	.64-.47	.61-.55	.70-.63	.75-.66	.81-.77	.45-.32	.44-.32	.47-.33	.52-.33	.52-.33
300	30	.63-.47	.62-.56	.72-.66	.76-.69	.84-.79	.48-.2	.48-.32	.50-.32	.50-.32	.59-.33
400	40	.62-.48	.63-.52	.74-.67	.77-.72	**.85-.81**	.45-.31	.44-.31	.50-.31	.53-.33	.62-.33
500	50	.63-.48	.64-.56	.74-.69	.78-.74	**.86-.82**	.62-.33	.40-.31	.54-.32	.53-.31	.62-.33
1000	100	.64-.49	.66-.60	.79-.72	.80-.76	**.88-.84**	.43-.31	.35-.31	.57-.31	.63-.32	.63-.33
2000	200	.65-.50	.68-.61	.82-.72	.80-.77	**.88-.84**	.49-.31	.45-.31	.55-.31	.57-.32	.67-.34
5000	500	.65-.51	.69-.63	.81-.73	.80-.77	**.89-.84**	.44-.32	.59-.31	.44-.32	.63-.34	.64-.39

Note. Values to the left and right sides of "-" represent classification accuracy rates for unequal and equal class proportions respectively. N_c indicates the smallest class size in the true model.

Table 4-12. ANOVA table for logit transformed overall classification accuracy: dichotomous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	**227.20**	**2**	**3228.82**	**.00**	**.27**
SS(sample size)	24.68	8	87.68	.00	.03
ND(#indicators)	**72.15**	**4**	**512.64**	**.00**	**.09**
CS(class separation)	**358.36**	**1**	**10185.40**	**.00**	**.43**
CP(class proportion)	27.72	1	787.99	.00	.03
NC×SS	6.56	16	11.66	.00	.01
NC×ND	2.63	8	9.34	.00	.00
NC×CS	25.98	2	369.22	.00	.03
NC×CP	11.51	2	163.59	.00	.01
SS×ND	3.77	32	3.35	.00	.00
SS×CS	5.11	8	18.15	.00	.01
SS×CP	5.76	8	20.46	.00	.01
ND×CS	18.49	4	131.39	.00	.02
ND×CP	0.35	4	2.51	.05	.00
CS×CP	1.59	1	45.22	.00	.00
NC×SS×ND	2.40	64	1.07	.40	.00
NC×SS×CS	2.85	16	5.07	.00	.00
NC×SS×CP	5.82	16	10.34	.00	.01
SS×ND×CS	2.43	32	2.16	.00	.00
SS×ND×CP	1.33	32	1.18	.28	.00
ND×CS×CP	0.97	4	6.90	.00	.00
NC×ND×CS	2.10	8	7.45	.00	.00
NC×ND×CP	1.63	8	5.80	.00	.00
NC×CS×CP	1.35	2	19.15	.00	.00
SS×CS×CP	0.48	8	1.72	.11	.00
NC×SS×ND×CS	2.75	64	1.22	.21	.00
NC×SS×ND×CP	2.48	64	1.10	.35	.00
NC×ND×CS×CP	1.40	8	4.97	.00	.00
SS×ND×CS×CP	1.31	32	1.16	.30	.00
NC×SS×CS×CP	4.09	16	7.26	.00	.00
Error	2.25	64			
Correct Total	827.52				

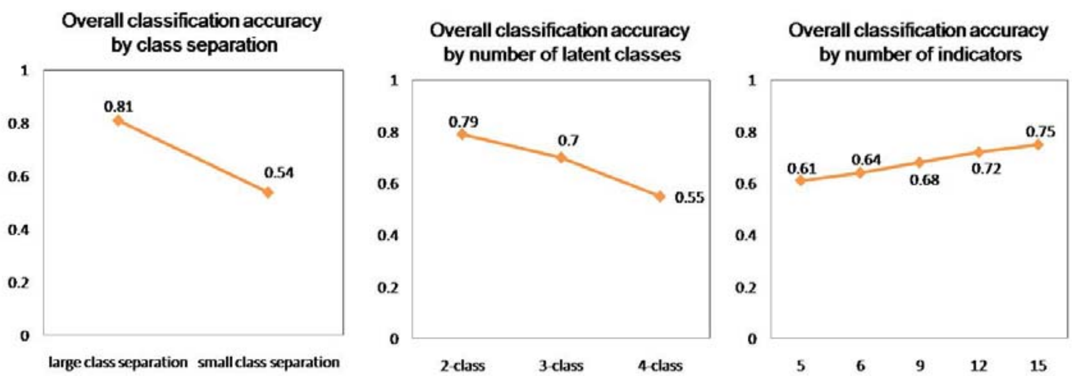

Figure 4-3. Main effects of class separation, number of latent classes, and number of indicators on overall classification accuracy for dichotomous indicators

True number of latent classes contributed an additional 27% of the total variance, with increased number of true number of latent classes associated with decreased overall classification accuracy: average overall classification accuracy was .79 for 2-class models, .70 for 3-class models, and .55 for 4-class models.

Number of indicators was the third largest factor, explaining 9% of the total variance in the transformed overall classification accuracy. Table values suggested a general pattern of positive effects from a larger number of indicators. Average overall correct classification rates ranged from .60 for 5 indicators to .75 for 15 indicators. The main effects of the three statistically significant factors are also presented in Figure 4-3.

In sum, overall classification accuracy tended to be higher for larger class separation, fewer latent classes, and more indicators. Note that average overall classification accuracy rates were indeed higher for unequal class proportion conditions (.72) than for equal class proportion conditions (.63), however, the effect size was small with eta squared of 3%. This suggested that individuals were more likely to be assigned to the larger group in the case of unequal class proportions. The remaining main effect, sample size, also had an eta squared of 3%.

Group classification accuracy

Table 4-14 reports results from ANOVA on group classification accuracy for dichotomous indicators. Consistent with results from overall classification accuracy, class separation and true number of latent classes continued to show the largest effects, accounting for 26% and 35% of the total variance respectively. Average group classification accuracy for small and large class separation conditions was .35 and .66 respectively. For models with 2, 3, and 4 latent classes, average group classification accuracy was .71 and .50 and .31.

Table 4-13. Lowest group classification accuracy for dichotomous indicators

		Number of indicators									
		5	6	9	12	15	5	6	9	12	15
		Large class separation					Small class separation				
N	N_c	2-class									
50	5	**.67-.79**	**.78-.88**	**.71-.91**	**.83-.97**	**.74-.97**	.51-.56	.50-.55	.46-.61	.55-.63	.47-.64
100	10	**.71-.86**	**.75-.89**	**.84-.94**	**.91-.97**	**.91-.97**	.52-.53	.45-.59	.53-.57	.44-.64	.59-.60
200	20	**.72-.88**	**.71-.90**	**.83-.94**	**.90-.97**	**.93-.98**	.47-.54	.50-.60	.58-.59	.51-.65	.57-.68
300	30	**.75-.89**	**.71-.89**	**.82-.95**	**.90-.97**	**.94-.98**	.57-.59	.51-.59	.46-.59	.54-.68	.56-.70
400	40	**.74-.89**	**.70-.89**	**.82-.95**	**.90-.97**	**.94-.98**	.50-.56	.44-.55	.58-.65	.54-.70	.39-.74
500	50	**.74-.89**	**.69-.90**	**.82-.95**	**.90-.97**	**.94-.98**	.42-.58	.38-.57	.55-.63	.50-.70	.51-.75
1000	100	**.67-.89**	**.69-.90**	**.83-.95**	**.90-.97**	**. 95-.98**	.20-.56	.54-.61	.36-.68	.41-.74	.40-.77
2000	200	**.66-.89**	**.69-.89**	**.84-.95**	**.90-.97**	**.95-.98**	.51-.60	.13-.65	.39-.69	.20-.75	.32-.78
5000	500	**.66-.89**	**.71-.90**	**.84-.95**	**.91-.97**	**. 95-.98**	.45-.65	.26-.67	.17-.71	.21-.75	.27-.79
		3-class									
50	5	.37-.51	.42-.61	**.50-.73**	**.57-.73**	**.45-.77**	.29-.40	.31-.42	.28-.43	.24-.43	.22-.39
100	10	.45-.54	.41-.67	**.55-.74**	**.65-.84**	**.48-.88**	.28-.40	.28-.40	.21-.37	.23-.41	.17-.43
200	20	.43-.61	**.47-.72**	**.62-.80**	**.73-.87**	**.77-.90**	.25-.38	.29-.41	.20-.39	.15-.39	.15-.43
300	30	.45-.62	**.45-.72**	**.62-.82**	**.74-.87**	**.80-.91**	.26-.36	.28-.35	.16-.39	.16-.40	.15-.42
400	40	.44-.64	**.50-.74**	**.65-.82**	**.74-.87**	**.81-.91**	.25-.39	.21-.38	.16-.41	.17-.42	.14-.44
500	50	.44-.64	**.50-.73**	**.50-.73**	**.75-.88**	**.82-.91**	.28-.36	.22-.40	.16-.41	.13-.39	.14-.43
1000	100	**.48-.71**	**.50-.74**	**.50-.74**	**.76-.88**	**.82-.91**	.28-.35	.22-.40	.12-.40	.13-.44	.14-.44
2000	200	.45-.71	**.49-.73**	**.65-.83**	**.76-.88**	**.82-.91**	.31-.33	.27-.35	.14-.41	.13-.50	.17-.53
5000	500	.43-.71	**.50-.74**	**.66-.83**	**.77-.88**	**.83-.91**	.32-.37	.17-.40	.10-.47	.16-.53	.18-.59
		4-class									
50	5	.06-.09	**.38-.41**	**.39-.45**	.31-.51	.29-.49	.25-.26	.20-.28	.30-.31	.17-.32	.21-.33
100	10	.05-.08	**.37-.41**	**.40-.47**	.32-.51	**.38-.60**	.17-.26	.20-.30	.18-.27	.06-.31	.14-.31
200	20	.04-.06	.34-.49	**.43-.50**	**.43-.64**	**.50-.77**	.13-.29	.16-.27	.15-.28	.09-.30	.15-.31
300	30	.05-.06	.30-.49	**.42-.58**	**.47-.65**	**.62-.76**	.14-.27	.18-.30	.11-.29	.03-.30	.09-.26
400	40	.05-.06	.35-.53	**.45-.63**	**.49-.70**	**.63-.79**	.12-.25	.09-.25	.10-.29	.09-.25	.11-.25
500	50	.04-.04	.36-.47	**.48-.66**	**.45-.70**	**.64-.79**	.14-.28	.10-.28	.08-.29	.12-.25	.05-.30
1000	100	.02-.04	**.37-.53**	**.54-.69**	**.52-.74**	**.67-.82**	.16-.28	.19-.24	.07-.28	.05-.28	.03-.32
2000	200	.02-.03	**.40-.55**	**.54-.69**	**.47-.76**	**.70-.82**	.10-.28	0-.23	.09-.27	.10-.31	.04-.31
5000	500	.03-.04	**.39-.58**	**.51-.69**	**.59-.77**	**.70-.82**	.07-.20	.07-.12	.16-.29	.06-.32	.02-.37

Note. Values to the left and right sides of "-" represent classification accuracy rates for unequal and equal class proportions respectively. N_c indicates the smallest class size in the true model.

Class proportion had the third largest effect on group classification accuracy, explaining 10% of the total variance. Contrary to a seemingly adverse effect of equal class proportions on overall classification accuracy, the smallest group classification rates were smaller for unequal class proportion conditions than for equal class proportion conditions. The pattern of smaller group classification accuracy rates for smaller classes was consistent with what was observed in previous study with logistic regression (e.g., Finch & Schneider, 2006). On average, smallest group classification rates for equal and unequal class proportions were .59 and .42 respectively. This pointed to the fact that although overall classification appeared to be higher for unequal class proportions, equal class proportion should be the preferred condition for classification purposes because group classification rates may be vastly different in the unequal class proportion conditions. In Table 4-12, cells were bolded when the smallest group classification accuracy rates were statistically significantly higher than random classification rates. Many cells in the large class separation conditions showed performance superior to random classification. Note that the smallest group classification rates were rarely statistically significantly higher than random classification when unequal class proportion condition was combined with small class separation condition.

Table 4-14. ANOVA table for logit transformed group classification accuracy: dichotomous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	**471.88**	**2**	**3479.62**	**.00**	**.35**
SS(sample size)	1.58	8	2.91	.01	.00
ND(#indicators)	80.76	4	297.77	.00	.06
CS(class separation)	**343.35**	**1**	**5063.76**	**.00**	**.26**
CP(class proportion)	**144.36**	**1**	**2128.98**	**.00**	**.11**
NC×SS	2.65	16	2.44	.01	.00
NC×ND	29.66	8	54.69	.00	.02
NC×CS	28.53	2	210.40	.00	.02
NC×CP	0.19	2	1.40	.25	.00
SS×ND	7.81	32	3.60	.00	.01
SS×CS	16.17	8	29.82	.00	.01
SS×CP	4.02	8	7.41	.00	.00
ND×CS	85.63	4	315.71	.00	.06
ND×CP	11.14	4	41.06	.00	.01
CS×CP	0.02	1	0.23	.63	.00
NC×SS×ND	9.52	64	2.19	.00	.01
NC×SS×CS	4.46	16	4.11	.00	.00
NC×SS×CP	2.67	16	2.46	.01	.00
SS×ND×CS	6.54	32	3.01	.00	.00
SS×ND×CP	3.08	32	1.42	.12	.00
ND×CS×CP	1.52	4	5.61	.00	.00
NC×ND×CS	30.17	8	55.62	.00	.02
NC×ND×CP	5.48	8	10.09	.00	.00
NC×CS×CP	2.94	2	21.69	.00	.00
SS×CS×CP	6.68	8	12.31	.00	.01
NC×SS×ND×CS	7.96	64	1.83	.01	.01
NC×SS×ND×CP	7.45	64	1.72	.02	.01
NC×ND×CS×CP	5.09	8	9.38	.00	.00
SS×ND×CS×CP	5.22	32	2.40	.00	.00
NC×SS×CS×CP	1.93	16	1.78	.05	.00
Error	4.34	64			
Correct Total	1332.78				

Number of indicators explained 6% of the total variance, with average group classification accuracy ranging from .60 to .75 for the 5- to 15-indicator conditions. In sum, group classification accuracy tended to be higher for larger class separation, fewer latent classes, equal class proportions, and more indicators. Sample sizes had a negligible effect though, contributing to less than 1% of the total variance, indicating that it might be inefficient to increase group classification accuracy by increasing total sample size. The main effects of statistically significant effects are displayed in Figure 4-4.

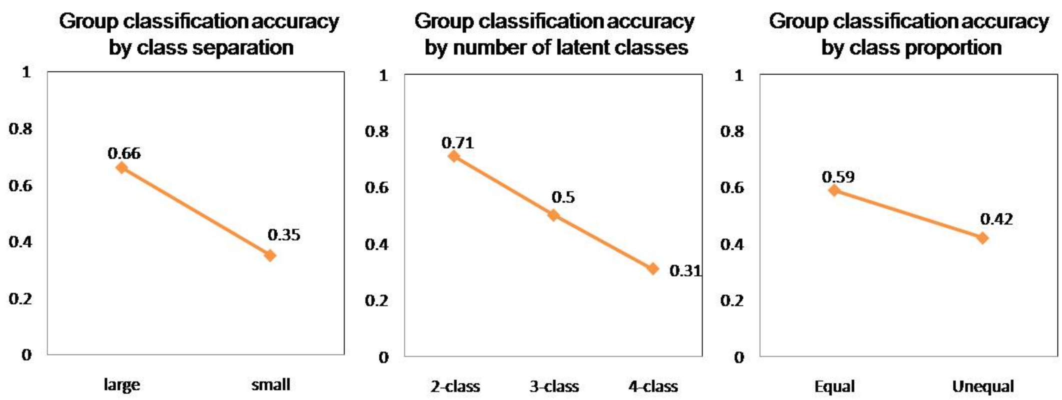

Figure 4-4. Main effects of class separation, number of latent classes and class proportion on group classification accuracy for dichotomous indicators

Entropy and classification accuracy

In applied research, there is no way to know the true overall or group classification accuracy; researchers have been using entropy to indicate the certainty of class membership assignment in LCM. Table 4-15 reports the average entropy values over the 100 replications for each studied condition. Entropy values were relatively homogeneous for the 100 replications within each simulation condition. Most of the standard deviations were smaller than .05; standard deviations tended to be larger with smaller sample size conditions, but rarely exceeded .10. Cells with high entropy values (i.e., ≥ .80) were bolded to indicate high entropy values. In the meanwhile, low entropy values (i.e.,< .60) were italicized.

