Title
The Intersection of Sample Size, Number of Indicators, and Class Enumeration in LCA: A Monte Carlo Study

Author
Morovati, Diane

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

The Intersection of Sample Size, Number of Indicators, and Class Enumeration in LCA: A Monte Carlo Study

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Education

by

Diane Morovati

Dissertation Committee:

Professor Karen Nylund-Gibson

Professor John Yun

Professor Nancy Collins

June 2014

The dissertation of Diane Morovati is approved.

John T. Yun

Nancy L. Collins

Karen Nylund-Gibson, Committee Chair

June 2014

The Intersection of Sample Size, Number of Indicators, and

Class Enumeration in LCA: A Monte Carlo Study

by

Diane Morovati

To my Mom,

Thank you for teaching me to work hard and never give up.

This is just the beginning…

ACKNOWLEDGEMENTS

First and foremost, I acknowledge my parents, Lia and Morad Morovati, for unconditionally loving me and providing me with so much to be grateful for. This dissertation is dedicated to my mom because she deserves special praise for always pushing me to strive for the best, even when I wanted to give up. Mom, thank you for always believing in me. Thank you for raising me to be strong and independent. I wouldn't be where I am today if it wasn't for you. I love you and I am so proud to be your daughter.

To my younger sister and best friend, Dahlia Morovati, and my little brother, Daniel Morovati, words cannot express my love for you both. I am so proud of you and look forward to your many future accomplishments. I want to thank my grandmother, Faezehjoon, for being such a strong and supportive woman in my life. I am very fortunate for your wisdom and guidance. You truly mean the world to me. Thank you for being unconditionally proud of me. I also want to thank Vicki and Shahram Moradzadeh, Jack Lahidjani, and my beautiful cousins for being a huge part of my strong support system. I am so lucky to have you all in my corner. I know I can always count on you.

To my best friends and extended family, Jill and Tamara, thank you for the laughs and good times we have shared for the past decade. Thank you for listening to me vent, keeping me sane and grounded, and for giving me much needed distractions and reality checks. I am so fortunate to have wonderful friends like you. Jill, thank you for being such a good and supportive friend, for reading countless drafts of my work, and for just being there for me. To my amazing boyfriend Chris, I am so thankful for your daily love, support, and encouragement, even from 150 miles away. Thank you for never making me feel like we were far apart, for sticking by my side, and for just being you.

I thank my dissertation committee, Karen Nylund-Gibson, Nancy Collins, and John Yun, for their support and feedback, all of which helped make this dissertation stronger. Nancy, thank you for being such an encouraging committee member. I appreciate your support, feedback, and positive attitude. John, thank you for supporting me from my master's defense through my dissertation defense. I appreciate every minute of your time and all of your feedback. Karen, thank you for being an amazing and brilliant advisor. I appreciate all of the knowledge, support, and mentorship you have given me for the past 5 years. As I approach graduation, I realize how much I have truly learned from you, and how much you have positively contributed to my graduate school experience and professional development. I can't express into words how appreciative I am for everything you have done for me. I am so very thankful for the opportunity to work with you and learn from you.

I also want to thank the Latent Variable Group (LVG). I learned so much by doing research with you all and appreciate every single one of you. Thank you for being supportive and providing me with a strong sense of community at UCSB. I want to especially thank Lauren Taylor for being the best office buddy, friend, and colleague. I will miss working next to you every single day. Lastly, I have a wonderful group of friends who I have met throughout graduate school, each of whom supported me in some way through this process: Brittany Koegel, Chloe Lenow, Amber Gonzalez, Shelley Hart, Ann Kim, and Myley Dang. Thank you so much for being great friends and peers. Lastly, I want to thank UCSB for an amazing 5 years. I will forever value the knowledge and experience that I gained while attending the Gevirtz Graduate School of Education (GGSE). Go Gauchos!

VITA OF DIANE MOROVATI

June 2014

EDUCATION

May 2009 Bachelor of Arts in Psychology, California State University, Northridge

June 2012 Master of Arts in Education, University of California, Santa Barbara

June 2014 Doctor of Philosophy in Education, University of California, Santa Barbara

FIELD OF RESEARCH

Research Methods in Education, Department of Education, University of California, Santa Barbara

Quantitative Methods in the Social Sciences, Interdisciplinary Ph.D. Emphasis, University of California, Santa Barbara

RESEARCH EXPERIENCE

2007-2009 Student Research Assistant, Dr. Ellie Kazemi, Department of Psychology, California State University, Northridge

2008-2009 Student Research Assistant, Dr. Howard B. Lee, Department of Psychology, California State University, Northridge

2009- Present Graduate Student Researcher, Dr. Karen Nylund-Gibson, Department of Education, University of California, Santa Barbara

2012-2013 Research Consultant, Online General UCSB Catalog Survey, University of California, Santa Barbara

EMPLOYMENT

2009-2010 Behavioral Clinician, Koegel Autism Consultants, University of California, Santa Barbara

2009-2010 Research Consultant, Dr. Russell Rumberger, EAR (Engagement, Alignment, & Rigor) Project, Department of Education, University of California, Santa Barbara

2010-2014 Teaching Assistant, Department of Communication, University of California, Santa Barbara

2013-2014 Market Research Intern, Screen Engine LLC, Santa Monica, CA

2014-Present   Sr. Research Scientist, The Nielsen Company, Los Angeles, CA

<u>CONFERENCE PRESENTATIONS</u>

Gottfried, A. E., Nylund-Gibson, K., Gottfried, A. W., Morovati, D., & Gonzalez, A. M. (April, 2013). *Modeling transitions from intrinsic motivation to need for cognition and educational attainment: Childhood through adulthood*. Presented at the 2013 American Educational Research Association (AERA) conference, San Francisco, California.

Morovati, D., Nylund-Gibson, K., & Hart, S. (April, 2011). *Comparing latent class models with and without sampling weights using the NCVS-SCS dataset.* Paper presented at the 2011 Western Psychological Association (WPA) conference, Los Angeles, California.

Morovati, D., Nylund-Gibson, K., & Maysn, K. (February, 2012). *The consequences of latent classes: Approaches to specifying the effects of latent class membership on distal outcomes.* Paper presented at the 2012 Society for Research in Child Development (SRCD) conference, Tampa, Florida.

Morovati, D., Rylander, A., Norstedt, J., & Kazemi, E. (April, 2009). *The relationship between training satisfaction, job satisfaction, and burnout*. Presented at the 2009 Western Psychological Association Conference (WPA), Portland, Oregon.

Rylander, A., Norstedt, J., Morovati, D. & Kazemi, E. (May, 2009). *The relationship of training and job satisfaction in behavioral therapist retention*. Presented at the annual convention of the Association for Behavior Analysis International (ABAI), Phoenix Convention Center, Phoenix, Arizona.

<u>PUBLICATIONS</u>

Askew, J., Jones-Wiley, D. G., Morovati, D. R., & Lee, H. B. (2008). Evidence for validity of the revised Levinson and Sanford anti-Semitism scale. *Psychological Reports, 103,* 604-606.

Dowdy, E., Nylund-Gibson, K., Felix, E., Morovati, D., Carnazzo, K., & Dever, B. V. (2013). Long-term stability of screening for behavioral and emotional risk. *Educational and Psychological Measurement, 73,* 453-472.

Gottfried, A. E., Nylund-Gibson, K., Gottfried, A. W., Morovati, D., & Gonzalez, A. M. Longitudinal trajectories from academic intrinsic motivation to need for cognition and educational attainment: Childhood through adulthood. *Manuscript submitted for publication.*

Morovati, D. R., Steinberg, A. L., Taylor, L. C., & Lee, H. B. (2008). Further validation evidence for the Pet Attitude Scale. *North American Journal of Psychology*, *10*, 543-552.