Table 4-15. Average entropy for dichotomous indicators

	Number of indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	.75-.81	**.88-.82**	**.84-.89**	**.92-.93**	**.90-.96**	.78-.78	**.81-.81**	**.82-.82**	**.88-.87**	**.88-.88**
100	.73-.71	.86-.75	**.84-.84**	**.95-.91**	**.95-.95**	.77-.74	.77-.77	.76-.74	.82-.76	.82-.78
200	.66-.66	.87-.71	**.91-.83**	**.95-.90**	**.97-.94**	.74-.66	.72-.68	.73-.65	.79-.65	.80-.61
300	.74-.64	.85-.70	**.91-.82**	**.95-.89**	**.97-.94**	.75-.61	.71-.63	.69-.75	.72-*.53*	.77-*.52*
400	.73-.64	.85-.69	**.91-.82**	**.94-.89**	**.97-.94**	.75-.60	.74-*.58*	.72-*.50*	.72-*.46*	.76-*.45*
500	.70-.63	.85-.69	**.90-.82**	**.94-.89**	**.97-.94**	.73-*.56*	.72-*.53*	.74-*.46*	.75-*.43*	.72-*.44*
1000	.77-.63	.84-.69	**.90-.81**	**.94-.89**	**.96-.93**	.69-*.40*	.68-*.39*	.70-*.34*	.71-*.34*	.74-*.38*
2000	.78-.62	.84-.69	**.90-.81**	**.94-.89**	**.96-.93**	.71-*.29*	.67-*.27*	.66-*.27*	.67-*.32*	.74-*.37*
5000	.79-.62	.83-.68	**.90-.81**	**.94-.89**	**.96-.93**	*.59-.20*	*.58-.20*	.65-*.25*	.66-*.31*	.70-*.36*
	3-class									
50	**.88-.84**	**.88-.85**	**.91-.89**	**.95-.91**	**.94-.95**	**.83-.83**	**.86-.84**	**.88-.86**	**.88-.90**	**.91-.91**
100	.79-.78	.84-.76	.87-.79	**.92-.84**	**.90-.88**	**.80-.81**	**.80-.80**	**.82-.82**	**.84-.93**	**.83-.82**
200	.75-.66	.75-.65	.84-.71	.90-.78	**.93-.84**	.76-.75	.78-.74	.78-.74	.77-.75	.77-.76
300	.73-.62	.75-.60	.83-.68	.88-.77	**.92-.83**	.74-.73	.73-.74	.73-.71	.75-.72	.76-.72
400	.72-.57	.74-.57	.83-.67	.88-.77	**.91-.83**	.75-.71	.73-.69	.73-.71	.74-.67	.73-.65
500	.74-.55	.73-.55	.82-.66	.87-.76	**.91-.83**	.75-.70	.72-.72	.72-.66	.71-.63	.72-.61
1000	.70-.49	.73-.53	.81-.65	.86-.75	**.90-.82**	.75-.69	.74-.65	.71-.57	.71-*.52*	.66-*.48*
2000	.69-.46	.74-.51	.80-.65	.85-.75	**.89-.82**	.78-.65	.72-*.57*	.66-*.41*	.66-*.34*	.63-*.33*
5000	.67-.45	.72-.51	.79-.65	.85-.75	**.89-.82**	.73-*.53*	.67-*.41*	.60-*.25*	.55-*.24*	.60-*.25*
	4-class									
50	**.88-.89**	**.88-.88**	**.91-.92**	**.93-.93**	**.95-.95**	**.88-.87**	**.86-.88**	**.91-.91**	**.94-.93**	**.94-.94**
100	**.83-.83**	**.81-.81**	**.84-.82**	**.86-.86**	**.91-.88**	.81-.79	**.83-.82**	**.87-.83**	**.85-.86**	**.83-.85**
200	.80-.74	.77-.73	.78-.74	.76-.77	.86-.79	.80-.76	.78-.77	.80-.77	**.80-.80**	.78-.80
300	.78-.71	.74-.68	.75-.67	.73-.71	.84-.75	.78-.75	.75-.77	.74-.76	.78-.77	.73-.77
400	.74-.68	.73-.66	.72-.62	.71-.68	.82-.74	.74-.76	.74-.74	.72-.74	.75-.74	.79-.77
500	.73-.67	.71-.63	.71-.59	.69-.67	.82-.73	.78-.73	.78-.75	.71-.73	.72-.74	.80-.75
1000	.69-.62	.67-.55	.68-.55	.68-.64	.80-.71	.72-.73	.78-.73	.67-.68	.74-.67	.69-.68
2000	.67-.58	.62-.50	.66-.53	.66-.63	.95-.71	.74-.72	.73-.71	.67-.66	.61-.59	.73-.57
5000	.65-.56	.61-.49	.65-.52	.66-.62	.78-.70	.77-.70	.74-.67	.57-*.52*	.61-*.40*	.62-*.31*

Note. Values to the left and right sides of "-" represent entropy for unequal and equal class proportions respectively.

The usefulness of entropy was evaluated based on its consistency with measures of overall and group classification accuracy from Tables 4-11 and 4-13. Looking across the three tables, there was very limited evidence for such consistency. A very striking difference in the pattern between entropy and the computed classification accuracy measures was that bolded values tended to cluster in the upper right corner in each section for entropy, whereas they tended to cluster in the lower right corner in each section for Tables 4-11 and 4-13. This indicated the differential effects of sample size for entropy and the computed classification accuracy measures: entropy tended to decrease when sample sizes increased, while such pattern was reversed for classification accuracy. ANOVA on entropy (refer to Table 4-16) also revealed sample size to be the largest contributing factor to the variance in entropy, explaining 28% of its total variance. As a result, cells with large entropy values appeared to be a reflection of small sample sizes in most conditions and did not well correspond with cells with high classification accuracy measures.

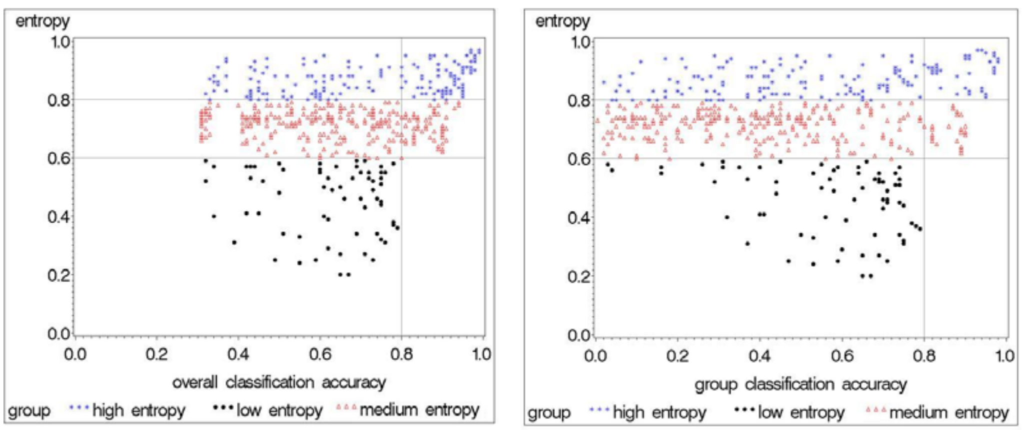

Figure 4-5. Scatterplot between entropy and overall/group classification accuracy for dichotomous indicators

Figure 4-5 presents a visual display of the lack of consistency between entropy and the actual overall (the left panel of Figure 4-4) and group classification accuracy (the right panel of Figure 4-5). Entropy is categorized into the “large”, “medium” and “low” groups using cutoffs of .80 and .60, and the three groups are indicated by the blue, red and black colors. Using the cutoff of .80 for overall and group classification, the vertical gray line divides the classification accuracy into “small” and “large” groups. Note that for each level of the entropy level, there is a wide spread of the classification accuracy: for large entropy values, overall classification accuracy ranges from around .30 to close to 1 and group classification accuracy ranges from close to 0 to 1; for small entropy values, overall classification accuracy ranges from around .30 to around .80 and group classification accuracy ranges from 0 to .80.

In addition to sample size, interaction between class separation and number of indicators and between class separation and true number of latent classes contributed to another 9% and 12% of total variance. For large class separation, entropy tended to increase as number of indicators increased, ranging from .70 for 5 indicators to .88 for 15 indicators; also, entropy tended to decrease as number of latent classes increased. However for small class separation, entropy remained largely constant at an average of .70 across different numbers of indicators; and as number of latent classes increased, entropy tended to increase as well, ranging from an average of .63 for 2-class models to .75 for 4-class models.

Class proportion showed a moderate effect with eta squared of .09. Average entropy was higher for unequal class proportions (.78) than for equal class proportions (.69). This pattern was consistent with overall classification accuracy, which was also an aggregate measure of all latent classes, rather than an individual measure for a certain latent class, such as group classification accuracy. This suggested that entropy may be more useful in indicating the overall classification accuracy than the group classification accuracy.

In sum, entropy tended to be lower with larger sample sizes and unequal class proportions. The pattern of high and low entropy values did not necessarily indicate high and low classification accuracy, especially when sample sizes were small.

Table 4-16. ANOVA table for logit transformed entropy: dichotomous Indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	1.44	2	35.65	.00	.00
SS(sample size)	**102.75**	**8**	**634.36**	**.00**	**.28**
ND(#indicators)	**31.07**	**4**	**383.58**	**.00**	**.08**
CS(class separation)	**41.60**	**1**	**2054.47**	**.00**	**.11**
CP(class proportion)	**32.32**	**1**	**1596.45**	**.00**	**.09**
NC×SS	5.87	16	18.12	.00	.02
NC×ND	9.85	8	60.84	.00	.03
NC×CS	**43.60**	**2**	**1076.67**	**.00**	**.12**
NC×CP	5.76	2	142.17	.00	.02
SS×ND	2.46	32	3.80	.00	.01
SS×CS	8.64	8	53.34	.00	.02
SS×CP	10.53	8	64.99	.00	.03
ND×CS	**33.72**	**4**	**416.38**	**.00**	**.09**
ND×CP	0.31	4	3.86	.01	.00
CS×CP	0.06	1	3.07	.08	.00
NC×SS×ND	3.21	64	2.48	.00	.01
NC×SS×CS	7.46	16	23.03	.00	.02
NC×SS×CP	2.00	16	6.17	.00	.01
SS×ND×CS	4.53	32	6.99	.00	.01
SS×ND×CP	0.66	32	1.02	.46	.00
ND×CS×CP	0.39	4	4.85	.00	.00
NC×ND×CS	6.24	8	38.55	.00	.02
NC×ND×CP	0.47	8	2.88	.01	.00
NC×CS×CP	2.69	2	66.37	.00	.01
SS×CS×CP	2.43	8	14.98	.00	.01
NC×SS×ND×CS	4.56	64	3.52	.00	.01
NC×SS×ND×CP	1.62	64	1.25	.19	.00
NC×ND×CS×CP	0.72	8	4.43	.00	.00
SS×ND×CS×CP	0.83	32	1.29	.19	.00
NC×SS×CS×CP	1.39	16	4.30	.00	.00
Error	1.30	64			
Correct Total	370.48				

Recommendations for practitioners

Since class separation and true number of latent classes have consistently shown strong effects on correct class extraction rates and classification accuracy, recommendation on sample size and number of indicators for LCM should take into account researchers' level of knowledge about those two factors.

First of all, the clear message from both analyses of correct class extraction rates and of classification accuracy is that LCM is not highly recommended when the classes are not well separated. In this simulation study, small class separation is defined as .20 difference in indicator probabilities across latent classes. In the presence of small class separation, LCM is incapable of either distinguishing among models of different numbers of latent classes (e.g., right panels of Tables 4-3 and 4-4), or correctly assigning individuals to appropriate latent classes with high level of accuracy (e.g., right panels of Tables 4-11 and 4-12). The only few situations LCM may be useful for small class separation conditions is that when the purpose is to classify individuals into two classes with relatively equal class sizes. Under such circumstances, sample sizes of at least 400 and 12 or more dichotomous indicators are recommended, which would produce overall and group classification rates of .70 or above.

LCM is much more useful when the defined latent classes are well separated, which in this simulation study is operationalized as .50 difference in indicator probabilities. Requirement on sample sizes and number of indicators differs depending on whether the purpose is to distinguish among models of different numbers of latent classes and choose the correct number of latent classes, or, to classify individuals into latent classes.

If the purpose is to distinguish among models of different numbers of latent classes, it is important that analysis should yield satisfactory correct class extraction rates regardless of the true number of latent classes. Therefore, one should be searching for common conditions that show high correct class extraction rates across 2-, 3- and 4-class models. Based on ABIC or BLRT (left panels of Tables 4-3 and 4-4), cells with sample sizes of 1,000 or above and 9 or more indicators fit the purpose. With larger sample sizes (e.g., 5000), fewer indicators may be needed (e.g., 6).

The above requirement may be difficult to meet with smaller scale studies. There are a few possibilities that such requirement may be somewhat relaxed. One such condition would be that the researcher has some ideas about the few alternatives he/she is considering. For example, if he or she, based on prior knowledge, believes that the true model contains either 2 or 3 latent classes, and is interested in testing between the 2-class model and 3-class model, then sample size requirement for ABIC can be drastically reduced from 1,000 down to only 300 with 9 or more indicators (e.g., left panel of Table 4-3). Another example is when researchers anticipate the latent classes to be of relatively equal sizes. When this is the case, sample size requirement reduces to 400 with 9 or more indicators to distinguish among 2-, 3- and 4-class models (e.g., Table 4-4).

If the purpose is to classify individuals rather than distinguish models with different numbers of latent classes, recommendation is made when overall classification accuracy is high ($\geq .80$) and the smallest group classification accuracy is significantly higher than random classification. It should be noted that the requirement is established based on the assumption that the researcher has a clear idea about the true number of latent classes based on prior knowledge. However, when this is not the case, requirement is conditional on the correct identification of number of latent classes first.

To classify individuals into 2 latent classes, sample sizes of 100 and 6 indicators seem sufficient to produce overall classification higher than .80; for all those conditions, group classification accuracy is around or above .70, significantly larger than the random classification rate of .50.

For classification into 3 latent classes, sample sizes of at least 200 and 9 or more indicators are recommended; under the studied conditions, this produces overall classification accuracy above .80, and group classification accuracy above .50, significantly higher than the random classification rate of .33.

For 4-latent class classification, sample sizes of at least 300 and 15 or more indicators are recommended; under the studied conditions, those combinations produce overall classification accuracy above .80, and group classification above .60, significantly higher than the random classification rate of .25.

In order to make the requirement more easily adopted in research practices, an attempt was made to simplify the recommendation. Since number of latent classes appears to have such a large effect, a simplified requirement focuses on sample sizes and number of indicators per latent class and is listed in Table 4-17. A tentative guideline is made at 3 to 4 indicators per latent class and a class size of at least 40 to 100 for class extraction purposes. More indicators and larger class sizes are preferred for models with more latent classes: that is, when anticipated number of latent classes is 3 or fewer, class size of 50 and 3 indicators per latent class seem adequate; when anticipated number of latent classes is 4 or higher, class size of at least 80 and 4 indicators per latent class are needed. For classification purposes, if the researcher has little idea about the number of latent classes, then requirement is conditional on the correct identification of latent classes. It turns out that with 3 indicators per latent class and class size of at least 50 (requirement for acceptable correct class extraction with 2- or 3- latent class models), overall classification accuracy rates are at or above .80 while group classification accuracy rates at or above .70 for 2-class models, and at or above .50 for 3-class models. For 4-class models, class size of at least 80 and 4 indicators per latent class produces overall classification accuracy at or above .80 and group classification accuracy rates at or above .60. In sum, when conditions meet requirement for correct class extraction purposes, those conditions also produce acceptable level of overall and group classification accuracy rates.

In instances when the researcher has a clear idea about how many latent classes to extract, 3 indicators per latent class and class size of 20 or sample size of 200, whichever is larger, are needed for classification into 3 or fewer latent classes; for classification into 4 latent classes, 4 indicators per latent classes and class size of 30 or sample sizes of 300, whichever is larger, are needed.

In sum, there are some general guidelines on recommended model condition in LCM. First, LCM is not recommended for latent classes with small class separation. The current study suggests that a difference of class-specific indicator probabilities of .20 may be too small for LCM, but a difference is .50 seems sufficiently large. Second, larger sample sizes and indicators are required for more latent classes, and the ratio of sample size to number of latent class has to increase with increasing number of latent classes to maintain the desirable performance of LCM. In other words, larger class sizes are required for models with more latent classes. Third, requirement on sample sizes varies depending on the purpose of the research study; usually larger sample sizes are required for the purpose of deciding on the optimal number of latent classes than to classify individuals with a known number of latent classes. Fourth, entropy should be used with extra caution as an indication of classification accuracy: high entropy may not necessarily mean good classification accuracy, especially when sample sizes are small (e.g., <100).

Table 4-17. Recommendations for LCM with dichotomous Indicators: large class separation condition[a]

Research Purpose	Number of latent classes	Number of indicators per latent class	
Deciding on the number of latent classes	4	3	Class size of 100 or larger
		≥ 4	Class size of 80 or larger
	≤ 3	2	Class size of 100 or larger
		≥ 3	Class size of 40 or larger
Classification[b]	2	3	Class size of 10 or sample size of 100, whichever is larger
	3	3	Class size of 20 or sample size of 200, whichever is larger
	4	>3	Class size of 30 or sample size of 300, whichever is larger

Note.

[a] Large class separation is operationally defined as .5 difference in indicator probabilities for dichotomous indicators.

[b] Recommendation for classification purposes may be conditional on the correct identification of number of latent classes if the researcher does not have a clear idea of how many latent classes to classify to.

Continuous indicators

Correct class extraction

Tables 4-18 to 4-22 present correct class extraction rates for AIC, BIC, ABIC, BLRT and entropy under all studied conditions for continuous indicators. Cells with correct class extraction rates at or above .80 were bolded. Type I error rates of MST and MKT are presented in Tables 4-23 and 4-24.

Comparing fit indices

As was introduced in Chapter 2, MKT and MST was calculated by first approximating the distribution of the two statistics using re-sampling techniques, which involved very intensive computation. Running on a computer with a CPU of 2.10 GHz and 2.00 GB memory, it took hours for one replication to yield final estimates. Suppose on average, it takes an hour to complete estimation for one replication, with the study's full set of 100 replications for each of the 540 simulation conditions, it would take 2250 days. As a result, only 10 replications were estimated for each simulation condition. This may lead to somewhat coarse measures of Type I error rates. However, a large number of simulation conditions (i.e., 540) somewhat compensated this flaw, and provided chances for inflated Type I error to present itself if it were present. For example, the overall Type I error rate is .02 across all conditions. Furthermore, Muthén, du Toit, & Spisic (in press) considered a Type I error rate close to or lower than .10 to be acceptable in social and behavioral science research, when .05 was the nominal level for statistical significance.