COMMITTEE MEMBERSHIP

2012-2013 Quantitative Faculty Hiring Committee, University of California, Santa Barbara

ABSTRACT

The Intersection of Sample Size, Number of Indicators, and Class Enumeration in LCA: A Monte Carlo Study

by

Diane Morovati

This Monte Carlo simulation study examined the performance of the most commonly used fit indices in selecting the "correct" latent class model while varying factors such as: the true number of latent classes, the size of the latent classes (i.e., class prevalence), the nature of the latent classes, the number of indicators, and sample size. Specifically, the fit indices examined in this simulation study were the Akaike Information Criterion (AIC), the Consistent Akaike Information Criterion (CAIC), the Bayesian Information Criterion (BIC), the adjusted Bayesian Information Criterion (ABIC), the adjusted Lo-Mendell-Rubin likelihood ratio test (LMR-LRT), the parametric bootstrapped likelihood ratio test (BLRT), the approximate Bayes Factor (BF), and the correct model probability (cmP). No study to date has examined the performance of the BF and cmP in recovering the correct latent class model.

This simulation study also aimed to simultaneously examine and understand how sample size, the number of observed indicators, and class enumeration intersect in latent class analysis (LCA) models. In other words, when sampling observations from a larger population, is there a critical point where the size of the sample and the number of indicators cannot uncover all existing heterogeneity? That is, at what point is specificity of the emerging latent classes lost?

All data were generated and analyzed using Mplus latent variable software (Muthén & Muthén, 1998-2013). The specific data generation and analysis conditions in this dissertation were created based on a literature search of Education and Psychology related databases. Results from this study will help applied researchers using LCA models further understand which fit index to trust under various conditions when going through the class enumeration process in practice. Specifically, the ABIC and BLRT indices emerged as being the highest performing across a variety of conditions considered in this study. Results also highlight the practical importance of thoughtfully considering sample size and the number of indicators included when estimating and interpreting LCA models. Findings of this dissertation provide evidence for a relatively strong interplay between sample size, number of indicators, and class enumeration in LCA models.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Overview of Mixture Models and LCA

Mixture modeling is a data analysis technique used to uncover and model unobserved heterogeneity in a population by identifying different groups of individuals based on their observed response patterns (Nylund, Asparouhov, & Muthén, 2007). Membership into these groups (or latent classes) is often not known *a priori*, but is instead inferred from the data (Muthén & Muthén, 1998-2011). This type of modeling is also known as finite mixture modeling in that it expresses the overall distribution of one or more variables as a *mixture* of a finite number of component distributions or subpopulations, usually more homogenous in nature when compared to the overall distribution (Masyn, 2013; McLachlan & Peel, 2000). One type of cross-sectional[1] mixture model is latent class analysis (LCA). The main objective of LCA is to categorize people into latent groups, called classes, based on their response patterns to binary (or categorical) indicators. Additionally, LCA aims to simultaneously identify indicators that best distinguish between these identified latent classes (Nylund et al., 2007a). Furthermore, LCA treats the underlying class variable as an unobserved (or unknown), categorical latent variable.

1.1.1 Applications and Advantages of LCA

These statistical models that classify individuals into homogenous subgroups have been shown to have important applications in many substantive research areas,

[1] Data collected from a population at one specific point in time.

such as Education and Psychology. For example, LCA has been used to classify adolescents in middle school based on their victimization experiences (Nylund, Bellmore, Nishina, & Graham, 2007). Specifically, this study empirically derived three subgroups of students that differed in victimization severity rather than the specific type of victimization they experienced (e.g., physical versus verbal victimization) (Nylund et al., 2007b). In other words, one class was composed of adolescents who were characterized by consistently high probabilities of victimization, while a second class included adolescents who were characterized by moderate probabilities of victimization, and a third class included adolescents who were characterized by extremely low probabilities of victimization (Nylund, et al., 2007b). Moreover, this study also compared these three empirically derived latent classes to more traditional methods of creating groups based on predetermined cut scores (i.e., both raw and standardized cut scores[2]). LCA and creating groups based on cut scores are both ways to classify people into smaller, more homogenous groups. However, a key difference between LCA and various cut score methods is that class membership is treated as unknown in LCA. Alternatively, the groups that are identified using cut score methods are based on some determined cutoff criteria. Findings of this study revealed that there was some overlap in classification between the LCA groups and the sample dependent cut score groups (Nylund et al., 2007b).

[2] Standardized cutoffs are often utilized in practice however this is potentially problematic because an individual's group membership is sample dependent. This means that the same individual with the same score on an observed variable can end up in different groups depending on varying characteristics in different samples (i.e., means and standard deviations of the observed variable) (Nylund, Bellmore, Nishina, & Graham, 2007). LCA, on the other hand, is a model based or probabilistic approach, which implies that the latent classes can be replicated with independent samples (Muthén & Muthén, 2000).

However, compared to the LCA groups, the raw score cutoffs tended to underestimate students that were frequently victimized and the standardized cutoffs overestimated students that were frequently victimized (Nylund et al., 2007b).

In another study, LCA was used to identify three subgroups of individuals that suffered from various eating disorders based on their observed symptoms and psychopathology (Eddy et al., 2010). An advantage of using LCA over other classification methods in this study was that LCA provided a more holistic approach to creating groups that considered an array of variables and estimated different latent classes based on the patterns that emerged among these set of variables. Therefore, the researchers were able to take a *multivariate* approach to empirically identifying and deriving subgroups of individuals that have an eating disorder. Results indicated one latent class including individuals who reported binge eating and purging behaviors, another class comprised of individuals who reported excessive amounts of exercise and extreme eating cognitions, and a third class that was characterized by minimal eating behaviors (Eddy et al., 2010). A second advantage of using LCA in this context was that researchers were then able to compare the empirically derived subgroups to established eating disorder categories in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). This was an important advantage for the researchers because the empirically derived latent classes presented a more nuanced picture of eating disorder phenotypes when compared to the established DSM-IV-TR categories (Eddy et al., 2010). Specifically, the authors concluded that the emergent latent classes could be used to improve the DSM-IV-TR categories.

In a study looking at patterns of violence against women, LCA was used to understand and predict mental health outcomes later in life, such as depression and posttraumatic stress disorder (Cavanaugh et al., 2012). The results indicated four different patterns of violence, which included a class experiencing low amounts of violence, a class experiencing high physical and psychological intimate partner violence, a class experiencing high physical and psychological workplace violence, and a class experiencing moderate to high childhood abuse. When compared to the low violence class, those that experienced intimate partner violence and childhood abuse displayed more depressive symptoms later in life. Additionally, those in the intimate partner violence class also experienced more symptoms of posttraumatic stress at a six month follow up assessment (Cavanaugh et al., 2012). An advantage of using LCA in this context was that it allowed researchers to investigate and identify different combinations of violent experiences and how those experiences collectively related to various mental health outcomes later in life (i.e., depression and posttraumatic stress disorder), rather than how each experience related separately.

Taken together, these various mixture model applications illustrate how LCA can be used to understand, classify, and summarize individuals based on their observed profiles. Accurate classification of individuals into these various latent classes provides potential advantages. Specifically, it allows researchers to tailor intervention and educational programs differently for individuals that vary in their observed characteristics and experiences. For example, individuals reporting minimal eating behaviors in the Eddy et al. (2010) study might benefit from a different type of intervention program than individuals reporting excessive binging and purging behaviors (Eddy et al., 2010). Similarly, individuals that are victimized daily by their peers would benefit from a more intensive, frequent intervention program when compared to individuals who report only being victimized occasionally (Nylund et al., 2007b).

Furthermore, researchers may want to classify people into more homogenous groups with the goal of comparing the empirically derived classes to proposed theoretical subgroups. For example, a previous study empirically derived six latent classes of individuals diagnosed with attention deficit hyperactivity disorder (ADHD) and then compared them with already established DSM-IV ADHD subtypes (Elia et al., 2009). Results showed that some of the latent classes corresponded to specific DSM subtypes, while other latent classes could be used to further define ADHD phenotypes (Elia et al., 2009). Lastly, classifying individuals into homogenous groups allows researchers to explore whether the groups differ on various outcome measures (e.g., proximal or distal outcomes). For example, research has shown that adolescents who are victimized display higher levels of depression later in life when compared to adolescents who are not victimized (Nylund et al., 2007b).