Based on table values of Tables 4-23 and 4-24, MKT and MST both showed some reasonable control over Type I error in most conditions, with many table values having 0 incorrect rejection rates. In general, MKT appeared to have better control of its Type I error than MST. With the presence of many 0 cells, there did not appear to be much variation in MKT type I error rates, but ANOVA on the logit transformed MKT Type I error rates pointed out class proportion to be the only term with *p* value smaller than .05: Type I error rates appeared to be lower for unequal class proportion than for equal class proportion conditions. However, effect size was relatively small, with eta squared of .05.

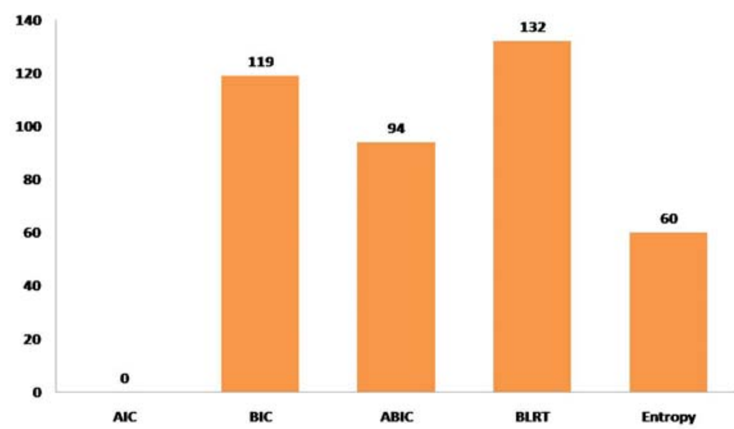

Figure 4-6. Number of cells with acceptable correct class extraction rates ($\geq$ 80%) by fit indices for continuous indicators

Figure 4-6 presents for each fit index, number of cells that show acceptable correct class extraction rates for the remaining fit indices. Among all the remaining fit indices, AIC showed the worst overall performance, with no cell having correct class extraction rates above .80 (see Table 4-18). Entropy, as displayed in Table 4-22, showed better performance with continuous indicators than it was with dichotomous indicators, however, its performance was still not comparable to a few other fit indices discussed below.

Similar to the situation with dichotomous indicators, BLRT outperformed all other fit indices. For large class separation conditions, BLRT showed high correct class extraction rates in most studied conditions with sample sizes of 100 or above, however, its correct class extraction rates for small class separation condition still remained very low (see Table 4-21). BIC from Table 4-19 showed the second best overall performance, and was only inferior to BLRT under a few conditions with sample sizes of 100 or 200 for large class separation conditions. ABIC, as shown in Table 4-20, was a close third, showing high correct class extraction rates at sample sizes of 300 or above.

Table 4-18. Correct class extraction based on AIC for continuous indicators

	Number of Indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	55-60	50-48	45-53	33-49	53-61	22-36	18-37	41-34	0-33	11-46
100	58-58	64-52	41-53	38-47	47-61	29-24	28-39	46-39	42-51	41-50
200	59-61	59-58	51-49	44-50	49-58	9-32	22-37	31-43	38-56	34-59
300	56-62	64-55	48-52	37-38	46-41	23-34	19-41	12-49	39-53	64-51
400	67-54	64-62	50-47	35-39	40-37	34-37	24-51	45-54	44-47	55-58
500	67-54	64-62	49-52	35-38	45-37	29-28	22-56	25-56	63-57	49-55
1000	67-53	58-54	31-38	27-28	31-23	18-44	18-62	53-50	52-47	59-56
2000	63-48	60-38	33-26	21-21	19-24	32-40	43-59	37-60	36-46	39-53
5000	47-45	64-63	26-24	20-19	12-17	49-49	48-52	63-44	50-51	48-50
	3-class									
50	33-50	46-46	41-47	54-81	59-67	14-26	13-6	42-21	96-25	16-31
100	53-55	58-54	57-45	49-78	60-59	13-14	13-57	48-18	94-24	25-12
200	55-60	44-54	61-44	54-65	56-65	13-18	7-5	58-15	4-47	28-42
300	59-61	57-58	59-54	55-71	58-64	11-31	3-24	45-31	1-31	23-40
400	66-63	61-61	54-56	54-70	59-57	10-16	8-24	40-40	3-32	24-39
500	66-61	48-55	59-54	62-68	52-59	14-24	3-26	36-19	94-63	31-58
1000	56-63	57-59	54-41	45-69	46-51	17-51	3-24	28-50	5-51	43-77
2000	60-60	47-60	59-44	45-70	44-52	21-44	16-31	26-7	1-33	39-30
5000	65-64	54-58	53-47	46-67	54-43	39-77	34-54	4-22	5-54	0-31
	4-class									
50	29-41	38-42	43-50	62-47	65-75	22-44	22-22	20-20	32-40	42-10
100	46-57	51-51	52-50	54-59	60-62	34-52	44-34	30-22	23-42	42-31
200	45-60	54-50	56-46	58-52	46-54	15-4	12-15	22-40	13-10	26-41
300	62-62	57-64	59-39	59-57	52-46	22-27	39-19	50-32	30-23	37-42
400	62-62	57-57	54-44	61-59	59-56	28-32	16-32	67-42	32-24	42-38
500	64-60	54-57	58-41	59-50	51-54	5-28	32-16	32-50	27-48	45-23
1000	60-70	64-60	50-51	54-40	52-47	38-25	15-0	20-42	33-32	60-42
2000	60-68	67-46	51-44	56-45	56-43	22-32	22-25	30-37	30-44	45-55
5000	63-69	61-46	45-42	59-38	51-37	36-46	16-32	15-51	60-39	20-30

Note. Values to the left and right sides of “-” represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-19. Correct class extraction based on BIC for continuous indicators

	Number of Indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	**91-100**	**95-99**	**95-100**	**99-100**	**99-100**	0-20	14-24	0-2	0	0
100	**98-100**	**100**	**100-99**	**100**	**100**	19-10	18-17	0-1	0-3	0-1
200	**100-99**	**100**	**100**	**100**	**100**	0-6	9-8	0-1	0-14	0-20
300	**100**	**100**	**100**	**100**	**100**	0-2	0-10	0-9	0-39	0-58
400	100	**100**	**100**	**100**	**100**	0	0-12	0-30	0-59	0-95
500	**100**	**100**	**100**	**100**	**100**	0	16-20	0-48	0-82	0-99
1000	**100**	**100**	**100**	**100**	**100**	11-3	0-64	0-100	0-100	0-100
2000	**100**	**100**	**100**	**100**	**100**	0-34	0-100	0-100	0-100	79-100
5000	**100**	**100**	**100**	**100**	**100**	0-99	39-100	**100**	**100**	**100**
	3-class									
50	30-87	32-97	61-100	57-100	76-100	11-21	2-0	29-0	0	0
100	62-100	72-100	**91-100**	**96-100**	**96-100**	9-0	0-41	35-0	0	0
200	**95-100**	**98-100**	**100**	**100**	**100**	2-0	0	38-0	0	0
300	**100**	**100**	**100**	**100**	**100**	7-25	0	30-0	0	0
400	**100**	**100**	**100**	**100**	**100**	7-11	0-28	28-0	0	0
500	**100**	**100**	**100**	**100**	**100**	1-35	0-14	12-0	0	0
1000	**100**	**100**	**100**	**100**	**100**	8-22	0	2-0	0	0
2000	**100**	**100**	**100**	**100**	**100**	6-69	0	1-0	0-12	0-61
5000	**100**	**100**	**100**	**100**	**100**	17-100	0	2-30	0-100	0-100
	4-class									
50	3-13	2-20	22-77	20-95	35-99	0	13-22	0	0	0
100	4-28	9-52	59-99	73-100	**85-100**	0	34-33	0	0	0
200	35-86	59-92	**99-100**	**100**	**100**	0	17-16	0	0	0
300	70-98	**87-99**	**100**	**100**	**100**	0	12-14	0	0	0
400	**86-100**	**97-100**	**100**	**100**	**100**	0	13-5	0	0	0
500	**94-100**	**100**	**100**	**100**	**100**	0	17-0	0	0	0
1000	**100**	**100**	**100**	**100**	**100**	0	12-0	0	0	0
2000	**100**	**100**	**100**	**100**	**99-100**	0	0	0	0	0
5000	**100**	**100**	**100**	**100**	**99-100**	0	0	0	0	0

Note. Values to the left and right sides of “-” represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-20. Correct class extraction based on ABIC for continuous indicators

	Number of Indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
2-class										
50	1-0	0	0	0	0	0-23	11-23	0	0	0
100	20-14	21-13	2-5	12-2	21-0	35-24	32-21	21-12	16-9	0-10
200	64-68	67-65	61-61	61-63	64-74	24-31	24-43	34-52	44-62	45-69
300	**90-81**	**88-83**	**85-93**	**84-90**	**96-96**	25-30	23-59	34-79	54-80	36-96
400	**98-90**	**90-91**	**94-97**	**96-97**	**99-100**	38-33	34-75	23-95	54-96	39-99
500	**99-94**	**96-98**	**99-99**	**98-100**	**99-100**	0-31	11-82	0-96	22-98	42-100
1000	**97-100**	**100**	**99-100**	**100**	**100**	0-63	19-98	21-100	51-100	**74-100**
2000	**100**	**100**	**100**	**100**	**100**	0-93	41-100	**71-100**	**72-100**	**100**
5000	**100**	**100**	**100**	**100**	**100**	20-98	**90-100**	**100**	**100**	**100**
3-class										
50	1-1	1-0	0	0	0	0-11	0	50-21	0	1-0
100	21-21	12-16	10-8	5-11	43-6	8-16	7-50	52-20	1-31	25-12
200	66-66	51-67	74-60	65-81	74-79	14-17	5-0	57-14	1-32	21-33
300	**85-89**	**81-87**	**89-89**	**86-97**	**93-94**	10-47	2-13	43-26	0-15	6-45
400	**95-96**	**92-90**	**96-93**	**97-97**	**98-99**	15-20	2-27	38-22	1-14	2-32
500	**98-92**	**91-89**	**96-97**	**98-99**	**100-98**	13-39	0-29	26-19	0-33	1-57
1000	**98-100**	**99-99**	**100-99**	**100**	**100**	16-59	.1-0	12-0	0-48	0-100
2000	**100**	**100**	**100**	**100**	**100**	16-90	0-12	2-37	0-100	0-100
5000	**100**	**100**	**100**	**100**	**100**	34-100	0-91	26-23	0-100	1-100
4-class										
50	24-1	12-1	4-0	0	2-0	10-8	10-25	0-10	0-12	10-0
100	18-30	17-18	15-5	10-8	10-4	20-23	32-38	32-34	61-21	32-42
200	51-67	61-57	66-59	70-66	63-79	10-0	12-12	20-32	12-22	11-27
300	75-82	**83-87**	**84-82**	**87-91**	**93-96**	10-12	23-22	33-12	0-12	11-22
400	**95-89**	**90-97**	**91-93**	**97-96**	**100**	20-0	17-0	37-0	0	0
500	**95-94**	**93-96**	**95-96**	**99-100**	**99-99**	0	15-0	15-0	0	0-10
1000	**99-97**	**98-100**	**99-100**	**100**	**100**	0	11-0	0	0	0
2000	**100**	**99-100**	**100**	**100**	**100-99**	0	0	0	0	0
5000	**100**	**100**	**100**	**100**	**99-100**	0	0	0	0-50	0-98

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-21. Correct class extraction based on BLRT for continuous indicators

		Number of Indicators									
		5	6	9	12	15	5	6	9	12	15
		Large class separation					Small class separation				
N	N_c	2-class									
50	5	**91-100**	**89-91**	**94-91**	**93-91**	**98-97**	11-24	14-30	26-24	0-23	0-30
100	10	**95-96**	**98-93**	**90-89**	**92-91**	**90-95**	23-18	29-28	24-27	14-56	11-65
200	20	**96-97**	**94-91**	**95-95**	**94-93**	**99-95**	9-18	32-37	14-64	54-83	12-94
300	30	**94-95**	**96-93**	**93-99**	**88-92**	**96-97**	12-20	0-49	9-77	51-88	43-97
400	40	**99-89**	**93-95**	**94-93**	**93-95**	**96-98**	14-17	14-60	14-96	33-93	43-98
500	50	**97-93**	**92-98**	**92-98**	**88-91**	**94-97**	21-29	23-80	0-97	54-94	61-96
1000	100	**94-93**	**97-97**	**87-88**	**85-93**	**92-95**	0-77	31-99	52-84	**81-87**	**90-99**
2000	200	**97-93**	**94-96**	**91-96**	**90-91**	**90-95**	0-98	69-99	**89-93**	**100-95**	**100-98**
5000	500	**87-98**	**94-96**	**92-84**	**86-91**	**91-88**	72-99	**100-98**	**88-98**	**100-96**	**92-95**
		3-class									
50	5	36-93	44-98	75-87	79-82	**86-97**	2-0	0	38-11	0	0
100	10	**80-93**	**92-96**	**89-93**	**87-91**	**89-99**	0	0-40	50-11	0	0
200	20	**91-95**	**93-95**	**94-95**	**93-95**	**100-99**	0	0	54-0	0-12	1-0
300	30	**93-99**	**95-92**	**95-98**	**90-95**	**94-95**	0	0	59-0	0-12	3-27
400	40	**97-100**	**94-97**	**94-93**	**96-97**	**95-99**	1-0	0-22	57-0	1-30	2-22
500	50	**98-97**	**90-92**	**94-97**	**93-95**	**98-94**	0	0-13	55-25	1-56	4-62
1000	100	**95-98**	**98-94**	**87-85**	**92-98**	**93-95**	0	1-5	39-0	1-10	2-100
2000	200	**98-98**	**94-97**	**100-95**	**93-99**	**98-95**	1-20	0-51	16-22	0-100	25-100
5000	500	**97-99**	**95-93**	**96-92**	**96-98**	**99-94**	3-80	13-100	1-100	4-100	**97-70**
		4-class									
50	5	6-18	5-36	36-89	51-94	62-98	0	15-22	0	0	0
100	10	23-58	40-81	**89-97**	**84-88**	**92-98**	0	32-33	0	0	0
200	20	78-93	**89-93**	**98-96**	**96-95**	**96-100**	0	1-12	0	0	0
300	30	**90-98**	**93-94**	**94-93**	**94-93**	**96-98**	0	10-15	0	0	0
400	40	**97-96**	**96-99**	**90-96**	**93-96**	**96-99**	0	0-16	0	0	0
500	50	**96-97**	**96-97**	**93-94**	**96-97**	**96-96**	0	0-12	0	0-12	0
1000	100	**94-95**	**95-97**	**87-93**	**95-99**	**93-96**	0	10-0	0	0-13	0
2000	200	**95-98**	**95-91**	**97-98**	**99-95**	**96-89**	0	0	0-19	0-47	0-13
5000	500	**94-96**	**95-97**	**92-96**	**98-95**	**96-93**	0	0-10	10-32	0-81	22-92

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively. N_c denotes the smallest class size for unequal class proportion conditions.