1.2 Establishing Best Practices in Mixture Modeling

Since their gain in popularity (Muthén & Muthén, 2000; Nagin, 1999), there have been a lot of advances in the type and complexity of mixture models, but not nearly as much attention has been given to understanding how we best use these models in practice, and the consequences of different model specifications. While the applications of mixture models have provided insight into a wide range of substantive areas, there are still methodological questions about the best practices of these models that have yet to be addressed. Although some progress has been made in this area, there are still many unanswered questions that would add to researcher confidence and knowledge about how to effectively apply LCA models in practice. Thus, there is still work to be done in terms of establishing best practices in the application of mixture models, which can most effectively be demonstrated through the use of Monte Carlo simulation studies.

1.2.1 Monte Carlo Simulation Studies

Monte Carlo simulation studies allow researchers to generate data with a known set of truths or population parameters and then analyze it under different modeling conditions, which provide an ideal context to establish best practices in methodology. Simulation studies are advantageous because knowing the underlying *Truth* in the data gives researchers an understanding of the utility of the statistical models employed, their ability to precisely recover model parameters under different modeling conditions (i.e., different sample sizes), and the accuracy of the fit indices that are used to evaluate model fit. This differs from the process of working with real data where the *Truth* is unknown. Many of the simulation studies that have focused on examining mixture models have aimed to understand how fit indices perform in identifying the correct number of latent classes under various models and data conditions, and to improve the overall accuracy of classifying individuals into these latent classes. One influential study by Bauer and Curran (2003a) used simulation study methodology and focused on the implications of having non-normal data when fitting growth mixture models using continuous outcomes. They concluded that when data is non-normal, it might be difficult to distinguish between a single-class model with observed variables that are not normally distributed and a true mixture model. The findings of this study revealed that data drawn from a non-normal distribution may be more susceptible to over extracting latent classes and in turn may produce latent classes that are not substantively interpretable.

These findings from Bauer and Curran (2003a) had important implications for research because they were the first to provide some cautions when using mixture models, skepticism of modeling results, while also offering recommendations for use. The recommendations made in this paper are useful since they discouraged poor applications of mixture models and also highlighted the possibility of alternative models that may fit the data equally well (Muthén, 2003). In a rejoinder, Muthén (2003) argued that appropriate model checking procedures would help identify and differentiate between data that is truly composed of a mixture of normal distributions from data that is non-normal and homogenous in nature. Additionally, Muthén (2003) argued that there are cases where alterative models exist, and that substantive theory should be used to guide researchers in selecting which final model to retain. He labeled this process of deciding on the best fitting model as "substantive and statistical checking" (Muthén, 2003), a modeling recommendation that is widely used across a host of mixture models today. However, in instances where substantive theory cannot distinguish between two alternative models, mixture models may still provide useful insights into the data (Bauer & Curran, 2003b). It is now suggested that applied researchers think critically about the mixture models they are fitting and to always be open to the existing possibility of alternative models (Bauer & Curran, 2003b; Muthén, 2003). While these studies focused on growth mixture models, much of the modeling recommendations apply to what is currently considered to be best practices for all mixture models.

Since Bauer and Curran's (2003a) study, other simulation studies in LCA contexts often generate data with a set of known population parameters (e.g., a known number of latent classes, item distributions, class size, etc.) and then analyze these data under various modeling conditions (e.g., models with a different number of latent classes than the generated data). The goal of these studies is to ultimately evaluate the performance of the various fit indices that are used to identify the "true" number of latent classes. The majority of LCA simulation research thus far has focused on understanding the performance of various fit indices in deciding on the correct number of latent classes (Nylund et al., 2007a; Tofighi & Enders, 2007; Yang, 2006). Unfortunately this research is not substantial and there is still a need to generalize what we currently know about fit index performance across a variety of different contexts. Another area of simulation research has also examined the connection between retaining the correct number of latent classes, within-class sample size, and mean differences between latent classes (i.e., class separation[3]) in

[3] For the purposes of this study, class separation can be defined as the physical distance between the latent classes and will be discussed in more detail in Chapter 3.4.

latent profile analysis (LPA) (Lubke & Neale, 2006). LPA[4] can be thought of as a latent class model with continuous, instead of binary indicators. Findings from Lubke and Neal (2006) revealed that the number of latent classes retained was dependent on both within-class sample size and class separation. That is, when class separation was small, increasing the within-class sample size increased the likelihood of recovering the correct number of latent classes. Moreover, results from a more recent simulation study concluded that power to detect a true latent class solution is dependent on factors such as class separation (Tein, Coxe, & Cham, 2013).

1.3 Practical and Unanswered LCA Questions

Despite the contributions of previous simulation studies, there are still several important areas regarding the application of LCA models that have unanswered questions. For example, researchers often ask "what is the required sample size needed to fit an LCA model?" Unfortunately, there is no single answer to this question since there are many factors that need to be considered when answering this. Previous studies have concluded that an adequate sample size depends on several different factors, including but not limited to, the number of latent classes a researcher can foresee retaining and the number of indicators they wish to include in the model (Lubke & Muthén, 2005; Muthén & Muthén, 2002). Another practical question researchers commonly ask is, "How many indicators can I use in an LCA model?" Again, the answer to this question usually depends on factors such as

[4] LPA can be viewed as the same type of analyses as LCA, with the only difference being the nature of the outcome variables and the type of estimated produced (Nylund et al., 2007a). Specifically, LPA estimates class specific means and variances while LCA estimates class specific item probabilities.

sample size and the number of latent classes a researcher can foresee emerging. In other words, the answer to practical questions related to sample size and the number of indicators in latent class models is confounded and thus there is no single answer. To date, there are some suggested modeling recommendations, however, there are no established best practices related to sample size and the ideal number of indicators to include in an LCA model.

Having established best practices related to sample size and number of indicators is important because the class enumeration[5] process in LCA is known to be highly dependent on sample size and the number of indicators included in the model (Lubke & Neale, 2008; Masyn, 2013). Specifically, sample size and the number of observed indicators have been argued to play a critical role in the detection of what may be a less prevalent class in the population (Masyn, 2013). As sample size and the number of indicators included increases, the number of latent classes a researcher identifies can also potentially increase. This is because smaller samples may be underpowered and therefore less able to detect smaller and/or not well-separated latent classes (Masyn, 2013). Additionally, applied researchers are often left unsure about what final latent class solution to retain because the fit indices may point to many different solutions. This is due to the fact that the fit indices (specifically, the information criteria [ICs], described in more detail in Chapter 2.3) used to help decide on the number of latent classes are also influenced by sample size and the number of indicators included (i.e., the number of parameters estimated) in the model. In fact,

[5] Class enumeration is defined as the process of determining the number of latent classes (e.g. unobserved subgroups) to retain in a latent class model (Nylund et al., 2007a).

the information criteria are calculated by applying "penalties" for the number of model parameters, sample size, or both (Nylund et al., 2007a).

Researchers who have small samples but a large number of indicators often make decisions to delete items so that there is a better balance between the number of indicators, sample size, and the number of classes estimated. The decision on which indicators to remove can be quite varied. Some researchers will decide to remove redundant indicators (either by evaluating the content of the indicators, or using item correlations and deleting redundant items), some may remove indicators they deem unimportant, and some will use modeling results to delete indicators that do not seem to differentiate classes well. Furthermore, while researchers do not know *a priori* how many classes will emerge, they sometimes hypothesize about the number and type of classes that will emerge based on theory and will use that prediction to guide how many indicators to keep. While there are no current recommendations about sample size and number of indicators, there are different strategies that researchers use to make these decisions, none of which are grounded in empirical research.

1.4 The Current Study

Taken together, these practical issues that arise in the application of LCA models are important because they highlight the potential intersection between class enumeration, sample size, and the number of indicators included in these models. To date, there is no published research that directly examines these issues together within the LCA framework. Previous work has begun to establish a link between sample size and the number of latent classes retained in mixture models (Lubke & Neale, 2006; Masyn, 2013; Tein, Coxe, & Cham, 2013); however the current study explores the connection between sample size, the number of indicators, and class enumeration in LCA models.