Table 4-22. Correct class extraction based on entropy for continuous indicators

	Number of Indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	84-76	90-76	**93-87**	**100-97**	**100**	51-47	22-11	64-38	52-25	42-26
100	**81-78**	**89-84**	**94-88**	**100-99**	**100**	23-45	0-5	55-33	11-25	36-16
200	90-69	91-77	**91-94**	**99-94**	**100-99**	62-43	12-9	68-32	64-13	49-9
300	88-73	89-76	**92-86**	**99-98**	**100-99**	68-32	9-8	17-8	29-10	61-10
400	85-72	88-74	**91-90**	**96-98**	**100**	61-31	35-4	54-13	25-11	45-11
500	93-76	92-75	**93-93**	**98-97**	**100-99**	28-25	0-3	16-5	56-12	65-11
1000	88-61	**90-87**	**95-93**	**99-98**	**100**	26-16	8-4	36-5	51-19	57-18
2000	83-63	87-67	**91-89**	**99-97**	**100**	48-11	5-6	47-19	42-17	51-19
5000	60-60	**81-90**	**92-92**	**99-99**	**100**	35-9	14-13	41-10	51-27	36-17
	3-class									
50	36-19	48-36	65-47	58-57	51-74	4-10	7-20	18-20	1-21	16-13
100	33-45	55-56	77-73	83-73	**84-82**	3-0	2-10	15-20	1-29	30-12
200	38-64	69-70	**82-85**	**87-80**	**92-95**	9-0	3-0	26-20	2-23	24-33
300	34-65	81-74	**90-92**	**92-85**	**99-94**	4-0	1-0	19-30	93-14	23-7
400	39-67	86-67	**88-88**	**96-88**	**98-93**	9-20	2-0	14-60	1-47	32-11
500	38-65	85-72	**89-93**	**97-90**	**98-95**	7-10	2-0	18-20	1-16	29-14
1000	43-71	75-68	**92-82**	**98-92**	**100-94**	5-10	0	20-10	4-1	26-13
2000	44-62	81-72	**94-87**	**98-93**	**99-97**	12-10	0	18-0	6-2	41-5
5000	31-61	74-64	**85-89**	**97-91**	**100-97**	9-0	8-0	27-20	93-0	4-2
	4-class									
50	22-17	27-20	54-25	60-42	67-52	20-32	0	10-19	0-30	20-0
100	37-17	37-26	59-47	77-60	83-61	43-32	0	10-22	38-11	22-39
200	47-27	33-25	70-60	87-72	87-72	52-29	0	45-32	32-46	20-27
300	39-30	30-35	63-72	88-78	**90-81**	0-45	0	34-36	46-43	32-20
400	54-43	24-31	70-78	92-74	**90-91**	25-36	0	52-25	45-45	20-12
500	59-49	29-41	76-81	96-79	**90-91**	75-26	0	30-0	22-37	56-0
1000	69-57	12-45	76-79	**91-81**	**93-90**	47-42	0	22-12	22-11	20-39
2000	80-52	4-47	78-72	92-78	**88-85**	47-33	0	22-12	31-32	17-22
5000	69-60	97-41	67-72	**90-81**	**94-86**	12-12	0-10	30-0	30-0	10-0

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-23. Proportion of replications rejected incorrectly at the 5% level by MST for continuous indicators

	Number of Indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
N	2-class									
50	0-.10	0-.10	0-0	.10-0	0-.10	0-.10	0-.10	0-0	0-0	0-.10
100	0-.10	0-.10	0-0	0-.30	0-0	0-0	0-0	0-0	0-.20	0-.10
200	0-.10	0-0	0-0	0-0	0-0	0-0	0-.10	0-.10	0-0	0-.10
300	0-.10	0-.10	0-.20	0-0	0-0	0-.10	0-.20	0-0	0-0	0-.10
400	0-0	0-.10	0-.10	0-0	0-0	0-0	0-0	0-.10	0-0	0-0
500	0-.10	0-.10	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-0
1000	0-0	0-0	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-0
2000	0-.10	0-.10	0-.10	0-.10	0-0	0-0	0-0	0-.10	0-0	0-0
5000	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0
	3-class									
50	0-.10	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-0	0-0
100	0-.10	0-.10	0-0	0-0	0-0	0-0	0-0	0-.20	0-0	0-.10
200	0-0	0-0	0-.10	0-0	0-0	0-.10	0-0	0-.20	0-.10	0-.10
300	0-.10	0-0	0-.30	0-0	0-0	0-0	0-.10	0-0	0-0	0-0
400	0-.10	0-.20	0-0	0-0	0-.10	0-0	0-.10	0-.10	0-0	0-0
500	0-.10	0-0	0-0	0-0	0-.10	0-0	0-0	0-.20	0-.10	0-0
1000	0-0	0-0	0-0	0-0	0-0	0-.10	0-0	0-.30	0-0	0-.10
2000	0-0	0-0	0-0	0-.10	0-.10	0-0	0-0	0-0	0-0	0-.10
5000	0-.10	0-0	0-0	0-0	0-0	0-.10	0-0	0-.10	0-0	0-0
	4-class									
50	0-.10	0-.10	.10-0	0-0	0-0	0-.10	.10-.20	.10-.10	0-0	0-0
100	0-.10	0-0	.10-0	.20-0	0-.10	.10-.20	.10-.10	.20-0	.20-.10	.10-0
200	0-.10	0-.10	0-0	0-.10	0-0	.0-.10	.30-.20	.40-.20	.20-.50	.40-.30
300	.10-0	.20-0	0-0	0-.10	0-0	.20-0	.20-.10	.10-.30	.40-.10	.20-.20
400	0-0	.10-.10	.10-0	0-0	0-.10	.10-.20	.20-.30	.20-.10	.10-.20	.10-.10
500	0-.10	.20-.10	0-0	0-.10	0-0	0-0	.10-.10	0-.20	.10-.20	0-.20
1000	0-0	.20-0	0-0	0-0	0-0	0-.30	.10-.40	.10-.30	0-.20	.40-0
2000	0-0	0-.10	0-0	.10-.20	0-.10	.20-.20	.40-.20	0-.10	.30-.10	0-0
5000	.10-.10	0-.10	0-.20	0-0	0-0	0-0	.10-.20	0-0	.10-.20	0-0

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively.

Table 4-24. Proportion of replications rejected at the 5% level by MKT for continuous indicators

	Number of Indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
	2-class									
50	0-0	0-0	0-.10	.10-0	0-.10	0-0	0-0	0-0	0-0	0-.10
100	0-0	0-.10	0-.10	.10-0	0-0	0-0	0-.10	0-0	0-0	0-.10
200	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-.10	0-0	0-0
300	0-0	0-.10	0-.10	0-0	0-0	0-0	0-.10	0-.10	0-0	0-0
400	0-0	0-.10	.10-0	0-0	0-0	0-0	0-.10	0-0	0-0	0-0
500	0-0	0-0	0-.20	0-0	.10-0	0-.10	0-.20	0-0	0-0	0-0
1000	0-.10	0-0	.10-0	.10-0	.10-0	0-0	0-0	0-.10	0-0	0-.10
2000	0-.10	0-.10	.10-.10	0-0	0-0	0-.10	0-0	0-0	0-.10	0-.10
5000	0-0	0-.10	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-0
	3-class									
50	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-.10	0-.10	0-0
100	0-.10	0-.10	0-.10	0-0	0-0	0-0	0-0	0-0	0-0	0-0
200	0-.10	0-.10	0-.10	0-0	0-0	0-0	0-0	0-0	0-0	0-0
300	0-.10	0-0	0-0	0-0	0-0	0-0	0-0	0-.20	0-0	0-.10
400	0-0	0-0	0-.10	0-0	0-0	0-0	0-.10	0-0	0-0	0-0
500	0-0	0-.20	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-0
1000	0-0	0-0	0-0	0-0	0-0	0-.10	0-.10	0-0	0-0	0-.20
2000	0-0	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-.10	0-0
5000	0-0	0-0	0-0	0-0	0-.10	0-0	0-0	0-.10	0-0	0-0
	4-class									
50	0-.10	0-0	0-0	0-.10	0-0	0-0	0-.10	0-.10	0-.10	0-0
100	0-0	0-.10	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0
200	0-0	0-0	.10-.10	0-0	0-0	0-0	0-.10	.10-0	0-0	0-0
300	0-0	.10-.10	0-.10	0-0	0-.10	0-0	.10-0	.20-0	.10-0	0-0
400	0-.10	0-0	.10-.10	0-0	0-0	0-0	0-0	0-0	.10-.20	0-0
500	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-0	0-0	.10-.10
1000	0-0	0-0	.10-0	0-0	0-0	0-0	0-.20	0-.10	.10-0	0-0
2000	.10-0	.10-0	.10-0	0-0	0-.10	0-0	0-0	0-0	.10-0	0-0
5000	0-0	0-0	0-0	0-0	0-0	0-.10	0-0	0-0	0-0	0-0

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions respectively.

Effects of manipulated factors on correct class extraction rates

ANOVA was conducted on the logit transformed correct class extraction rates based on BLRT, ABIC, and BIC respectively. Similar to dichotomous indicators, terms of at least moderate effect sizes (with eta squared larger than .06) were further interpreted.

Table 4-25. ANOVA table for logit transformed BLRT-based correct class extraction rates for continuous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	**553.15**	**2**	**138.52**	**.00**	**.08**
SS(sample size)	334.11	8	20.92	.00	.05
ND(#indicators)	78.07	4	9.78	.00	.01
CS(class separation)	**3497.65**	**1**	**1751.83**	**.00**	**.52**
CP(class proportion)	128.89	1	64.55	.00	.02
NC×SS	28.18	16	0.88	.59	.00
NC×ND	33.16	8	2.08	.05	.00
NC×CS	**493.26**	**2**	**123.53**	**.00**	**.07**
NC×CP	10.42	2	2.61	.08	.00
SS×ND	45.23	32	0.71	.86	.01
SS×CS	181.31	8	11.35	.00	.03
SS×CP	18.32	8	1.15	.35	.00
ND×CS	57.32	4	7.18	.00	.01
ND×CP	14.61	4	1.83	.13	.00
CS×CP	34.48	1	17.27	.00	.01
NC×SS×ND	104.86	64	0.82	.78	.02
NC×SS×CS	121.74	16	3.81	.00	.02
NC×SS×CP	44.49	16	1.39	.17	.01
SS×ND×CS	119.52	32	1.87	.02	.02
SS×ND×CP	42.99	32	0.67	.89	.01
ND×CS×CP	15.87	4	1.99	.11	.00
NC×ND×CS	74.83	8	4.68	.00	.01
NC×ND×CP	49.23	8	3.08	.01	.01
NC×CS×CP	31.88	2	7.98	.00	.00
SS×CS×CP	56.88	8	3.56	.00	.01
NC×SS×ND×CS	103.56	64	0.81	.80	.02
NC×SS×ND×CP	136.37	64	1.07	.40	.02
NC×ND×CS×CP	40.23	8	2.52	.02	.01
SS×ND×CS×CP	26.72	32	0.42	.99	.00
NC×SS×CS×CP	68.29	16	2.14	.02	.01
Error	127.78	64			
Correct Total	6673.39				

Table 4-25 reports ANOVA results for the logit transformed BLRT-based correct extraction rates. Three terms displayed at least moderate effect sizes: class separation with eta squared of .52, number of latent classes with eta squared of .08 and the interaction between them with eta squared of .07. The interaction effects between class separation and number of latent classes is presented in Figure 4-7. For large class separation conditions, increase in number of latent classes showed a small negative effect on correct class extraction rates, with average correct class extraction rates of .93, .92, and .88 for 2- to 4-class models; for small class separation, this effect was more drastic, with average correct class extraction rates of .53, .19, and .05 for 2- to 4-class models. The smaller difference for large class separation conditions was possibly due to the ceiling effects of correct class extraction rates when class separation was large for continuous indicators.

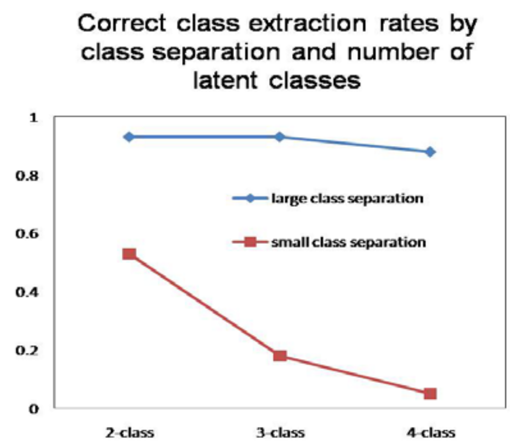

Figure 4-7. Effects of class separation and number of latent classes on correct class extraction rates for continuous indicators for continuous indicators

Tables 4-26 and 4-27 display AVOVA results for the logit transformed correct extraction rates based on ABIC and BIC. For ABIC, class separation, sample size and their interaction showed at least moderate effect sizes, with eta squared of .24, .28 and .15 respectively. For large class separation, correct classification rates increased steadily with increasing sample size, but this pattern was less consistent for small class separation. In terms of BIC, class separation shows a dominating effect, explaining 73% of the total variance.

In sum, across the three fit indices (i.e., BLRT, ABIC, & BIC), class separation showed a consistent effect with larger class separation contributing to higher correct class extraction rates. In addition, BLRT was mostly affected by number of true latent classes, while ABIC was mostly affected by sample sizes.

Table 4-26. ANOVA table for logit transformed ABIC-based correct class extraction rates: continuous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	303.27	2	130.31	.00	.04
SS(sample size)	**2208.74**	**8**	**237.26**	**.00**	**.28**
ND(#indicators)	30.52	4	6.56	.00	.00
CS(class separation)	**1876.24**	**1**	**1612.38**	**.00**	**.24**
CP(class proportion)	102.16	1	87.79	.00	.01
NC×SS	301.44	16	16.19	.00	.04
NC×ND	53.29	8	5.72	.00	.01
NC×CS	321.53	2	138.16	.00	.04
NC×CP	54.27	2	23.32	.00	.01
SS×ND	98.58	32	2.65	.00	.01
SS×CS	**1157.24**	**8**	**124.31**	**.00**	**.15**
SS×CP	63.29	8	6.80	.00	.01
ND×CS	10.01	4	2.15	.08	.00
ND×CP	31.16	4	6.69	.00	.00
CS×CP	93.63	1	80.47	.00	.01
NC×SS×ND	95.46	64	1.28	.16	.01
NC×SS×CS	203.70	16	10.94	.00	.03
NC×SS×CP	62.17	16	3.34	.00	.01
SS×ND×CS	60.00	32	1.61	.05	.01
SS×ND×CP	31.69	32	0.85	.69	.00
ND×CS×CP	22.37	4	4.81	.00	.00
NC×ND×CS	52.91	8	5.68	.00	.01
NC×ND×CP	37.82	8	4.06	.00	.00
NC×CS×CP	46.11	2	19.81	.00	.01
SS×CS×CP	55.99	8	6.01	.00	.01
NC×SS×ND×CS	99.06	64	1.33	.13	.01
NC×SS×ND×CP	77.29	64	1.04	.44	.01
NC×ND×CS×CP	32.73	8	3.52	.00	.00
SS×ND×CS×CP	31.68	32	0.85	.69	.00
NC×SS×CS×CP	67.93	16	3.65	.00	.01
Error	74.47	64			
Correct Total	7756.75	539			

Table 4-27. ANOVA table for logit transformed BIC-based correct class extraction rates: continuous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	424.93	2	139.99	.00	.04
SS(sample size)	427.99	8	35.25	.00	.04
ND(#indicators)	18.52	4	3.05	.02	.00
CS(class separation)	**8764.88**	**1**	**5775.10**	**.00**	**.73**
CP(class proportion)	150.71	1	99.30	.00	.01
NC×SS	113.71	16	4.68	.00	.01
NC×ND	98.97	8	8.15	.00	.01
NC×CS	68.28	2	22.49	.00	.01
NC×CP	57.76	2	19.03	.00	.00
SS×ND	32.59	32	0.67	.89	.00
SS×CS	182.86	8	15.06	.00	.02
SS×CP	20.83	8	1.72	.11	.00
ND×CS	41.42	4	6.82	.00	.00
ND×CP	13.96	4	2.30	.07	.00
CS×CP	12.80	1	8.43	.01	.00
NC×SS×ND	113.71	64	1.17	.27	.01
NC×SS×CS	355.79	16	14.65	.00	.03
NC×SS×CP	98.45	16	4.05	.00	.01
SS×ND×CS	152.57	32	3.14	.00	.01
SS×ND×CP	23.79	32	0.49	.99	.00
ND×CS×CP	18.17	4	2.99	.03	.00
NC×ND×CS	151.48	8	12.48	.00	.01
NC×ND×CP	31.48	8	2.59	.02	.00
NC×CS×CP	152.32	2	50.18	.00	.01
SS×CS×CP	136.41	8	11.23	.00	.01
NC×SS×ND×CS	72.84	64	0.75	.87	.01
NC×SS×ND×CP	110.26	64	1.14	.31	.01
NC×ND×CS×CP	32.19	8	2.65	.01	.00
SS×ND×CS×CP	33.14	32	0.68	.88	.00
NC×SS×CS×CP	55.48	16	2.28	.01	.00
Error	97.13	64			
Correct Total	12065.41	539			

Classification accuracy

Overall classification accuracy

Table 4-28 reports the overall classification rates under each of the studied conditions for continuous indicators. Results from ANOVA on overall classification accuracy are displayed in Table 4-29. Class separation was the most influential factor on overall classification accuracy, and it accounted for 75% of the total variance in overall classification accuracy. As displayed in the left panel of Figure 4-8, average overall classification accuracy was .97 for large class separation conditions, in contrast to .58 for small class separation conditions. In terms of cells with acceptable overall classification accuracy (i.e., $\geq$ .80), 134 out of the total of 135 cells showed high overall classification accuracy, compared to only 7 for small class separation conditions.

In addition, true number of latent classes explained an additional 9% of the total variance: as true number of latent classes increased from 2 to 4, overall classification accuracy decreased from .84 to .79, and then down to .69, as shown in the right panel of Figure 4-8.

In sum, overall classification accuracy was higher for large class separation and fewer latent classes. Effects of other main effects remained small, with eta squared of .01, .05, and .01 for sample size, number of indicators and class proportions.

Table 4-28. Overall classification accuracy rates by class separation, number of latent classes, sample sizes, number of indicators and class proportions: continuous indicators

		Number of Indicators									
		5	6	9	12	15	5	6	9	12	15
		Large class separation					Small class separation				
N	N_c	2-class									
50	5	**.98-.99**	**.99-.99**	**1**	**1**	**1**	.22-.52	.49-.61	.49-.63	.20-.68	.86-.69
100	10	**.99-.99**	**.99-.99**	**1**	**1**	**1**	.59-.55	.82-.62	.83-.65	.58-.69	.87-.75
200	20	**.99-.99**	**1-.99**	**1**	**1**	**1**	.15-.55	.85-.64	.22-.67	.87-.75	.61-.79
300	30	**.99-.99**	**.99-.99**	**1**	**1**	**1**	.87-.57	.87-.65	.84-.71	.17-.76	.17-.81
400	40	**.99-.99**	**1-.99**	**1**	**1**	**1**	.66-.58	.13-.67	.76-.72	.78-.76	**.81-.91**
500	50	**.99-.99**	**.99-.99**	**1**	**1**	**1**	.38-.60	.57-.68	.90-.74	.92-.79	.17-.82
1000	100	**.99-.99**	**1-.99**	**1**	**1**	**1**	.36-.64	.88-.70	.91-.76	**.91-.80**	**.92-.83**
2000	200	**.99-.99**	**.99-.99**	**1**	**1**	**1**	.15-.67	.90-.71	.90-.76	**.93-.80**	**.93-.83**
5000	500	**.99-.99**	**.99-.99**	**1**	**1**	**1**	.90-.70	.90-.72	.91-.77	**.92-.80**	**.93-.83**
		3-class									
50	5	**.91-.91**	**.95-.94**	**.98-.98**	**.99-.99**	**.99-1**	.59-.37	.63-.37	.75-.36	.74-.41	.74-.54
100	10	**.94-.92**	**.97-.95**	**.99-.98**	**1-.99**	**1**	.64-.45	.66-.45	.79-.37	.61-.45	.75-.52
200	20	**.95-.93**	**.97-.95**	**.99-.98**	**.99-.99**	**1**	.63-.47	.69-.43	.81-.36	.78-.50	.76-.43
300	30	**.95-.93**	**.97-.96**	**.99-.99**	**.99-.99**	**1**	.68-.41	.70-.48	.81-.30	.79-.53	.76-.68
400	40	**.96-.93**	**.97-.96**	**.99-.99**	**.99-.99**	**1**	.68-.41	.72-.44	.80-.35	.80-.47	.77-.61
500	50	**.96-.93**	**.97-.96**	**.99-.99**	**.99**	**1**	.69-.40	.71-.47	.81-.33	.78-.52	.76-.58
1000	100	**.96-.93**	**.98-.96**	**.99-.99**	**.99-1**	**1**	.66-.42	.71-.44	.81-.44	.35-.57	.78-.62
2000	200	**.96-.94**	**.98-.96**	**.99-.99**	**.99-1**	**1**	.69-.46	.70-.45	.81-.54	.79-.62	.80-.64
5000	500	**.96-.94**	**.98-.96**	**.99-.99**	**.99-1**	**1**	.71-.51	.74-.47	.74-.57	.82-.62	.77-.67
		4-class									
50	5	.79-.76	**.83-.82**	**.93-.93**	**.96-.97**	**.98-.99**	.42-.38	.49-.34	.52-.35	.55-.43	.49-.39
100	10	**.83-.80**	**.87-.85**	**.95-.94**	**.98-.97**	**.99-.99**	.45-.33	.45-.35	.46-.36	.44-.37	.56-.39
200	20	**.86-.84**	**.90-.88**	**.96-.95**	**.98-.98**	**.99-.99**	.47-.31	.57-.33	.51-.32	.61-.38	.59-.37
300	30	**.88-.84**	**.90-.88**	**.96-.95**	**.98-.98**	**.99-.99**	.49-.30	.55-.32	.53-.32	.55-.36	.58-.36
400	40	**.88-.85**	**.91-.89**	**.96-.95**	**.98-.98**	**.99-.99**	.50-.32	.46-.31	.54-.35	.59-.33	.58-.37
500	50	**.89-.85**	**.91-.89**	**.96-.95**	**.98-.98**	**.99-.99**	.57-.30	.51-.32	.61-.34	.56-.35	.63-.39
1000	100	**.89-.86**	**.91-.89**	**.96-.95**	**.99-.98**	**.99-.99**	.47-.31	.54-.30	.66-.36	.57-.37	.63-.43
2000	200	**.89-.86**	**.91-.89**	**.96-.95**	**.99-.98**	**.99-.93**	.50-.32	.51-.33	.55-.38	.62-.41	.63-.48
5000	500	**.89-.86**	**.92-.89**	**.96-.95**	**.99-.98**	**.97-.99**	.45-.33	.59-.30	.52-.41	.65-.47	.69-.51

Note. Values to the left and right sides of “-” represent correct class extraction rates for unequal and equal class proportions respectively. N_c denotes the smallest class size for unequal class proportion conditions.

Table 4-29. ANOVA table for logit transformed overall classification accuracy: continuous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	**219.91**	**2**	**578.45**	**.00**	**0.09**
SS(sample size)	13.53	8	8.90	.00	0.01
ND(#indicators)	126.55	4	166.44	.00	0.05
CS(class separation)	**1814.52**	**1**	**9545.79**	**.00**	**0.75**
CP(class proportion)	18.62	1	97.95	.00	0.01
NC×SS	3.37	16	1.11	.37	0.00
NC×ND	12.07	8	7.93	.00	0.00
NC×CS	17.76	2	46.73	.00	0.01
NC×CP	9.24	2	24.31	.00	0.00
SS×ND	6.25	32	1.03	.45	0.00
SS×CS	3.67	8	2.41	.02	0.00
SS×CP	1.80	8	1.19	.32	0.00
ND×CS	37.07	4	48.75	.00	0.02
ND×CP	2.88	4	3.79	.01	0.00
CS×CP	8.09	1	42.58	.00	0.00
NC×SS×ND	11.94	64	0.98	.63	0.00
NC×SS×CS	7.95	16	2.61	.00	0.00
NC×SS×CP	4.63	16	1.52	.12	0.00
SS×ND×CS	8.38	32	1.38	.14	0.00
SS×ND×CP	7.09	32	1.17	.30	0.00
ND×CS×CP	1.69	4	2.22	.08	0.00
NC×ND×CS	21.76	8	14.31	.00	0.01
NC×ND×CP	2.00	8	1.32	.25	0.00
NC×CS×CP	5.43	2	14.29	.00	0.00
SS×CS×CP	0.67	8	0.44	.89	0.00
NC×SS×ND×CS	12.97	64	1.07	.40	0.01
NC×SS×ND×CP	13.38	64	1.10	.35	0.01
NC×ND×CS×CP	1.42	8	0.93	.50	0.00
SS×ND×CS×CP	7.85	32	1.29	.19	0.00
NC×SS×CS×CP	5.07	16	1.67	.08	0.00
Error	12.17	64			
Correct Total	2419.76				

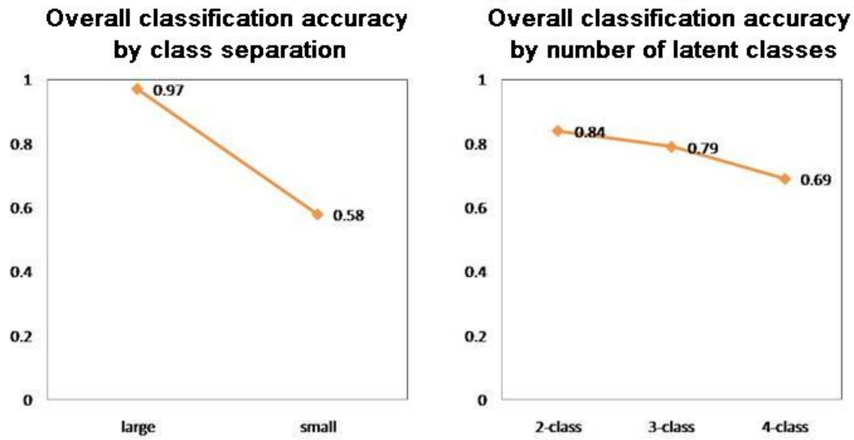

Figure 4-8. Main effects of class separation and number of latent classes on the overall classification accuracy for continuous indicators

Group classification accuracy

Group classification accuracy rates are shown in Table 4-30, while the ANOVA results are reported in Table 4-31. Cells with group classification rates statistically significantly higher than random classification rates were bolded in Table 4-30 (i.e., >.64 for 2-class, >.46 for 3-class, and >.37 for 4-class models). Note that all conditions with large class separation achieved that level of significance. In contrast, conditions with small class separation rarely achieved acceptable group classification accuracy.

Based on ANOVA results form Table 4-31, class separation again showed the largest effect on group classification accuracy, accounting for 75% of the total variance. This effect is visually presented in the left panel of Figure 4-9: for large class separation, average group classification accuracy was .94; in contrast, average group classification accuracy was only .29 for small class separation conditions. True number of latent classes was the second largest factor, explaining 8% of total variance: average group classification accuracy decreased from .74 for 2-class models to .57 for 3-class models, and then further down to .54 for 4-class models. The effect of number of latent classes is shown in the right panel of Figure 4-9.

Table 4-30. Lowest group classification accuracy for continuous indicators

		Number of Indicators									
		5	6	9	12	15	5	6	9	12	15
		Large class separation					Small class separation				
N	N_c	2-class									
50	5	**.95-.98**	**.98-.99**	**1**	**1**	**1**	.10-.52	.20-.53	.44-.59	.07-.66	.70-.65
100	10	**.96-.98**	**.98-.99**	**1**	**1**	**1**	.58-.51	.07-.54	.80-.65	.57-.68	.86-.74
200	20	**.96-.99**	**.98-.99**	**1**	**1**	**1**	.10-.47	.16-.60	.13-.64	.37-.73	.61-.79
300	30	**.96-.99**	**.98-.99**	**1**	**1**	**1**	.16-.54	.14-.62	.42-.71	.10-.75	.12-.80
400	40	**.96-.99**	**.98-.99**	**1**	**1**	**1**	.65-.53	.06-.62	.39-.71	.17-.77	.36-.81
500	50	**.96-.99**	**.98-.99**	**1**	**1**	**1**	.33-.56	.56-.65	.26-.72	.34-.78	.09-.82
1000	100	**.96-.99**	**.98-.99**	**1**	**1**	**1**	.35-.61	.20-.70	.24-.73	.30-.79	.50-.83
2000	200	**.96-.99**	**.97-.99**	**1**	**1**	**1**	.07-.61	.13-.71	.20-.75	.34-.80	.39-.83
5000	500	**.96-.99**	**.98-.99**	**1**	**1**	**1**	.10-.69	.13-.71	.30-.76	.41-.80	.39-.83
		3-class									
50	5	**.75-.88**	**.78-.94**	**.94-.97**	**.97-.99**	**.98-1**	.30-.04	.25-.09	.12-.03	.02-.05	.22-.24
100	10	**.81-.91**	**.89-.95**	**.96-.98**	**.98-.99**	**1**	.21-.04	.22-.05	.05-.02	.02-.36	.19-.08
200	20	**.84-.92**	**.90-.95**	**.97-.98**	**.99-.99**	**1**	.22-.32	.18-.21	.01-.08	.01-.23	.19-.04
300	30	**.85-.92**	**.91-.96**	**.97-.98**	**.99-.99**	**1**	.15-.07	.17-.18	.01-.04	.02-.41	.21-.64
400	40	**.85-.92**	**.91-.96**	**.97-.98**	**.99-.99**	**1**	.15-.05	.15-.29	.01-.04	.02-.05	.20-.58
500	50	**.86-.92**	**.91-.96**	**.97-.98**	**.99-.99**	**1**	.14-.07	.16-.40	.02-.02	.02-.35	.21-.36
1000	100	**.86-.92**	**.91-.96**	**.97-.99**	**.99-1**	**1**	.16-.16	.14-.12	.01-.14	.30-.45	.25-.49
2000	200	**.85-.92**	**.91-.96**	**.97-.99**	**.99-1**	**1**	.16-.18	.15-.12	.04-.45	.19-.53	.28-.58
5000	500	**.85-.92**	**.91-.96**	**.97-.99**	**.99-1**	**1**	.15-.40	.16-.29	.06-.38	.16-.54	.07-.66
		4-class									
50	5	**.57-.70**	**.57-.74**	**.79-.90**	**.91-.96**	**.94-.98**	.13-.26	.15-.21	.21-.28	.22-.32	.18-.26
100	10	**.60-.75**	**.72-.79**	**.90-.93**	**.96-.97**	**.96-.98**	.11-.21	.13-.15	.19-.23	.22-.29	.18-.33
200	20	**.69-.81**	**.75-.83**	**.90-.94**	**.95-.98**	**.98-.99**	.12-.13	.12-.09	.08-.17	.12-.18	.11-.24
300	30	**.72-.81**	**.75-.85**	**.91-.94**	**.95-.98**	**.98-.99**	.19-.26	.13-.15	.18-.19	.18-.23	.18-.25
400	40	**.73-.82**	**.76-.85**	**.91-.94**	**.96-.98**	**.98-.99**	.17-.17	.22-.24	.19-.22	.08-.13	.10-.33
500	50	**.73-.83**	**.75-.85**	**.91-.95**	**.96-.98**	**.98-.99**	.10-.21	.19-.26	.06-.25	.15-.23	.06-.26
1000	100	**.74-.83**	**.75-.86**	**.92-.95**	**.96-.98**	**.98-.99**	.07-.22	.07-.26	.06-.31	.13-.21	.09-.35
2000	200	**.74-.84**	**.76-.86**	**.92-.95**	**.96-.98**	**.98-.93**	.13-.21	.13-.19	.07-.31	.12-.39	.20-.47
5000	500	**.74-.84**	**.76-.86**	**.92-.95**	**.96-.98**	**.95-.99**	.10-.20	.07-.27	.20-.27	.09-.41	.12-.49

Note. Values to the left and right sides of "-" represent correct class extraction rates for unequal and equal class proportions.

Table 4-31. ANOVA table for logit transformed group classification accuracy: continuous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	**352.46**	**2**	**895.96**	**.00**	**.08**
SS(sample size)	11.16	8	7.09	.00	.00
ND(#indicators)	190.29	4	241.85	.00	.05
CS(class separation)	**3107.69**	**1**	**15799.7**	**.00**	**.75**
CP(class proportion)	92.49	1	470.23	.00	.02
NC×SS	10.60	16	3.37	.00	.00
NC×ND	25.22	8	16.02	.00	.01
NC×CS	55.71	2	141.63	.00	.01
NC×CP	5.92	2	15.04	.00	.00
SS×ND	10.82	32	1.72	.03	.00
SS×CS	6.44	8	4.09	.00	.00
SS×CP	9.15	8	5.82	.00	.00
ND×CS	86.30	4	109.69	.00	.02
ND×CP	1.66	4	2.11	.09	.00
CS×CP	11.12	1	56.55	.00	.00
NC×SS×ND	17.24	64	1.37	.11	.00
NC×SS×CS	9.28	16	2.95	.00	.00
NC×SS×CP	3.92	16	1.24	.26	.00
SS×ND×CS	9.98	32	1.59	.06	.00
SS×ND×CP	9.46	32	1.50	.08	.00
ND×CS×CP	11.03	4	14.03	.00	.00
NC×ND×CS	32.35	8	20.56	.00	.01
NC×ND×CP	11.61	8	7.38	.00	.00
NC×CS×CP	12.36	2	31.43	.00	.00
SS×CS×CP	9.85	8	6.26	.00	.00
NC×SS×ND×CS	16.97	64	1.35	.12	.00
NC×SS×ND×CP	14.73	64	1.17	.27	.00
NC×ND×CS×CP	7.13	8	4.53	.00	.00
SS×ND×CS×CP	9.33	32	1.48	.09	.00
NC×SS×CS×CP	6.25	16	1.98	.03	.00
Error	12.59	64			
Correct Total	4171.10				

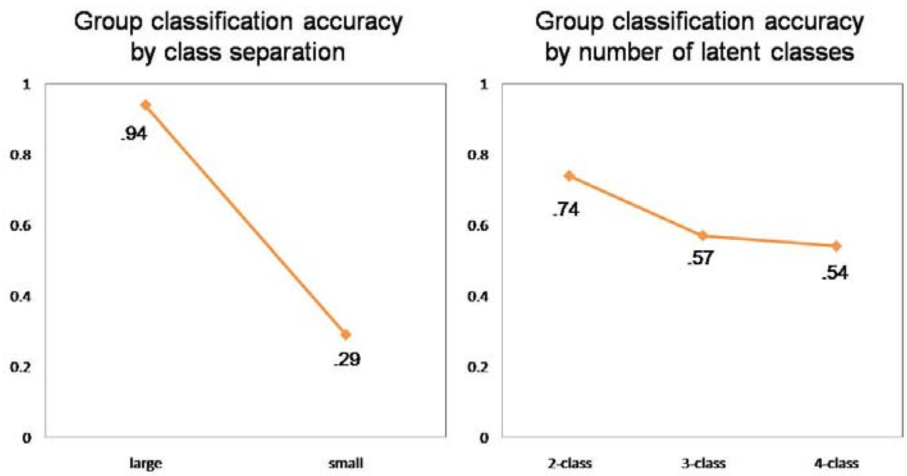

Figure 4-9. Main effects of class separation and number of latent classes on group classification accuracy for continuous indicators

Entropy and classification accuracy

Table 4-32 reports the average entropy values over the 100 replications for each studied condition. Cells with values above .80 were bolded, and cells with values below .60 were italicized. Figure 4-10 presents a visual display of the lack of consistency between entropy and the actual overall and group classification accuracy. Entropy is categorized into the “large”, “medium” and “low” groups (indicated by the blue, red and black colors) using cutoffs of .80 and .60. At each level of the entropy level, there is a wide spread of the classification accuracy: for large entropy values, overall classification accuracy ranges from around .20 to close to 1 and group classification accuracy ranges from close to 0 to 1; for small entropy values, overall classification accuracy ranges from around .30 to around .90 and group classification accuracy ranges from 0 to .80.