Given the known link between sample size, the number of observed indicators, and the number of latent classes extracted in factor mixture models (Lubke & Neale, 2008) and latent profile models (Tein, Coxe, & Cham, 2013), and what we currently know about fit index performance in mixture models (Nylund et al., 2007a; Tofighi & Enders, 2007; Yang, 2006), the purpose of this dissertation is twofold. First, this dissertation aims to examine and understand the interplay between sample size, the number of observed indicators, and class enumeration in LCA models. In other words, this study wants to determine if there is a point where, with decreasing sample sizes and potentially limited indicators, the existing heterogeneity in a population can no longer be fully uncovered. Specifically, at what point is specificity of the emerging latent classes lost? Second, while empirically examining the interplay of these aforementioned factors, this study also investigates the performance of the most commonly used measures of model fit in recovering the "correct" latent class solution when factors such as sample size and number of observed indicators are varied. This study also varies other population factors such as the true number of latent classes (either a 3 or 4-class solution), and the size and nature of the latent classes.

The results of this study are important for a number of reasons. First, they help researchers understand how factors such as sample size, and type of data collected can directly influence the classes that emerge in LCA models. Additionally, the results of this study help us to better understand the performance of the various fit indices commonly used to guide the class enumeration process in LCA models. Lastly, based on how the fit indices perform under a range of LCA conditions, this study provides “best practice” recommendations for researchers that currently use LCA models in their work.

1.5 Preview of Dissertation

What follows is a description of the simulation study used in this dissertation. First, previous research findings are summarized and discussed to serve as a foundation and rationale for the empirical conditions that were explored in this Monte Carlo simulation study. Additionally, the literature review (i.e., Chapter 2) discusses LCA models in more detail, as well as the fit indices that are commonly used in the class enumeration process of mixture models. These fit indices include the Akaike Information Criterion (AIC), the Consistent Akaike Information Criterion (CAIC), the Bayesian Information Criterion (BIC), the adjusted Bayesian Information Criterion (ABIC), the adjusted Lo-Mendell-Rubin likelihood ratio test (LMR-LRT), the parametric bootstrapped likelihood ratio test (BLRT), the approximate Bayes Factor (BF), and the correct model probability (cmP).

Second, the methods section (i.e., Chapter 3) presents and explains all data generation and analysis conditions considered in this study. Next, Chapter 4 reports the results of this simulation study in regards to the recovery rates of the various fit indices (i.e., how the fit indices performed when deciding on the correct number of latent classes). The results section also reports findings related to how sample size, the number of indicators, and class enumeration intersect in LCA models. Finally, the discussion (i.e., Chapter 5) recaps and summarizes the findings of this dissertation, provides recommendations for applied researchers, and discusses directions for future Monte Carlo simulation research.

Chapter 2

Literature Review

2.1 Overview of Latent Class Analysis (LCA)

Lazarfeld and Henry (1968) first introduced latent class analysis as a way to relate a single categorical latent variable to a number of observed categorical indicators. The main objective of LCA is to group people into classes based on multivariate response patterns to observed indicators and to identify indicators that best distinguish between classes (Nylund, Asparouhov, & Muthén, 2007). The purpose of LCA is conceptually similar to other traditionally used classification methods (e.g., cluster analysis) however, LCA is a model based approach that operates in a latent variable framework where the underlying class variable is treated as an unobserved, categorical latent variable. The LCA model with observed binary indicators[6], *u*, has an unordered categorical latent variable *c* with *K* classes ($c = k$; $k = 1, 2, 3,\ldots, K$). The *K* classes are exhaustive and mutually exclusive such that each individual in the population has membership in exactly one of the *K* latent classes (Masyn, 2013; see Figure 1). The marginal probability for item $u_j = 1$ is

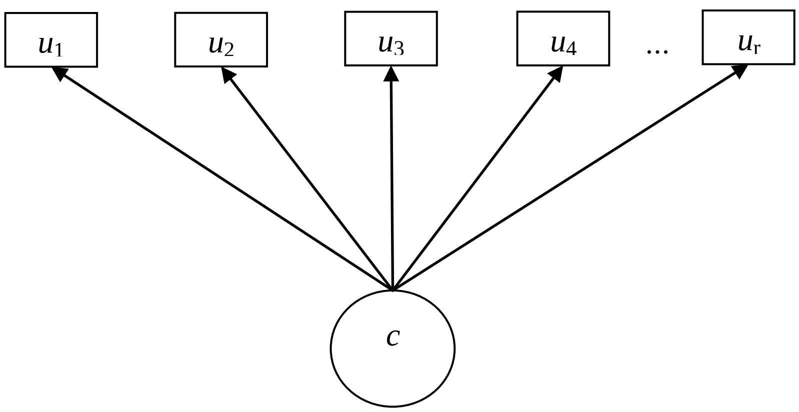
$$P(u_j = 1) = \sum_{k=1}^{K} P(c = k) P(u_j = 1 | c = k).$$

[6] Indicators do not need to be binary. They can be ordinal, multinomial, and/or continuous. A LCA with continuous indicators (i.e., LPA) will not be considered in this simulation study.

The joint probability for all of the observed *u*s (i.e., u_1, u_2,…,u_r), assuming conditional independence[7] is

$$P(u_1, u_2, \dots, u_r) = \sum_{k=1}^{K} P\,(c = k)P(u_1 = 1|c = k)P(u_2 = 1|c = k) \dots P(u_r = 1|c = k).$$

Figure 1. Standard latent class analysis model diagram



2.1.1 Estimated LCA Parameters

There are two types of parameters in the LCA model – class probability parameters and item parameters. *Class probabilities* specify the relative size of each latent class (i.e., how many individuals are in each class), also known as class prevalence. *Item parameters*, in LCA models with categorical outcomes, correspond to the conditional item probabilities for each latent class. Specifically, these parameters are unique to a given latent class and provide information on the probability of an individual in a latent class endorsing a particular item. For example,

[7] The assumption of conditional independence implies that the correlation among the indicators (i.e., u_1, u_2,…,u_r) is completely explained by the latent class variable *c* (Nylund et al., 2007a), thus they are uncorrelated conditioned on class.

a class specific conditional item probability of .90 indicates that 90% of individuals in that given class will endorse that particular item, while only 10% will not (Masyn, 2013). In practice, it is common to graph the conditional item probabilities in an *item probability plot* to get a clearer, more holistic understanding of the patterns that emerge within the data. In fact, the item probability plots often aide in the substantive interpretation of the latent classes that emerge. Some item probability plots are ordered latent class solutions and others are unordered latent class solutions. Specifically, ordered latent class solutions have latent classes that do not cross (see Figures 2 and 5) whereas the unordered latent class solutions do (see Figures 3 and 4). In other words, the latent classes in ordered solutions are differentiated by the degree to which individuals within a given class endorse the indicators. On the other hand, the type of indicators individuals within a class do or do not endorse differentiated latent classes in unordered solutions.

2.2 Assumptions of LCA Models

The overall goal of LCA models is to group or classify similar individuals into one of *K* latent groups (or classes). For this reason, an overarching assumption of these models is the existence of a latent exogenous variable (Heinen, 1996; Tueller & Lubke, 2010). Therefore, if a researcher does not hypothesize that there are underlying subgroups (or latent classes) present in the data, this type of analysis is not justified. LCA models also assume that an individuals' class membership is discrete and mutually exclusive (Clogg, 1995), and that class membership is exhaustive, meaning that the latent classes account for 100% of the individuals in the observed data. Conditional (or local) independence is another fundamental assumption of LCA models. This assumption implies that the underlying latent variable, c, accounts for all relationships between the observed variables (Clogg, 1995; McCutcheon, 1987; Nylund et al., 2007a). In other words, conditional independence implies that there is no remaining relationship between the observed variables after controlling for class membership in the data.