Table 4-32. Entropy for continuous indicators

	Number of Indicators									
	5	6	9	12	15	5	6	9	12	15
	Large class separation					Small class separation				
2-class										
50	**.98-.96**	**.99-.98**	**1**	**1**	**1**	**.96-.92**	.79-.79	**.85-.82**	**.86-.82**	**.83-.84**
100	**.97-.96**	**.99-.98**	**1**	**1**	**1**	**.84-.86**	.70-.69	.78-.68	.70-.73	.77-.69
200	**.97-.95**	**.98-.97**	**1-.99**	**1**	**1**	**.86-.84**	.86-*.57*	.74-*.57*	.76-*.55*	.73-*.57*
300	**.97-.95**	**.98-.97**	**1-.99**	**1**	**1**	.96-.76	.72-*.47*	*.59-.45*	.69-*.49*	.74-*.53*
400	**.97-.95**	**.98-.97**	**1-.99**	**1**	**1**	.86-.69	.68-*.41*	.67-*.43*	.71-*.46*	.68-*.51*
500	**.97-.95**	**.98-.97**	**1-.99**	**1**	**1**	.66-*.58*	.68-*.38*	.63-*.38*	.70-*.44*	.78-*.50*
1000	**.97-.95**	**.98-.97**	**1-.99**	**1**	**1**	.81-*.40*	.69-*.29*	.67-*.35*	.69-*.42*	.70-*.48*
2000	**.97-.95**	**.98-.97**	**1-.99**	**1**	**1**	.75-*.29*	.66-*.26*	.62-*.33*	.65-*.41*	.75-*.48*
5000	**.97-.95**	**.98-.97**	**1-.99**	**1**	**1**	*.58-.22*	*.57-.24*	.66-*.32*	.68-*.40*	.75-*.47*
3-class										
50	**.92-.90**	**.95-.93**	**.98-.98**	**1-.99**	**1**	.81-.76	**.83-.84**	**.95-.98**	**.83-.88**	**.90-.89**
100	**.90-.87**	**.95-.91**	**.98-.97**	**.99-.99**	**1**	.73-.71	.74-.72	**.96-.94**	.72-.77	.81-.77
200	**.90-.86**	**.94-.90**	**.98-.97**	**.99-.99**	**1**	.68-.61	.68-.63	**.98-.96**	.81-.68	.74-.63
300	**.90-.85**	**.94-.90**	**.98-.97**	**.99-.99**	**1**	.67-.67	.65-.62	**.97-.95**	.72-*.55*	.69-*.56*
400	**.90-.85**	**.94-.90**	**.98-.97**	**.99-.99**	**1**	.63-.63	.64-*.55*	**.98-.95**	.70-*.55*	.68-*.52*
500	**.90-.85**	**.94-.90**	**.98-.97**	**.99-.99**	**1**	*.59-.61*	.63-*.54*	**.98-.94**	*.52-.47*	.65-*.49*
1000	**.90-.85**	**.94-.90**	**.98-.97**	**.99-.99**	**1**	*.56-.50*	*.59-.50*	.98-.61	*.31-.35*	.61-*.39*
2000	**.90-.85**	**.94-.90**	**.98-.97**	**.99-.99**	**1**	*.55-.33*	*.56-.31*	.88-*.42*	*.44-.29*	.60-*.34*
5000	**.90-.85**	**.94-.90**	**.98-.97**	**.99-.99**	**1**	*.52-.23*	*.52-.22*	*.42-.24*	*.55-.27*	.62-*.32*
4-class										
50	**.88-.87**	**.90-.90**	**.96-.96**	**.99-.98**	**.99-.99**	**.85-.86**	**.82-.84**	**.87-.86**	**.89-.89**	**.92-.90**
100	**.83-.80**	**.86-.84**	**.95-.93**	**.98-.97**	**.99-.99**	.78-.74	.75-.75	.77-.75	.80-.79	**.83-.81**
200	.81-.76	**.85-.81**	**.94-.92**	**.98-.97**	**.99-.98**	.77-.65	.70-.69	.74-.67	.72-.69	.72-.71
300	.81-.74	**.84-.81**	**.94-.91**	**.97-.96**	**.99-.98**	.67-.64	.67-.63	.68-.65	.67-.68	.70-.66
400	.80-.74	**.84-.80**	**.93-.91**	**.97-.96**	**.99-.98**	*.59-.61*	.60-*.58*	.64-.61	.70-.70	.67-*.59*
500	.80-.74	**.84-.80**	**.93-.91**	**.97-.96**	**.99-.98**	.67-*.56*	*.58-.58*	.67-*.56*	.63-.60	.62-*.54*
1000	.79-.73	.84-.79	**.93-.91**	**.97-.96**	**.99-.98**	*.56-.55*	*.58-.59*	.65-*.46*	*.51-.43*	*.52-.42*
2000	.79-.73	.84-.79	**.93-.91**	**.97-.96**	**.99-.98**	*.50-.43*	*.43-.46*	*.48-.37*	*.48-.33*	.78-*.31*
5000	.79-.73	.84-.79	**.93-.91**	**.97-.96**	**.99-.98**	*.38-.33*	*.53-.37*	*.40-.24*	*.43-.24*	*.48-.26*

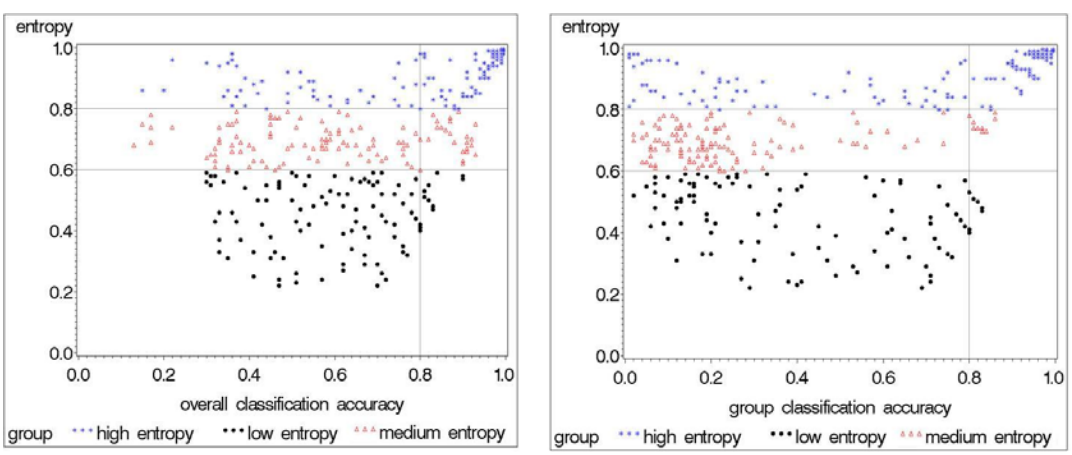

Figure 4-10. Scatterplot between entropy and overall/group classification accuracy for continuous indicators

ANOVA on entropy (reported in Table 4-33) revealed 3 terms to have at least moderate effect sizes: class separation explaining 60% of the total variance, number of indicators explaining 8% of the total variance, and the interaction between the two explaining another 8% of the total variance. The adverse effect of large sample sizes on entropy was stronger for small class separation conditions than for large class separation conditions. Take the smallest and largest sample sizes for example, for large class separation conditions, entropy for sample sizes of 50 and 5,000 was .96 and .94 respectively; however, for small class separation conditions, entropy for those two sample sizes were .86 and .42 respectively.

Table 4-33. ANOVA table for logit transformed entropy: continuous indicators

Source	SS	df	F	*P* value	eta squared
NC (#classes)	86.04	2	759.54	.00	.04
SS(sample size)	80.17	8	176.94	.00	.04
ND(#indicators)	**160.55**	**4**	**708.67**	**.00**	**.08**
CS(class separation)	**1161.17**	**1**	**20501.40**	**.00**	**.60**
CP(class proportion)	27.78	1	490.49	.00	.01
NC×SS	3.63	16	4.01	.00	.00
NC×ND	29.66	8	65.47	.00	.02
NC×CS	74.22	2	655.18	.00	.04
NC×CP	2.27	2	20.08	.00	.00
SS×ND	4.29	32	2.37	.00	.00
SS×CS	42.55	8	93.92	.00	.02
SS×CP	5.02	8	11.07	.00	.00
ND×CS	**154.17**	**4**	**680.49**	**.00**	**.08**
ND×CP	.99	4	4.35	.00	.00
CS×CP	2.71	1	47.90	.00	.00
NC×SS×ND	10.06	64	2.78	.00	.01
NC×SS×CS	2.63	16	2.90	.00	.00
NC×SS×CP	.60	16	.66	.82	.00
SS×ND×CS	4.93	32	2.72	.00	.00
SS×ND×CP	1.87	32	1.03	.45	.00
ND×CS×CP	.71	4	3.13	.02	.00
NC×ND×CS	36.76	8	81.14	.00	.02
NC×ND×CP	2.31	8	5.09	.00	.00
NC×CS×CP	2.76	2	24.36	.00	.00
SS×CS×CP	3.45	8	7.61	.00	.00
NC×SS×ND×CS	9.73	64	2.68	.00	.01
NC×SS×ND×CP	3.21	64	0.88	.69	.00
NC×ND×CS×CP	1.06	8	2.34	.03	.00
SS×ND×CS×CP	1.42	32	0.78	.77	.00
NC×SS×CS×CP	.96	16	1.06	.41	.00
Error	3.62	64			
Correct Total	1921.31				

In sum, with continuous indicators, high entropy values did not necessarily mean high overall or group classification accuracy. Entropy tended to be large with small sample sizes despite small class separation.

Additional analysis with moderate class separation conditions

Findings from the above analyses pointed out class separation as the most significant factor in the two criterion measures of model performance. The large difference in both correct class extraction rates and classification accuracy for large and small class separation motivated an investigation of the conditions in between. Additional analyses were conducted for data representing moderate class separation, operationally defined as indicator mean difference of 1 standard deviation across latent classes. Due to the exploratory feature of this additional set of analysis, only 10 replications were conducted, and only for the unequal class proportion conditions, which proved to be the more challenging conditions for satisfactory model performance than equal class proportions.

Tables 4-34 and 4-35 report correct class extraction rates based on BIC, ABIC, and BLRT, and the overall and group classification accuracy of the studied conditions. Manipulated factors continued to present similar pattern of effects on class extraction rates and classification accuracy as they were with larger class separation. Consistent with the author’s expectation, larger sample sizes and number of indicators would be needed to achieve the same level of correct class extraction rates and classification accuracy with moderate class separation conditions than those with large class separation conditions. The detailed conditions were discussed in the next section on recommendation for practitioners.

Table 4-34. Correct class extraction rates for moderate class separation and unequal class proportion: continuous indicators

		BIC					ABIC					BLRT				
		Number of indicators														
		5	6	9	12	15	5	6	9	12	15	5	6	9	12	15
N	N_c	2-class														
50	5	10	20	20	30	30	0	0	0	0	0	10	30	**80**	40	**90**
100	10	10	40	30	50	60	0	20	0	0	10	60	50	70	70	**90**
200	20	0	40	**80**	**90**	**90**	0	**80**	40	40	50	**80**	**90**	**90**	**90**	**100**
300	30	0	**80**	**90**	**90**	**90**	0	**80**	60	60	**90**	**90**	**90**	**100**	**80**	**100**
400	40	0	**90**	**90**	**90**	**90**	0	**80**	60	60	**90**	**90**	**100**	**90**	**80**	**100**
500	50	0	**90**	**90**	**90**	**90**	0	**90**	**90**	**90**	**90**	**90**	**100**	**100**	**80**	**100**
1000	100	0	**90**	**90**	**90**	**90**	0	**90**	**90**	**90**	**90**	**100**	**90**	**100**	**90**	**100**
2000	200	0	**90**	**90**	**90**	**90**	0	**90**	**90**	**90**	**90**	**100**	**80**	**90**	**100**	**90**
5000	500	0	**90**	**90**	**90**	**90**	10	**90**	**90**	**90**	**90**	**90**	**100**	**90**	**90**	**90**
		3-class														
50	5	0	0	0	0	0	0	10	0	0	10	0	0	0	0	0
100	10	0	0	0	0	0	30	20	10	0	0	0	0	0	20	30
200	20	0	0	0	0	0	30	30	40	70	20	0	0	20	**80**	60
300	30	0	0	0	0	10	10	20	40	**80**	**80**	0	0	20	**80**	**80**
400	40	0	0	0	30	60	30	20	50	**90**	**90**	10	0	40	**90**	**90**
500	50	0	0	10	20	**90**	40	40	60	**100**	**100**	40	40	**80**	**90**	**90**
1000	100	10	30	60	**100**	**100**	60	60	**100**	**80**	**100**	**80**	60	**90**	**80**	**100**
2000	200	**80**	**80**	**100**	**100**	**100**	**100**	**100**	**100**	**90**	**100**	**100**	**90**	**100**	**90**	**100**
5000	500	**80**	**100**	**100**	**100**	**100**	**80**	**100**	**100**	**100**	**100**	**80**	**100**	**100**	**100**	**100**
		4-class														
50	5	10	10	10	10	10	10	10	0	0	10	0	0	0	0	0
100	10	0	10	10	10	10	10	10	30	30	0	0	0	0	0	0
200	20	10	10	10	0	10	20	0	50	30	50	10	0	0	20	0
300	30	10	10	0	10	10	40	20	10	50	40	0	0	0	20	30
400	40	0	0	10	10	10	10	30	50	80	70	0	0	30	50	60
500	50	0	10	10	0	10	0	20	60	70	70	0	0	70	70	**80**
1000	100	10	0	10	20	50	30	0	70	**90**	**90**	20	0	**80**	**80**	**90**
2000	200	20	10	40	70	**100**	40	20	**90**	**90**	**90**	30	50	**80**	**90**	**90**
5000	500	0	10	**90**	**100**	**90**	50	**80**	**90**	**100**	**90**	**90**	**90**	**90**	**90**	**80**

Note. N_c denotes the smallest class size for unequal class proportion conditions.

Table 4-35. Overall and group classification accuracy rates for moderate class separation and unequal class proportion: continuous indicators

		Number of Indicators									
		5	6	9	12	15	5	6	9	12	15
		Overall classification accuracy					Group classification accuracy				
N	N_c	2-class									
50	5	.71	.71	**.90**	**.91**	**.93**	.66	**.72**	**.85**	**.90**	**.79**
100	10	**.83**	**.83**	**.87**	**.97**	**.98**	.68	**.71**	**.83**	**.93**	**.88**
200	20	**.89**	**.92**	**.96**	**.98**	**.99**	.61	**.67**	**.84**	**.87**	**.94**
300	30	**.93**	**.94**	**.95**	**.98**	**.99**	.55	**.67**	**.78**	**.89**	**.94**
400	40	**.93**	**.94**	**.96**	**.98**	**.99**	.59	**.65**	**.75**	**.85**	**.94**
500	50	**.93**	**.94**	**.97**	**.98**	**.99**	.54	**.69**	**.81**	**.86**	**.92**
1000	100	**.94**	**.95**	**.97**	**.98**	**.99**	.53	**.66**	**.81**	**.85**	**.91**
2000	200	**.94**	**.95**	**.96**	**.98**	**.99**	.52	**.65**	**.78**	**.86**	**.92**
5000	500	**.94**	**.95**	**.97**	**.98**	**.99**	.55	**.65**	**.79**	**.86**	**.92**
		3-class									
50	5	.75	.73	**.84**	**.84**	**.88**	.30	.29	.29	.24	.56
100	10	.76	**.82**	**.86**	**.89**	**.89**	.30	.29	**.47**	**.52**	**.56**
200	20	.78	**.80**	**.84**	**.89**	**.91**	.34	.30	.45	**.65**	**.73**
300	30	.78	**.82**	**.87**	**.90**	**.92**	.33	.26	**.49**	**.68**	**.76**
400	40	**.80**	**.80**	**.87**	**.91**	**.94**	.35	.40	**.58**	**.70**	**.78**
500	50	.77	**.83**	**.88**	**.91**	**.94**	.43	.31	**.56**	**.70**	**.80**
1000	100	**.83**	**.85**	**.89**	**.92**	**.94**	.39	.34	**.59**	**.70**	**.78**
2000	200	**.85**	**.86**	**.90**	**.92**	**.94**	.37	.40	**.58**	**.70**	**.79**
5000	500	**.85**	**.87**	**.90**	**.92**	**.94**	.34	.41	**.59**	**.70**	**.79**
		4-class									
50	5	.51	.60	.68	.71	.70	.20	.32	.21	.47	.40
100	10	.56	.57	.64	.74	.73	.27	.15	.27	.27	.15
200	20	.61	.63	.69	.72	.75	.15	.21	.33	**.42**	**.48**
300	30	.61	.62	.69	.76	.79	.19	.17	.39	**.45**	**.49**
400	40	.59	.62	.73	.78	**.83**	.25	.25	.31	**.51**	**.59**
500	50	.63	.65	.75	**.80**	**.84**	.23	.26	**.40**	**.56**	**.61**
1000	100	.61	.68	.78	**.83**	**.86**	.32	.35	**.43**	**.51**	**.65**
2000	200	.71	.73	.79	**.84**	**.86**	.24	.32	**.47**	**.54**	**.61**
5000	500	.73	.74	.80	**.84**	**.86**	.25	.31	**.48**	**.56**	**.62**

Note. N_c denotes the smallest class size for unequal class proportion conditions.

Recommendations for practitioners

Similar to the situation with dichotomous indicators, class separation and true number of latent classes continue to show strong effects on correct class extraction rates and classification accuracy. Recommendations of sample sizes and number of indicators should take into account researchers' level of knowledge about those factors.

Like in the case for dichotomous indicators, LCM is not recommended when the classes are not well separated for continuous indicators. In this simulation study, small class separation is defined as half a standard deviation difference in indicator means across classes. Under such circumstances, LCM is unable to either distinguish models of different number of latent classes (e.g., refer to right panel of Table 4-21), or correctly assign individuals to their true latent classes (e.g., refer to right panels of Tables 4-28 and 4-30). For small class separation conditions, LCM may be used to classify individuals into two classes with relatively equal class sizes. For this purpose, sample sizes of at least 500 with 12 indicators or 300 with 15 indicators are recommended with the objective overall classification accuracy rate reaching .80 and group classification accuracy rate close to or above .80.

LCM is much more useful when the defined latent classes are well separated, which in this simulation is operationalized as 2 standard deviation differences in indicator means. Requirement on sample sizes and number of indicators differ depending on whether the purpose is to distinguish among models of different numbers of latent classes, or to classify individuals to latent classes.

If the purpose is to distinguish among models of different number of latent classes, cells with sample sizes of 200 or larger and 6 or more indicators fit the purpose when classes are well separated. With smaller sample sizes (e.g., 100), more indicators may be needed (e.g., 9); with larger sample sizes (e.g., 200), fewer indicators may be needed (e.g., 5).

If the purpose is to classify individuals rather than distinguish models with different numbers of latent classes, recommendation is made for conditions with a high overall classification accuracy (i.e., $\geq$ .80) and group classification accuracy statistically significantly higher than the random classification rates. Based on the above criteria, sample sizes of 50 and 5 indicators seem sufficient to classify individuals into 2 or 3 latent classes, and those combinations produce an average overall classification accuracy above .90 and group classification accuracy above .75. For classification into 4 latent classes, sample sizes of at least 100 with 5 or more indicators are recommended, which is expected to produce an average overall correct classification rate above .80 and group classification accuracy rate above .60.

For models with moderate class separation (operationally defined as 1 standard deviation in class-specific indicator means), sample sizes of 1,000 and 9 indicators are recommended to distinguish models with up to 4 latent classes, but sample sizes may be reduced to 500 to distinguish models with up to 3 latent classes. For classification purposes, sample sizes of 100 and 5 indicators seem sufficient for classification into 2 latent classes; sample sizes of 100 and 9 indicators are recommended for classification into 3 latent classes; for 4 latent classes, sample sizes of 400 and 12 indicators are recommended. Note that this recommendation may be conditional on the correct identification of latent classes if the researcher has very vague idea about the true number of latent classes, and has to decide on the correct number of latent classes first.

In order to simplify the recommendation for easier adoption in applied research, Table 4-36 lists recommendations from the perspectives of requirement on number of indicators and sample sizes per latent class for large class separation. When there are 3 or fewer latent classes, number of indicators per latent class is recommended to be larger than 1 with sample size of 100 or class size of 10, whichever is larger. When there are more latent classes, number of indicators is recommended to be 2 per latent class, with sample size of 100 or class size of 10, whichever is larger; when fewer indicators are available (e.g., >1 but <2 per latent class), sample size of 300 or class size of 30 is recommended, whichever is larger. Classification accuracy for the large class separation condition is acceptable in all of the studied conditions; therefore, sample size of 50 and 5 indicators are sufficient for classification up to 4 latent classes. However, note that this requirement should be conditional on the correct identification of number of latent classes (i.e., requirement for correct class extraction rates has to be considered instead) if number of latent class is unknown to the researcher.

In the case of moderate class separation, it is still recommended that at least 3 indicators per latent class and class size of 50 are needed for acceptable correct class extraction rates. The above conditions would produce overall classification accuracy rates of at least .80 and group classification rates above .65, .56 and .40 for 2-, 3- and 4-class models. Class size may be reduced if the researcher has prior knowledge on the number of latent classes and is interested in classification. Under such circumstances, 3 indicators per latent class size and class size of at least 20 are recommended.

In sum, a few general guidelines can be established for LCM with continuous indicators. First, LCM is again not recommended for latent classes with small separation. The current study suggests that a difference of half a standard deviation in class-specific indicator means may be too small for LCM, but the method already appears to work well when differences in class-specific indicator means increase to 1 standard deviation; in the case of large class separation with mean difference increasing to 2 standard deviations, model shows superb performance in most conditions. Second, requirement on sample sizes and number of indicators vary depending on the purpose of the research study; usually larger sample sizes and more indicators are required for the purpose of deciding on the optimal number of latent classes than to classify individuals with a known number of latent classes; conditions that produce acceptable correct class extraction rates usually produce acceptable overall and group classification accuracy rates too. Third, high entropy may not necessarily mean good classification accuracy, especially when sample sizes are small (e.g., <100); entropy is not recommended for use to compare the fit of samples with different sizes.

Table 4-36. Recommendation for LCM with continuous Indicators: large class separation condition[a]

Research Purpose	Number of latent classes	Number of indicators Per latent class	Sample size
Deciding on the Optimal number of Latent classes	≤ 3	>1	Class size $\geq$ 10 or sample size ≥ 100, whichever is larger
	4	>1	Class size $\geq$ 30 or sample size ≥ 300, whichever is larger
		≥ 2	Class size $\geq$ 10 or sample size ≥ 100, whichever is larger
Classification[b]	2, 3, 4	Sample sizes $\geq$ 50, number of indicators ≥ 5	

Note.

[a] large class separation is operationally defined as 2 standard deviation difference in indicator means. For moderate class separation (1 standard deviation difference), 3 indicators per latent class and class sizes of 50 are recommended for class extraction purposes, and 3 indicators per latent class and class sizes of 20 are recommended for classification purposes with known number of latent classes.

[b] Recommendation for classification purposes may be conditional on the correct identification of number of latent classes if number of latent classes is unknown.

Examples of Applying Recommendations in Practice

This section provides a few examples of how recommendations in Tables 4-17 and 4-36 may be used in practice. Several scenarios are listed to illustrate different conditions a researcher may face, from the study design stage for determining sample sizes and number of indicators to the analysis stage for the understanding of results.

Scenario 1: Design Stage with an Existing Set of Indicators

This scenario represents the situation when a researcher is interested in verifying the existence of postulated latent classes with an existing set of indicators. The question for the researcher at this stage is how large a sample he/she needs to recruit for the postulated latent classes to emerge.

A concrete example on mathematical achievement is adopted from the McDonald's (1999) book of Test Theory. Suppose that a researcher is interested in exploring the heterogeneity in mathematics achievement among high school students. The existing instrument is a 15-multiple choice item test where the first five items focus on geometry, the next five items focus on algebra, and the last five items require skills related to both geometry and algebra. Further suppose that the researcher hypothesizes that there are four latent classes of students in terms of their mathematical achievement: those good at only geometry (G), those good at only algebra (A), those good at both (GA) and those good at neither (N).

The researcher has 15 dichotomous indicators and expects to extract four latent classes, so the number of indicators per latent class is between 3 and 4, which meets the condition listed in Table 4-17 for dichotomous indicators. The purpose at this stage is not on the accurate classification of individuals, but rather whether the 4-class structure holds, so we need to focus on the top panel of the table. If the four latent classes are clearly distinct from one another (i.e., large class separation), then class size of 100 should be sufficient for the purpose of class extraction. If the proportion of students in the smallest latent class accounts for 10% of the population, then a total sample size of 1,000 may be needed in order for the smallest class to emerge.

The successful application of the above recommendation depends on the viability of the assumptions made, for example, whether the classes are clearly separated, and whether smallest class proportion is large enough, and whether indicators can effectively distinguish latent classes. When those assumptions are not met, we may not observe the expected results.

Scenario 2: Design Stage with Known Sample Size

This scenario represents the situation when a researcher knows about the sample size he/she may be able to get, and is at the stage of deciding on the number of indicators.

Again take students' mathematical achievement as an example. Suppose the researcher is interested in testing the hypothesis about the existence of four latent classes of students in terms of their mathematical achievement: those good at only geometry (G), those good at only algebra (A), those good at both (GA) and those good at neither (N). Further suppose that the proportion of students in the smallest latent class only accounts for 10% of the population. The researcher expects a sample size between 300 and 500, which means that the smallest class size might be as small as 30.

Note that in the case of unequal class proportions, sample sizes of 300 to 500 is relatively small for dichotomous indicators based on Table 4-17 because minimal class sizes is expected to be above 80 for 4-class models. However, continuous indicators seem to work well at such sample size level (see Table 4-36). So the recommendation in terms of scale of the indicators would be the continuous scale. With class sizes of 30, number of indicators per latent class larger than 1, or 5 or more continuous indicators altogether, should be sufficient provided that classes are clearly distinct from one another. This recommendation assumes that the indicators are effective. An indicator is effective if effect size of the largest between-class difference is large (operationally defined as 2 SD difference in this dissertation study). Note that this large difference is only required for the largest between-class difference, not between any pair of latent classes. In conditions when the researcher expects a smaller between-class difference, more indicators per latent class is needed.

Continuous indicators are relatively more difficult to obtain than dichotomous indicators. Dichotomous indicators are often represented by items, but continuous indicators are often represented by tests or scales. Therefore, in the case of the current scenario, the researcher may collect respondents' scores on a number of tests related to algebra and geometry. Another possible way to obtain continuous scales is to form parcels of dichotomous items. However, there are two unanswered questions with this approach: first, there should be sufficient number of items within each parcel so the continuous scale assumption is reasonable, but at this point how many items is considered sufficient is not clear; second, the current findings seem to suggest a compensating effect of a larger number of indicators for smaller sample size, which means that class sizes smaller than 100 may also work well for dichotomous indicators if a researcher has many dichotomous indicators. In other words, it is not clear when and to what extent parcel scores would enhance the performance of LCM compared with using the original dichotomous indicators.

Scenario 3: Analysis stage

In addition to applying recommendations from Tables 4-17 and 4-36 at the design stage, findings from this study may also be used to understand results from LCM at the analysis stage with existing data. One common issue in analysis with LCM is that sometimes, extracted latent classes do not conform to the hypothesis, and one or more latent classes may fail to emerge from existing data. This could occur because the hypothesis somewhat deviates from the truth, and the results just reflect the truth; but another possibility is that existing data do not provide a desirable condition for class extraction. In other words, the results may be a statistical artifact with imperfect data. Findings from this study may help decide whether results from LCM are limited by the available data.

For example, a researcher has collected data from a sample of 300 students on the 15-item test introduced in Scenario 1. If a researcher applies LCM to the data, and finds out that the 2-class model provides the best fit among the 1- to 5-class models, contrary to the researcher's expectation of 4 latent classes. Further suppose that two extracted latent classes represent a class of students who are good at both geometry and algebra, and another class of students who are good at neither. According to Tables 4-2 to 4-4, with sample sizes of 300, LCM is very likely to produce incorrect class extraction (underestimation of number of classes in the case of small sample sizes). This indicates that even if the 4-class model reflects the truth in reality, the researcher may not be able to observe the correct number of classes with small sizes. However, if with the same test, the researcher manages to obtain a sample of 1,000 students, and still the 2-class model fits the data the best, then he/she can be much more confident in saying that students are either good at both geometry and algebra or good at neither. However, this only excludes the possibility of incorrect class extraction due to smaller sample sizes, but not by other factors such as small class separation or ineffective indicators or even too small latent classes. Researchers may review each of those factors to determine whether it is the plausible cause or whether it is of interest. For example, in the current scenario, ineffective indicators might not be very likely if all indicators have been demonstrated to distinguish between latent class GA and latent class N.

Chapter 5

Discussions

This dissertation uses simulated data to examine the performance of LCM at various combinations of sample sizes and number of indicators under different conditions varying by class separation, true number of latent classes, and class proportions. Evaluation criteria include both correct class extraction rates (defined as percentage of replications with the correct number of latent classes identified) and classification accuracy (defined as percentage of individuals classified into appropriate latent classes). This study represents an initial attempt to address the concern for sufficient sample size and number of indicators used in LCM; it also enriches the current literature on the less frequently studied classification accuracy issues in LCM. Results reported in Chapter 4 lead to several conclusions regarding the use of LCM.

When is LCM not recommended

Perhaps the most striking message from this dissertation study is that LCM has very limited capability in deciding on the optimal number of latent classes or classifying individuals when classes are not well separated. As the most significant contributing factor to variance in LCM performance, class separation is not given as much attention as it deserves in either methodological studies or applied research.

In methodological studies, Tofighi and Enders (2007) is the only study that has looked at the effect of this factor on correct class extraction rates. The authors found a positive effect of high class separation on correct class extraction rates; however, the difference was not very large (i.e., just exceeding .10 in average correct class extraction rates). Their relatively small difference in correct class extraction rates may be explained by the smaller differences in indicator means between their high and low class separation conditions than what was studied in the current study. Note that in their study, high class separation was defined as 2.40 and .55 standard deviation differences in intercept parameters among three adjacent growth trajectories, while low class separation was defined as 1.90 and .44 standard deviation differences.

The class separation factor was largely unmentioned in the three major studies reviewed in Chapter 2, but all three studies generated data based on models that represented large class separation. For dichotomous indicators, differences in class-specific indicator probabilities in those studies were .50 (Yang, 2005) and .70 (Nylund et al., 2007); for continuous indicators, differences in class-specific indicator means were fixed at 2 standard deviations (Nylund et al.; Wu & Lei, 2008). Considering the dominant effect of class separation on correct class extraction rates, findings from those studies have to be interpreted taking into account the large class separation assumed in those studies.

Another challenging condition for LCM is the presence of a large number of latent classes. The largest number of latent classes included in the current study is 4, and it already poses great challenge for either class extraction or classification in many conditions. It is interesting to note that number of latent classes has more than a linear effect on required sample size and number of indicators. In other words, increase in sample sizes and number of indicators proportional to number of latent classes does not meet the increasing difficulty from models with more latent classes. A similar effect was observed by Yang (2005). Therefore, when number of latent classes is large (i.e., 4 or more), researchers should make extra efforts to obtain large sample sizes and sufficient number of indicators.

Choice and report of fit indices

Findings on both dichotomous and continuous indicators have pointed out BLRT to be the best fit index in general. This is consistent with previous research with either dichotomous indicators or continuous indicators (e.g., Nylund et al., 2007; Wu & Lei, 2008). Therefore, BLRT should be reported whenever it is available. However, one limitation of BLRT is that it utilizes re-sampling technique and is much more computationally intensive than IC-based fit indices such as BIC and ABIC, and thus not available in all statistical software programs that estimate LCM. For example, SAS's LCM macro does not produce BLRT.

Recommendation between BIC and ABIC has not been a simple or consistent one (Yang, 2005; Nylund et al., 2007; Wu & Lei, 2008). Yang's recommendation of ABIC over BIC was made when number of latent classes was 4 or larger with dichotomous indicators. This is consistent with findings from the current study with dichotomous indicators, or continuous indicators with moderate class separation. This may be partly explained by ABIC and BIC's varying penalties for model complexity. BIC and ABIC differ in their penalized terms for model complexity. While BIC uses the original sample size n as the multiplier in the penalty term, ABIC replaces n with n^*, which is equal to $(n+2)/24$. With positive sample size, n is always larger than n^*, so BIC imposes heavier penalty on more complicated models. As a result, BIC is more likely to favor the more parsimonious (i.e., simpler) models than ABIC. This is consistent with what is observed with BIC for the 4-class models: among all the 180 conditions for 4-class models with dichotomous indicators, BIC on average underestimates the number of latent classes in 87 out of the 100 replications while only overestimates it in less than 1 replication. One attractive feature with ABIC is that it seems to be more robust to unequal class proportion conditions than BIC, while BIC shows an advantage for conditions involving continuous indicators and large class separation, in either the current study or previous studies (e.g., Nylund et al.; Wu & Lei). Synthesizing findings regarding BIC and ABIC from all those studies, a tentative conclusion is that BIC appears to have better performance than ABIC in the most desirable conditions (fewer latent classes, equal class proportions, large class separation), however, ABIC may be more advantageous in the more practical conditions that are often represented by unequal class proportions and smaller class separation.

In addition to relative fit indices, this study also provides partial support for the reporting of absolute fit indices. For dichotomous indicators, χ^2 appears to perform better than G^2, which is consistent with Collins et al. (1993). The adverse effects of sparse data on those two statistics are very strong in the current study. Both statistics tend to be conservative (i.e., too few rejections) with sparse data in the studied conditions, leading to underestimation of number of latent classes. Take conditions with 15 indicators for example, whenever a model is incorrectly identified, it almost always underestimate the number of latent classes. Therefore, practitioners should keep in mind that χ^2 is not functioning well when number of indicators is large. Since χ^2 tends to be conservative with sparse data, when χ^2 is not rejecting a model, it is unclear whether the model really fits the data; however, when the test is rejecting a model, that may be a strong sign for further investigation.

For continuous indicators, the fourth-moment statistic MKT appears to work better than the third-moment MST. This is consistent with what was observed in Tofighi & Enders (2007) for growth mixture models. The limited availability of this statistic poses obstacles to its use in applied research. So far, Mplus is the only popular software program that provides this statistic. In addition, evidence supporting the use of it remains sporadic, perhaps due to the excessive amount of time needed to run simulations with large number of replications under various conditions. Findings from this study only provide some limited evidence for its reasonably controlled Type I error, however, this is based on a small number of replications (i.e., 10). Furthermore, its power in detecting incorrect model is still unknown. That is, when the model is correct, MKT is likely to not reject it; however, when a model is incorrect, it is unclear whether MKT can successfully reject it. Perhaps a tentative recommendation with regard to the use of MKT is to view a significant *p* value as a motivation for modifying the model.

Lastly, entropy has limited capability as a model selection index and correct class extraction rates based on entropy is worse than those based on BIC, ABIC or BLRT. More importantly, contrary to what Pastor et al. (2007) has suggested, it may be very misleading to use entropy to compare the fit of the same model with different samples. This is because when samples are of different sample sizes, those with smaller sizes are more likely to show higher entropy regardless of their actual fit to the postulated model.

Sample size and class size

One unresolved question has been whether class proportion per se or the resulting class size negatively affects the correct class extraction rates. If it is the latter case, increasing total sample size would ensure the proper class extraction. The current results clearly suggest that class proportion is not an incurable problem with LCM; as long as class size is sufficient, correct class extraction rates are acceptable. For example, for 3-class models with dichotomous indicators, equal class proportion conditions need sample sizes of 100 to yield satisfactory correct class extraction rates with 3 indicators per class, which is equal to a class size of 33. For the corresponding unequal class proportion conditions, sample sizes of 400 are needed. The smallest class size is 40 given the smallest class proportion of 10%.

Another point worth noting is that class size requirement is contingent upon number of latent classes and smaller class size is required for models with fewer latent classes. Again take dichotomous indicators for example, class size of 50 is sufficient for 2- or 3-class models with an adequate number of indicators per class (e.g., 3 or more). However, for 4-class models, class size of 50 would be too small for all studied conditions, but class size of 100 seems to be adequate.