2.3 Fit Indices and Class Enumeration

Deciding on the best fitting model is often the most difficult part of the modeling process. In general, a researcher should consider both substantive theory and statistical fit when making this decision in practice (Muthén, 2003). Specifically, LCA models require the examination of fit indices along with congruence of the modeling results with substantive theory. It is recommended to begin the modeling process by specifying a one-class model and then fitting additional models, increasing the number of classes by one in each model, until the models are no longer well identified (Masyn, 2013). Once this is completed, fit information is collected from each fitted latent class model and aides the researcher in deciding on the statistically best fitting model. This process of deciding on the best fitting latent class model is also referred to as class enumeration. The following fit indices were considered in the current simulation study: Akaike's Information Criterion (AIC), Consistent Akaike's Information Criterion (CAIC), Bayesian Information Criterion (BIC), adjusted Bayesian Information Criterion (ABIC), adjusted Lo-Mendell-Rubin likelihood ratio test (LMR-LRT), the parametric bootstrapped likelihood ratio test

(BLRT), the approximate Bayes Factor (BF), and the correct model probability (cmP). The goal of examining these commonly used fit indices was to understand how well they perform in enumerating the correct latent class model.

2.3.1 Information Criteria (ICs)

Information Criteria (ICs) are fit indices that are commonly examined across a wide range of statistical models and are used to compare a set of models. The ICs take model complexity into account and are also used to evaluate statistical fit. These indices include the Akaike Information Criterion (AIC; Akaike, 1987), the Consistent Akaike Information Criterion (CAIC; Bozdogan, 1987), the Bayesian Information Criterion (BIC; Schwarz, 1978), and the adjusted Bayesian Information Criterion (ABIC), where lower values indicate a better fitting model. The AIC can be defined as:

$$\text{AIC} = -2(\text{log-likelihood}) + 2p,$$

where p is the number of free model parameters. The CAIC is a derivative of the AIC however it also penalizes the value of -2 times the log-likelihood of the model for the number of free model parameters and sample size (Bozdogan, 1987). The CAIC is defined as:

$$\text{CAIC} = -2(\text{log-likelihood}) + p\,[\log(n) + 1],$$

where p is the number of free parameters and n is the sample size. The BIC also includes an adjustment for the sample size and is defined as:

$$\text{BIC} = -2(\text{log-likelihood}) + p\log(n),$$

where p is the number of free parameters and n is the sample size. Lastly, the ABIC is a derivative of the BIC that reduces the penalty associated with sample size. The ABIC is defined as:

$$\text{ABIC} = -2(\text{log-likelihood}) + p\log[(n+2)/24],$$

where again p is the number of free parameters and n is the sample size.

Many simulation studies support the BIC as being the IC that consistently identifies the correct number of classes for mixture models (Jedidi, Jagpal, & DeSarbo, 1997; Magidson & Vermunt, 2004; Peugh & Fan, 2013; Roeder & Wasserman, 1997; Tein, Coxe, & Cham, 2013; Tueller & Lubke, 2010). In fact, a previous simulation study considered all of the aforementioned ICs and found the BIC to perform the best across various mixture models (Nylund et al., 2007a). A more recent simulation study considered latent class, latent profile, and factor mixture models and further confirmed these results. Specifically, findings revealed that the BIC tended to identify the correct solution with higher frequency than other indices, especially in models with more continuous than categorical indicators, or when rare classes were not present (Morgan, 2012). Other simulation studies have found strong evidence for the ABIC (Peugh & Fan, 2013; Tein, Coxe, & Cham, 2013; Tofighi & Enders, 2007), even in instances where the sample size was relatively small (Yang, 2006).

Lastly, there is a consensus in regards to the AIC overestimating the number of classes in mixture models (Celeux & Soromenho, 1996; Koehler & Murphree, 1988; Nylund et al., 2007a; Tein, Coxe, & Cham, 2013). Specifically, research has shown that the AIC overestimates the number of latent classes with larger sample sizes (Woodroofe, 1982). In fact, Nylund, Asparouhov, and Muthén (2007) found that the AIC accuracy decreased as sample size increased and suggested this is due to the fact that the AIC includes no adjustment for sample size. The CAIC however, has been shown to perform well across multiple conditions (Peugh & Fan, 2013), especially when the sample size is relatively large (i.e., n = 1000; Nylund et al., 2007a). This is likely due to the CAIC's adjustment for the number of parameters using the sample size, but more studies are needed to fully understand the range of use of the CAIC.

2.3.2 Likelihood Ratio Tests: Adjusted LMR-LRT and BLRT

The adjusted Lo-Mendell-Rubin likelihood ratio test (adjusted LMR-LRT; Lo, Mendell, & Rubin, 2001) and parametric bootstrapped likelihood ratio test (BLRT) are commonly used to compare nested models and are implemented within Mplus. These tests compare the K-1 class model (the null model) with the K class model. In other words, the null hypothesis for the adjusted LMR-LRT and BLRT states that the number of classes is equal to k-1 (H_0: $K = k-1$), and the alternative hypotheses states that the number of classes is equal to k (H_1: $K = k$) (Morgan, 2012). Therefore, statistically significant p-values suggest that the K class model fits the data significantly better than the K-1 class model (Masyn, 2013).

A previous simulation study examined the performance of these fit indices among others for 3 types of mixture models: LCA models, factor mixture models (FMM), and growth mixture models (GMM; Nylund et al., 2007a). Findings from this study showed the BLRT to be a very consistent indicator of classes across all of the models considered (Nylund et al., 2007a). In fact, studies show that the BLRT often outperforms the adjusted LMR-LRT (Nylund et al., 2007a; Tein, Coxe, & Cham, 2013). Other simulation studies however, have found strong evidence for the adjusted LMR-LRT (Lo, Mendell, & Rubin, 2001; Lubke & Muthén, 2007; Tofighi & Enders, 2007). Specifically, Tofighi & Enders (2007) examined a series of GMM analyses and concluded that the LMR-LRT was a relatively consistent indicator of the correct number of latent classes, however, this study did not consider the BLRT as well. Additionally, Lubke and Muthén (2007) explored a series of LPA models and found that the adjusted LMR-LRT performed extremely well in conditions where the latent classes were well separated.

2.3.3 Bayesian Fit Indices: BF and cmP

The approximate Bayes Factor (BF) and the approximate correct model probability (cmP) are two fit indices commonly used in the Bayesian framework that have more recently been suggested to be promising for mixture modeling (Masyn, 2013). The BF is a pair-wise comparison of relative fit between two competing models, Model A and Model B (Masyn, 2013). Specifically, Model B is the smaller model, nested in Model A. In practice, the BF is calculated by using the following equation:

$$\hat{BF}_{A,B} = \exp\left[SIC_A - SIC_B\right],$$

where SIC is the Schwarz Information Criterion (Schwarz, 1978), which is equal to -0.5BIC (Masyn, 2013). A BF greater than 1 and less than 3 is weak evidence for

Model A, greater than 3 and less than 10 is moderate evidence for Model A, and greater than 10 is strong evidence for Model A (Wasserman, 2000).

The approximate correct model probability (cmP) on the other hand, allows a researcher to compare a set of more than two latent class models. This statistic is calculated once all of the latent class models are fit and generally outside of the commonly used statistical software packages. Specifically, there is a cmP value for each of the latent class models. Model A (A=1, …, J) is calculated by using the following equation:

$$cm\hat{P}_A = \frac{\exp(SIC_A - SIC_{\max})}{\sum_{j=1}^{J}\exp(SIC_j - SIC_{\max})},$$

where SIC_{max} is the maximum SIC score of all the J models being considered (Masyn, 2013). If the sum of the cmP values across a set of models is equal to 1, then the "true" model is assumed to be one of the models in the set being compared (Masyn, 2013). No existing research to date has examined the performance of the BF and cmP in the class enumeration process with LCA models.

In practice, researchers using the set of fit indices described above to fit a LCA model will often end up with the fit indices indicating a few competing models. Thus, it has been recommended that researchers should use these indices in concert with substantive theory to decide on the final model to retain (Masyn, 2013; Muthén, 2003).

Item probability plots are also often used to help a researcher decide on the best substantively fitting model. These plots graphically show the various latent classes that emerge and can help a researcher understand how classes differ in terms of the patterns they exhibit. Additionally, item probability plots help researchers understand which indicators are most useful in producing meaningful latent classes. Specifically, "good" indicators should have both high within class homogeneity and high between-class separation (Masyn, 2013; discussed in more detail in Chapter 3.4). Lastly, model parsimony should also be considered while deciding on the statistically and substantively best fitting model. In general, the model with the fewest number of classes that fits the data both statistically and substantively well is favored (Masyn, 2013).

2.4 Previous Findings and Future Directions

Existing research thus far has highlighted the performance of these various fit indices in deciding on the statistically best fitting model, but has indicated some sensitivity to sample size. In general, research has shown that top performers include the BLRT (Nylund et al., 2007a), BIC (Jedidi, Jagpal, & DeSarbo, 1997; Magidson & Vermunt, 2004; Nylund et al., 2007a; Peugh & Fan, 2013; Roeder & Wasserman, 1997), and adjusted BIC fit indices (Peugh & Fan, 2013; Tofighi & Enders, 2007; Yang, 2006). Yang (2006) noted that in general, small sample sizes tend to cause instability in the performance of the ICs. That is, accurately deciding on the correct number of latent classes increases as sample size increases (Yang, 2006). Morgan (2012) concluded that the BIC identified the correct solution in models with more continuous than categorical indicators, or when rare classes were not present. Additionally, this study noted that the AIC tended to identify the correct class solution with higher frequency than the other indices when there was a small degree of separation between latent classes (Morgan, 2012).

Taken together, this research suggests a connection between sample size, the nature of the latent classes, the metric level of the observed indicators, and the final number of latent classes extracted. In fact, Lubke and Neale (2006) conducted a simulation study that revealed that retaining the correct number of latent classes is dependent on within-class sample size and between-class separation. This study however, did not consider the number of indicators included in the model. Furthermore, Lubke and Neale (2008) conducted a simulation study using factor mixture modeling (FMM) concluding the number of latent classes extracted was heavily dependent on sample size, and the number of indicators included in the model. Similarly, another simulation study examined a special case of FMM's labeled structural equation mixture models (SEMM), which includes regressions between latent factors within each latent class. Results indicated that sample size has a substantial effect on model performance, especially in cases where class separation is low (Tueller & Lubke, 2010). Lastly, a recent simulation study found that power to detect the true latent class solution in LPA models depended on the number of indicators included in the model, distance between the latent classes, and sample size (Tein, Coxe, & Cham, 2013). This study however, did not consider binary items, sample sizes greater than 1,000, and did not take the nature of the latent classes into account (i.e., ordered vs. unordered classes).

This interplay between sample size, number of indicators, and number of latent classes retained has been discussed to be true for latent class models as well (Masyn, 2013). One reason could be because small sample sizes may be underpowered to detect smaller latent classes (Masyn, 2013; Muthén & Muthén, 2002). Further, this interplay is crucially important in applied research because often times, in practice, researchers do not have access to large datasets. Social science researchers instead often collect data on a small sample of the larger population. Consequently, it is critical to understand how subsetting observations from a larger population influences fit index performance and thus the latent classes that emerge. Similarly, there is a practical need to understand how reducing the number of indicators (either because they do not help distinguish classes or because they are redundant) influences fit index performance and subsequently class enumeration in general.

Therefore, this study aims to empirically understand the extent to which sample size, number of observed indicators, and number of latent classes extracted intersect in LCA models. In other words, when sampling observations from a larger population, is there a critical point where the size of the sample and the number of indicators cannot uncover all of the existing heterogeneity? That is, at what point is specificity of the emerging latent classes lost? Additionally, this study aims to further understand the ability of the commonly used fit indices to uncover the correct latent class solution. The answers to these questions have important implications for researchers using LCA models in their research. Specifically, results will allow researchers to understand how sample size, and the number of observed indicators (among other factors) influences the class enumeration process. Results also highlight the importance of thoughtfully considering sample size and the number of indicators included when estimating and interpreting LCA models. Lastly, results from this study will help researchers using LCA models further understand which fit index to trust under various conditions when going through the class enumeration process in practice.

Chapter 3

Method

3.1 Overview of Current Monte Carlo Study

The purpose of this dissertation was to examine the performance of the most commonly used fit indices in selecting the “correct” latent class model while varying factors such as: the true number of latent classes, the size of the latent classes (i.e., class prevalence), the nature of the latent classes, the number of indicators, and sample size. Secondly, this study aimed to examine and understand the intersection of sample size, the number of observed indicators, and class enumeration in LCA models. The following methodological steps were taken in this Monte Carlo simulation study, and will be discussed in more detail below. First, a total of eight data conditions (each with 500 replications) were generated within Mplus Version 7.1 (Muthén & Muthén, 1998-2013), based on a known number of latent classes and a known set of population parameters. Next, this data was read into Mplus and analyzed with a different sample size, and number of indicators than the generated data. Modeling results were then tabulated to determine recovery rates for each of the fit indices considered. Specifically, for the purposes of this study, a recovery rate is defined as the percentage for which the various fit indices were able to correctly identify the “true” latent class solution. Finally, these tabulated results were further examined to understand the extent to which sample size, number of indicators, and class enumeration intersect in LCA models.

3.2 Creating Empirical Conditions: Meta-Analysis

The conditions for the current simulation study were created based on a review of recent (within the last 5 years) empirical research articles that used LCA. Specifically, articles were selected by searching the *Elton B Stephens Company (EBSCO)*, *Education Full Text*, *Education Resources Information Center (ERIC)*, *PsycINFO*, and *PsycARTICLES* databases. Keywords used to find research studies were originally solely "latent class analysis." However, these keywords alone yielded over 3,500 potential research articles. Therefore, keywords used to find relevant studies were expanded to include: latent class analysis, LCA, binary, and finite mixture models. The search was also limited to recent peer-reviewed, full text research articles available through the University of California, Santa Barbara library that were within the years of 2009 and 2013. It is important to mention that searching large databases does not ensure that all relevant studies will be found.

A total of twenty studies[8] were selected that used LCA between 2009 and 2013 in peer-reviewed education and psychology journals. Ten of the 20 articles (50%) using LCA reported identifying four-class solutions, followed by 6 out of 20 (30%) identifying three-class solutions, and only 4 out of 20 (20%) identifying five or more latent classes. Of these 20 latent class solutions, 10 out of 20 (50%) were unordered classes, 5 out of 20 (25%) were ordered classes, and another 5 out of 20

[8] The database search using the aforementioned keywords yielded a total of 24 potential articles. Of these 24 articles, four articles did not meet the predetermined criteria for inclusion. Specifically, some studies included continuous indicators and did not use latent class analysis procedures. Therefore, 20 final studies were selected.

(25%) did not provide enough information to decide (i.e., did not include item probabilities or item probability plots).

In terms of the most commonly reported fit indices and procedures used to decide on the best fitting latent class solution, 16 out of 20 articles (80%) reported using the BIC, 10 out of 20 (50%) reported using substantive theory, 9 out of 20 (45%) reported using the AIC, 5 out of 20 (25%) reported using the adjusted LMR-LRT, 3 out of 20 (15%) reported using the ABIC, and 2 out of 20 (10%) reported using the CAIC[9]. Sample sizes ranged from approximately 150 participants to approximately 20,000 participants (large-scale, nationally representative data). A Pearson product moment correlation was used to explore a potential relationship between sample size and the number of latent classes extracted in these 20 studies. Results indicated that the relationship between sample size and latent classes was not statistically significant (r = -.02, n = 20, p = .948). Lastly, these studies reported using anywhere from 2 to 33 indicators of class membership. It is important to note that these values for the number of indicators were outliers and the majority of the studies (90%) reported using between 5 and 10 indicators of class membership.