Overall and group classification accuracy

Overall classification accuracy rate is a weighted average of group classification accuracy rates with class size as the weights. When classes are of equal sizes, group classification accuracy rate (operationally defined as the lowest group classification accuracy rate among all latent classes) should be comparable to overall classification accuracy rates because the weights are equal for each class; when classes are of unequal sizes, group classification accuracy rate is likely to be lower than overall classification.

It is interesting to note that class proportion shows apparently different directions of effects on overall and group classification accuracy rates in the current study. Overall classification accuracy rates tend to be higher for unequal class proportion conditions whereas group classification accuracy rates tend to be higher for equal class proportion conditions. This seemingly contradictory pattern occurs because individuals from the smaller classes tend to be assigned to the larger classes, resulting in a lower group classification accuracy for the smaller group, but higher group classification for the larger group, which is consistent with what was observed in Finch & Schneider (2006) using logistic regression.

One thing worth mentioning is that although LCM, similar to logistic regression, tended to classify individuals from the smaller group to the larger group, it may hold some promise for classification because its group classification accuracy rates for the smaller group appear to be somewhat higher than results from logistic regression from previous research at a similar level of class separation (e.g., Finch & Schneider, 2006). However, the two studies’ conditions are not directly comparable (number of indicators are smaller in Finch & Schneider), so further research is needed to verify the relative strength of the two methods in classifying individuals from smaller groups.

A caveat on the use of recommendations

Inherent in any simulation study, the investigated conditions may represent an artificial and sometimes over-optimistic set of conditions that researchers hope to characterize their real data. Although the simulated conditions in the current study have been carefully selected to reflect a vast array of what might occur with applied research, real data often do not match the profile provided here. The best way to gauge the specific need of a particular study is to use the specific setting of the intended study and then conduct a simulation study to search for desirable experimental designs. This would be an ideal approach, however, applied researchers may not have the luxury of time or personnel to implement this.

Therefore, cautions are advised when applying the more general profile provided in the recommendation to the specific model condition. For example, number of indicators in this study includes only "effective" indicators that distinguish among latent classes. Adding indicators that do not distinguish well among classes probably does little help in improving model performance. Also, unequal class proportions are operationalized as the smallest class containing 10% of the total sample. Less or more stringent requirement on sample size and number of indicators may be used for class proportions that are less or more extreme. More importantly, appropriate choice of sample size and number of indictors relies on the practitioner's understanding of class separation and true number of latent classes. All in all, the use of this recommendation without close examination of the specific model conditions may do more harm than help.

Limitations and future directions

The preceding paragraphs have already mentioned at least two future directions. First, it may be interesting to compare the classification accuracy of LCM with that of logistic regression when class proportions are unequal. Second, evidence pertaining to the use of MST and MKT is limited with a small number of replications, and future research may focus on the power of these two statistics as well as its Type I error. The following lists a few more limitations.

The findings on LCM's undesirable performance in the case of small class separation is intended to be more of a motivation for further research on how to improve on its performance rather than a total discourage of the use of LCM under such conditions. For example, one possible solution may be the addition of covariates. Research has found that when covariates are properly specified in the model, they can significantly increase classification accuracy (Lobke & Muthén, 2007); positive evidence with regard to increase in correct class extraction rates is still lacking though. Secondly, class separation has been identified to be the most influential factor in model performance, but the current study only includes two levels of class separation. This limits the practical use of the recommendation for many possible conditions standing in between. Future research can delve into more detailed levels of class separation conditions to inform applied research. Furthermore, a meta-analysis on class separation would be very helpful for the designs of future methodological studies.

In conclusion, findings from this study highlight the problems with some of the current practice with LCM, including the unquestioned use of small sample size and number of indicators. Class separation stands out as the most influential factor in model performance, highlighting the importance of more attention to this factor in both applied research and methodological studies.

Tentative recommendation on desirable model condition is proposed as an initial effort to promote the best practice with LCM.

Appendix

Mplus Codes for the Monte Carlo Procedure

title:

this is an example of a 2-class LCM with five binary latent class indicators

montecarlo:

names are u1-u5;

generate = u1-u5(1); !specifying the number of categories in indicators

categorical = u1-u5;

genclasses = c(2); !number of true latent classes

classes = c(2);

nobs = 500; !sample size

seed = 3454367;

nrep = 100; !number of replications

repsave = all;

save=c:\dissertation\c2u4s500*.txt;

analysis:

type = mixture;

model population:

%overall%

[c#1*0]; !specifying class proportion, logit scale

%c#1%

[u1$1*1 u2$1*1 u3$1*-1 u4$1*-1 u5$1*1]; ! specifying class-specific means, logit scale for dichotomous indicators

```
    %c#2%
    [u1$1*-1 u2$1*-1 u3$1*1 u4$1*1 u5$1*-1];
model:
    %overall%
    [c#1*0];
    %c#1%
    [u1$1*1 u2$1*1 u3$1*-1 u4$1*-1 u5$1*1];
    %c#2%
    [u1$1*-1 u2$1*-1 u3$1*1 u4$1*1 u5$1*-1];
output:
    tech8 tech9;
```

Bibliography

*Acosta, M. T., Casterllanos, F. X., Bolton, K. L., Balog, J. Z., Eagen, P., Nee, L., et al. (2008). *Journal of the American Academy of Child & Adolescent Psychiatry, 47,* 797-807.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317-332.

*Aldridge, A. A., & Roesch, S. C. (2008). Developing coping typologies of minority adolescents: A latent profile analysis. *Journal of Adolescence, 31*, 499-517.

Banfield, J.D. and Raftery, A.E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics, 49*, 803-821.

Bauer D.J. & Curran P.J. (2003). Distributional assumptions of growth mixture models : implications for overextraction of latent trajectory classes. *Psychological Methods, 8(3),* 338-363.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

*Bouwmeester, S., & Sijtsma, K. (2007). Latent class modeling of phases in the development of transitive reasoning. *Multivariate Behavioral Research, 42*, 457-480.

*Bowen, N., Lee, J., & Weller, B. E. (2007). Social environmental risk and protection: a typology with implications for practice in elementary schools. *Children & Schools, 29*, 229-242.

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

Brown, R. S. (April 2000). Using latent class analysis to set academic performance standards. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.

*Buckley, J., & Schneider, M. (2005). Are charter school students harder to educate? Evidence from Washington, D.C. *Educational Evaluation and Policy Analysis, 27*, 365-380.

Celeux, G. and Soromenho, G. (1996). “An Entropy Criterion for Assessing the Number of Clusters in a Mixture Model”. *Journal of Classification*, 13 (2): 195-212.

*Chan, W., Leu, Y., & Chen, C. (2007). Exploring group-wise conceptual deficiencies of fractions for fifth and sixth graders in Taiwan. *The Journal of Experimental Education, 76*, 26-57.

*Chung, T., & Martin, C. S. (2005). Classification and short-term course of DSM-IV cannabis, hallucinogen, cocaine, and opioid disorders in treated adolescents. *Journal of Consulting and Clinical Psychology, 73,* 995-1004.

*Chung, T., Martin, C. S., Cornelius, J. R., & Clark, D. B. (2008). Cannabis withdrawal predicts severity of cannabis involvement at 1-year follow-up among treated adolescents. *Addiction, 103*, 787-799.

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* (Ch. 6; pp. 311-359). New York: Plenum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum, Hillsdale, New Jersey.

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(3)*, 375-389.

Cudeck, R., & Henley, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size. *Psychological Bulletin, 109*, 512–519.

D’Angiulli, A., Siegel, L. S., & Maggi, S. (2004). Literacy instruction, SES, and word-reading achievement in English-language learners and children with English as a first language: a longitudinal study. *Learning disabilities research & practice, 19*, 202-213.

Davey, B. & Macready, G.B. (1990). An exploration of latent structure underlying multiple-choice reading comprehension items for good and poor readers. *Applied Measurement in Education, 3*, 209-229.

*DeCarlo, L. T. (2005). A model of rater behavior in essay grading based on signal detection theory. *Journal of Educational Measurement, 42*, 53-76.

*DeCarlo, L. T., & Luthar, S. S. (2000) Analysis and class validation of a measure of parental values perceived by early adolescents: an application of a latent class model for rankings. *Educational and Psychological Measurement, 60*, 578-591.

*DiStefano, C., & Kamphaus, R. W. (2006). Investigating subtypes of child development: A comparison of cluster analysis and latent class cluster analysis in typology creation. *Educational and Psychological Measurement, 66*, 778-794.

*Dush, C. M. K., Taylor, M. G., & Kroeger, R. A. (2008). Marital happiness and psychological well-being across the life course. *Family relations, 57*, 211-226.

Fan, X. & Wang, L. (1999). Comparing linear discriminant function with logistic regression for the two-group classification problem. *Journal of Experimental Education, 67,* 265-286.

Finch, H. & Schneider, M.K. (2006). Misclassification rates for four methods of group classification: Impact of predictor distribution, covariance inequality, effect size, sample size and group. *Educational and Psychological Measurement, 66(2)*, 240-257.

*Geiser, C., Lehmann, W., & Eid, M. (2006). Separating “Rotators” from “Non-Rotators” in the Mental Rotations Test: A multigroup latent class analysis. *Multivariate Behavioral Research, 41*(3), 261-293.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215-231.

Goodman, L. A. (2002). Latent class analysis: the empirical study of latent types, latent variables, and latent structures. In Applied Latent Class Analysis (Eds. Hagenaars J. A., McCutcheon A. L), Cambridge University Press, Cambridge, UK.

*Guttentag, C. L., Pedrosa-Josic, C., Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Individual variability in parenting profiles and predictors of change: effects of an intervention with disadvantaged mothers. *Journal of Applied Developmental Psychology, 27*, 349-369.

Haberman, S. J. (1974). *The Analysis of Frequency Data.* Chicago: University of Chicago Press.

Hill, A. L., Degnan, K. A., Calkins, S. D., Keane, S. P. (2006). Profiles of externalizing behavior problems for boys and girls across preschool: the roles of emotion regulation and inattention. *Developmental Psychology, 42,* 913-928.

*Herman, K. C., Ostrander, R., Walkup, J. T., Silva, S. G., & March, J. S. (2007). Empirically derived subtypes of adolescent depression: latent profile analysis of co-occurring symptoms in the treatment for adolescents with depression study (TADS). *Journal of Consulting and Clinical Psychology, 75*, 716-728.

*Jackson, K. M. & Sher, K. J. (2008). Comparison of longitudinal phenotypes based on alternate heavy drinking cut scores: A systematic comparison of trajectory approaches III. *Psychology of Addictive Behaviors, 22*, 198-209.

*Jansen, B. R. J., & van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology, 81,* 383-416.

Kaplan, D. (2009). *Structural Equation Modeling: Foundations and Extensions (2nd Edition).* Newbury Park, CA: SAGE Publications.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guildford Press.

*Klonsky, E. D., & Olino, T. M. (2008). Identifying clinically distinct subgroups of self-injurers among young adults: A latent class analysis. *Journal of Consulting and Clinical Psychology, 76*, 22-27.

Lazarsfeld, P. F., and Henry, N. W. (1968), *Latent Structure Analysis*, Boston: Houghton Mifflin.

*Liao, P., & Tu, S. (2006). Examining the scalability of intimacy permissiveness in Taiwan. *Social Indicators Research, 76*, 207-232.

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

*Loken, E. (2004). Using latent class analysis to model temperament types. *Multivariate Behavioral Research, 39*, 625-652.

Lord F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale NJ: Erlbaum.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14*, 26-47.

*MacMillan, R., & Copher, R. (2005). Families in the life course: interdependency of roles, role configurations, and pathways. *Journal of Marriage and Family, 67*, 858-879.

*Macy, R. J., Nurius, P. S., & Norris, J. (2007). Latent profile among sexual assault survivors: Understanding survivors and their assault experiences. *Journal of Interpersonal Violence, 22*, 520-542.

McCutcheon, A. L. (1987). *Latent class analysis*. Thousand Oaks, CA: Sage.

McCutcheon, A. L. (2002) "Basic Concepts and Procedures in Single and Multiple Group Latent Class Analysis." In J.A. Hagenaars and A.L. McCutcheon (eds.) *Applied Latent Class Analysis*. New York: Cambridge University Press. Pgs. 56-88.

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Lawrence Erlbaum Associates.

McLachlan, G., & Peel, D. (2000). *Finite Mixture Models.* New York: John Wiley.

*Morgan-Lopez, A. A., & Fals-Stewart, W. (2007). Analytic methods for modeling longitudinal data from rolling therapy groups with membership turnover. *Journal of Consulting and Clinical Psychology, 75*, 580-593.

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1-34). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29,* 81-117.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling : Comment on Bauer and Curran. *Psychological Methods, 8*, 369-377.

Muthén, B. (2008). Latent variable hybrids: overview of old and new models. In G. R. Hancock, Samuelsen, K. M. (Eds.), *Advances in Latent variable mixture models* (pp. 1-24). Charlotte, NC: Information Age Publishing.

Muthén, B., & Asparouhov, T. (2002). *Mixture testing using multivariate skewness and kurtosis.* Manuscript in preparation.

Muthén, L. K., & Muthén, B. O. (2007). M*plus* 5.0 [Computer software]. Los Angeles, CA: Muthén & Muthén.

*Notenboom, A., & Reitsma, P. (2007). Spelling dutch doublets: Children's learning of a phonological and morphological spelling rule. *Scientific Studies of Reading, 11,* 133-150.

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.

*Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development, 78*, 1706-1722.

*Oxford, M. L., Gilchrist, L. D., Lohr, M. J., Gillmore, M. R., Morrison, D. M. & Spieker, S. J. (2005). Life course heterogeneity in the transition from adolescence to adulthood among adolescent mothers. *Journal of Research on adolescence, 15*, 479-504.

*Park, A., Sher, K. J., & Krull, J. L. (2008). Risky drinking in college changes as fraternity/sorority affiliation changes: A person-environment perspective. *Psychology of Addictive Behaviors, 22*, 219-229.

*Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology, 32,* 8-47.

Raftery, A.E. (1995). Bayesian model selection in social research (with Discussion). *Sociological Methodology*, 25, 111-196.

*Raijmakers, M. E. J., Jansen, B. R. J., & van der Maas, H. L. J. (2004). Rules and development in triad classification task performance. *Developmental Review, 24,* 289-321. Reierse

*Reiersen, A. M., Constantino, J. N., & Todd, R. D. (2008). Co-occurrence of motor problems and autistic symptoms in Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 47,* 662-672.

*Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically derived subtypes of child academic and behavior problems: co-occurrence and distal outcomes. *Journal of Abnormal Child Psychology, 36*, 759-770.

*Romano. E., Zoccolillo, M., & Paquette, D. (2006). Histories of child maltreatment and psychiatric disorder in pregnant adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 45,* 329-336.

Stone, M., & Yumoto, F. (2004). The effect of sample size for estimating Rasch/IRT parameters with dichotomous items. *Journal of Applied Measurement, 5*, 48-61.

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

*Striegel-Moore, R. H., Franko. D. L., Thompson, D., Affenito, S., May, A., & Kraemer, H. C. (2008). Exploring the typology of night eating syndrome. *International Journal of Eating Disorders, 41,* 411-418.

*Tapola, A. & Niemivirta, M. (2008). The role of achievement goal orientations in students' perceptions of and preferences for classroom environment. *British Journal of Educational Psychology, 78*, 291-312.

Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287-305.

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.

van de Pol, F., Langeheine, R., & De Jong, W. (1996). *PANMARK 3: Panel analysis using Markov chains—A latent class analysis program* [user manual]. Voorburg, The Netherlands.

van den Wittenboer, G., Hox, J.J. & de Leeuw, E.D.(1997). Aberrant response patterns in elderly respondents: latent class analysis of respondent scalability. Pp. 155-162 in J. Rost & R. Langeheine (eds). *Applications of latent class models in the social sciences.* Münster/New York: Waxmann.

*van Gaalen, R. I., & Dykstra, P. A. (2006). Solidarity and conflict between adult children and parents: A latent class analysis. *Journal of Marriage and Family, 68,* 947-960.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-106). Cambridge, UK: Cambridge University Press.

Vermunt, J. K. and Magidson, J. (2005), *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Belmont Massachusetts: Statistical Innovations Inc.

Volk, H. E., Neuman, R. J., Todd, R (2005). A systematic evaluation of ADHD and comorbid psychopathology in a population-based twin sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 44,* 768-775.

Vuong Q. H. (1989). Likelihood ratio test for model selection and non-nested hypothesis. *Econometrica, 57,* 307-333.

*Webb, M. L., Cohen, A. S., & Schwanenflugel, P. J. (2007). Latent class analysis of differential item functioning on the Peabody Picture Vocabulary Test III. *Educational and Psychological Measurement*, *68*, 335-351.

Whiteman, S., & Loken, E. (2006). Comparing analytic techniques to classify dyadic relationships: an example using siblings. *Journal of Marriage and Family, 68*, 1370-1382.

*Witkiewitz, K., & Masyn, K. E. (2008). Drinking trajectories following an initial lapse. *Psychology of Addictive Behavior, 22*, 157-167.

Wu, Q., & Lei, P. W. (2008). Comparing model fit indices for latent class models with continuous indicators. Paper presented at the Annual Meeting of American Educational Research Association, New York City, New York.

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

Yang, X., Shaftel, S., Glasnapp, D., & Poggio, J. (2005) Qualitative or quantitative differences? Latent class analysis of mathematical ability for special education students. *The Journal of Special Education, 38,* 194-207.

Vita

Qiong Wu was born in 1982 in Zongyang, China. She was raised and educated in the same town until she was eighteen. In 2000, she was enrolled in the University of Science and Technology of China, Hefei, China, where she received her dual degree in English and Computer Science in 2004. In the same year, she became a doctoral student in Educational Psychology at the Pennsylvania State University, University Park. She met her husband at the same university and got married in 2007.