3.3 Data Generation

3.3.1 Number of Latent Classes and Class Prevalence

All generated data included a total of 10,000 observations and 10 binary indicators. The true number of latent classes was either a three or four class model. The decision to generate data with a true three-class and four-class solution was based

[9] The usage of these fit indices is not mutually exclusive. That is, 90% of the articles reported using more than one of the aforementioned fit indices when deciding on the best fitting latent class model.

on the previously mentioned review of recent empirical research. In fact, 80% of the empirical studies retrieved in the meta-analysis identified either a three or four-class solution. The relative size of the latent classes (i.e., class prevalence) were varied to be either equal (π_1 = .33, π_2 = .33, π_3 = .33 for the three-class solution and π_1 = .25, π_2 = .25, π_3 = .25, π_4 = .25 for the four-class solution), which was modeled from a simulation study by Nylund, Asparouhov, & Muthén (2007), or unequal in nature (π_1 = .10, π_2 = .30, π_3 = .60 for the three-class solution and π_1 = .10, π_2 = .20, π_3 = .20, π_4 = .50 for the four-class solution). The unequal sized three-class condition aimed to represent one large/normative class, one moderate class, and one relatively small latent class, a configuration that is commonly seen in applied research (Shin, Hong, & Hazen, 2010; Stormont, Herman, Reinke, David, & Goel, 2013; Von Stumm, Chung, & Furnham, 2011). The unequal sized four-class condition was created to represent one large/normative latent class, one small latent class, and two relatively moderate classes, a configuration that has also been seen in previous applications (Bettencourt & Farrell, 2013; Cavanaugh et al., 2012; Grant et al., 2006; Shin, Hong, & Hazen, 2010). Previous simulation studies have also generated data with unequal class sizes but have done so in slightly different ways. For example, Morgan (2012) generated data with both three and four unequal class sizes however, the specific parameter values were marginally different (π_1 = .59, π_2 = .26, π_3 = .15 and π_1 = .59, π_2 = .19, π_3 = .13, π_4 = .09, respectively). Nylund, Asparouhov, & Muthén (2007) also generated data with four unequal class sizes however, there was only one large/normative class and the remaining three classes were relatively small ($\pi_1 = .75$, $\pi_2 = .05$, $\pi_3 = .10$, $\pi_4 = .15$).

3.3.2 Nature of the Latent Classes

The four population latent classes were varied to be either ordered or unordered in nature (see Figures 2-5 below). Figures 2 and 5 display a graphical representation of what the four and three ordered class conditions were generated to look like, respectively. Specifically, the classes in the ordered latent class solutions were generated to not overlap and were solely differentiated by the degree to which a given class endorsed the various indicators. There were also two types of unordered latent class conditions (see Figures 3 and 4 below). The first one (unordered latent class model A) was modeled after latent classes that were well separated by multiple indicators (see Figure 3 below) whereas the second set (unordered latent class model B) of unordered profiles were more overlapping, and not as well separated (see Figure 4 below). The four classes in these unordered latent class solutions were generated to overlap and were differentiated by the type of indicators endorsed or the various class specific conditional item probabilities.

After crossing all factors that were varied in this study, a total of eight generated data conditions were created (see Table 1 and Figure 6 below). All of the conditions with four classes were fully crossed, yielding the following six four-class conditions: four equal sized ordered classes, four unequal sized ordered classes, four equal sized unordered classes A, four unequal sized unordered classes A, four equal sized unordered classes B, and four unequal sized unordered classes B (see Figures 2-

4 below for graphical representations of the unequal class size conditions). Alternatively, the three-class models were not fully crossed. Specifically, there were only two conditions with true three-class solutions. One condition included three ordered and equal sized latent classes, and the other condition included three ordered and unequal sized latent classes (see Figure 5 below for an item probability plot of the unequal class size condition).

Generated data were based on the above mentioned population specifications, which allowed the current study to investigate the performance of fit indices under different modeling conditions. Data were randomly generated based on variations of these abovementioned structures using the Monte Carlo functions in Mplus (See appendices A-D for sample data generation input files). Each of the eight generated datasets of 10,000 observations included 500 replications, which has been demonstrated to be a sufficient number of replications in previous simulation research (Morgan, 2010; Nylund et al., 2007a). This resulted in a total of 4,000 generated datasets (8 conditions * 500 replications each = 4,000).

Figure 2. Item probability plot for four ordered latent classes

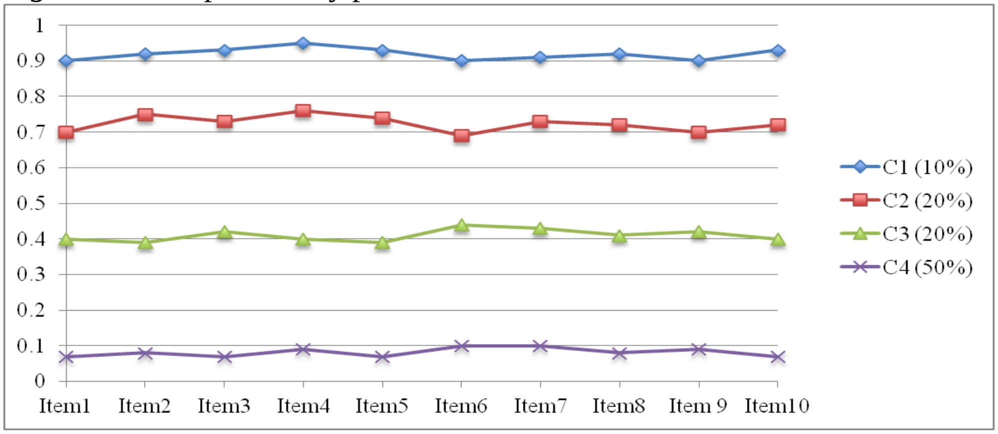


Figure 3. Item probability plot for four unordered latent classes A

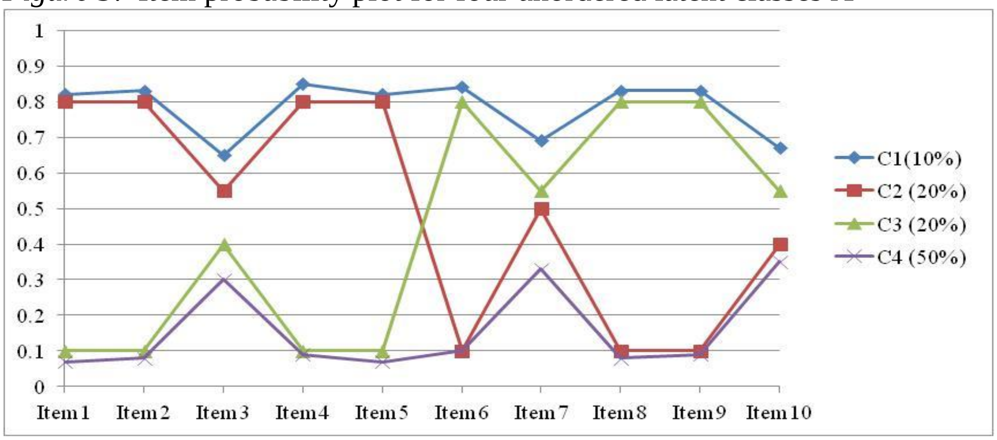


Figure 4. Item probability plot for four unordered latent classes B

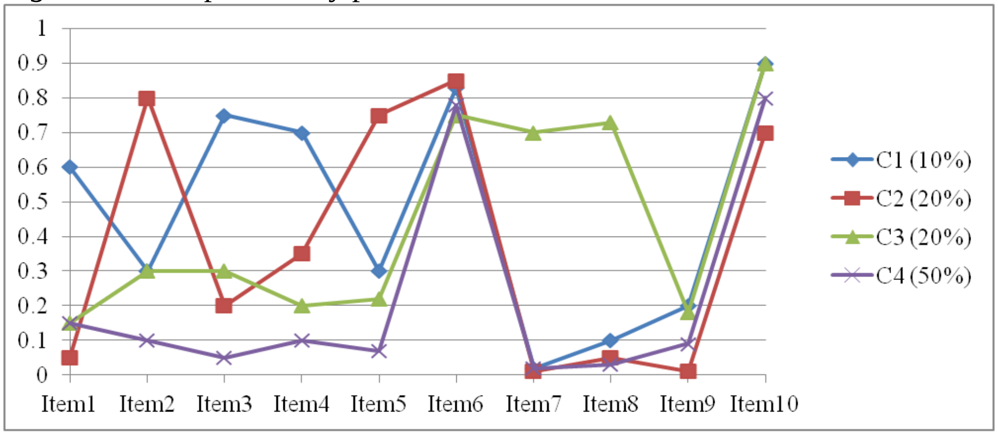


Figure 5. Item probability plot for three ordered latent classes

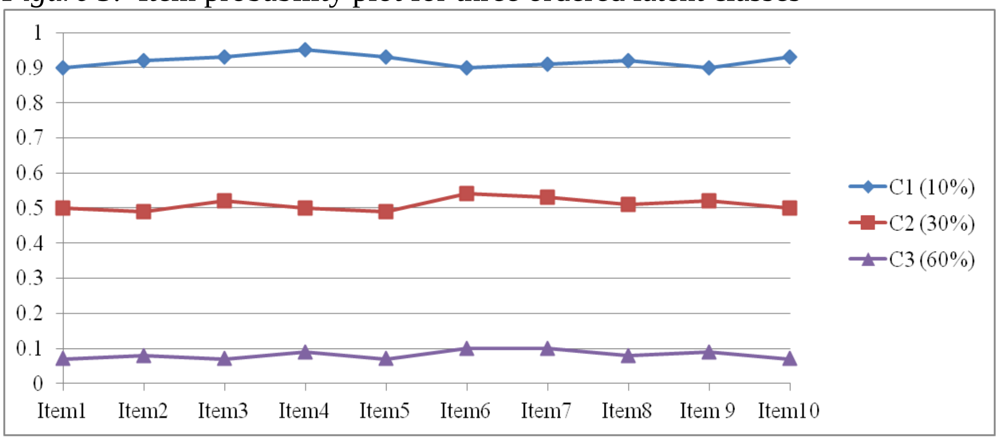


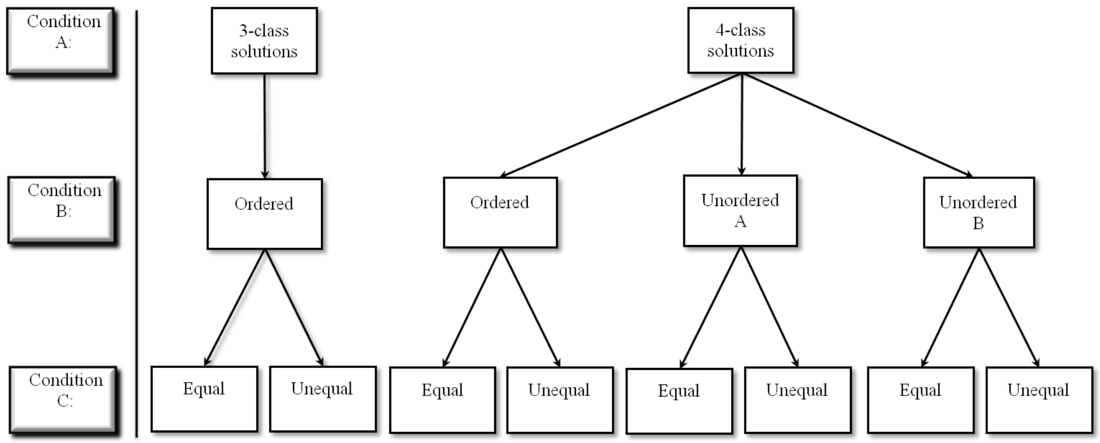
Table 1

Data Generation Conditions

	A	B	C
	Number of Latent Classes	Nature of Latent Classes	Class Size
Number of Levels	2	3	2
Details	Three-class solutions, Four-class solutions*	Ordered, Unordered A, Unordered B	*Equal class sizes*: 33.3% each for three-class solutions and 25% each for four-class solutions *Unequal class sizes*: large/normative 60%, 30%, and small 10% for three-class solutions; large/normative class 50%, 20%, 20%, and small 10% for four-class solutions

**Note*. All conditions with four classes were fully crossed; however, conditions with three classes were not. The eight generated data conditions were as follows: four equal sized ordered classes, four unequal sized ordered classes, four equal sized unordered classes A, four unequal sized unordered classes A, four equal sized unordered classes B, four unequal sized unordered classes B, three equal sized ordered classes, and three unequal sized ordered classes (see Figure 6 below).

Figure 6. Data Generation Conditions Simulation Map



3.4 Data Analysis

After all 4,000 datasets were generated; they were subsequently read in as data and analyzed in Mplus. All data were generated to have an original population sample size of 10,000 and to have latent classes that were created based on precise item probabilities across 10 binary indicators. The number of indicators and sample size were later systematically varied across analyses conditions explored in this simulation study.

3.4.1 Number of Indicators

The numbers of indicators were varied in each of the analysis conditions; specifically, for each condition there were models with either 7 or 10 indicators. The decision rule for eliminating indicators (i.e., in conditions where there were only 7 indicators) was to eliminate 3 items that were purposely generated to be “poor” latent class indicators. Specifically, “poor” indicators were defined as having low class homogeneity and/or low class separation (i.e., did not separate the latent classes well)

(Masyn, 2013). Indicators with low class homogeneity had average item probabilities ranging between approximately .3 and .7 (e.g., items 3, 7, and 10 in Figure 3), meaning that participants in that given class responded potentially differently to that particular indicator (Masyn, 2013). In other words, indicators that had low class homogeneity did not provide a good representation of the latent class since item endorsement was not "typical" or a "characteristic of" that latent class (Masyn, 2013, p.559). Alternatively, indicators that had poor/low class separation did not help distinguish or differentiate the various latent classes (e.g., items 6, 9, and 10 in Figure 4). In other words, the conditional item probabilities overlapped across classes and there was close to no between-class separation among these items. This process of eliminating items that have low class homogeneity and low class separation is also frequently done in practice with applied data (Collins & Lanza, 2010; Masyn, 2013). Given that the indicators in the ordered three and four-class solutions were designed to have relatively the same item endorsement probabilities within class (see Figures 2 and 5), items 1, 5, and 10 were eliminated in conditions where the ordered latent classes were defined by only 7 indicators. Eliminating items in the ordered latent class conditions aimed to represent removing potentially redundant latent class indicators. In general, varying the number of indicators that were defining the latent classes to be either 7 or 10 indicators allowed the current study to understand how removing "poor" or potentially redundant indicators, which is often times done in practice (especially with small samples in the interest of parsimony), influenced the class enumeration process.

3.4.2 Sample Size

Sample size was also varied in the analyses conditions. Sample sizes considered in this study were partially mirrored from a previous simulation study by Nylund et al. (2007a) and were also informed through the previously mentioned literature search. Specifically, sample sizes examined in this dissertation were 200, 500, 1,000, 5,000 and 10,000. Given that the generated data included 10,000 observations, a random subset of 200, 500, 1,000, and 5,000 observations were selected when analyzing the data to assess the effect of sample size on the ability of the fit indices to uncover all existing heterogeneity in the overall population. These subset sample sizes reflected 2%, 5%, 10%, and 50% of the overall population, respectively. This approach of subsetting the population sample size when analyzing the generated data has not been previously done in published research. All existing simulation studies that have examined the effects of sample size on class enumeration have generated populations that vary in their "true" population sample size.

3.5 Data Analytic Procedures

3.5.1 Subsetting Sample Size and Number of Indicators

The 4,000 generated datasets (8 conditions * 500 replications each = 4,000) were read in as data and analyzed in Mplus. Though all data were generated to be N = 10,000, sample sizes were varied when analyzing the generated data to reflect a random subset of 200, 500, 1,000, and 5,000 observations from the original data set. These sample sizes were purposely chosen to represent 2%, 5%, 10%, and 50% of the overall population, respectively. Specifically, an identification (ID) variable was generated in all 4,000 datasets using Stata Version 13 (StataCorp, 2013). This was done in Stata because Mplus does not currently include an option to generate an ID variable within the Monte Carlo context. The data was then imported back into Mplus and using the newly generated ID variable, a random subset of 200, 500, 1,000 and 5,000 observations were pulled from the original 10,000 observations via the "useobservations" command. Additionally, the number of indicators included in the model (i.e., either 7 or 10) was varied using the "usevariables" command. The rationale behind subsetting sample size and the number of indicators when analyzing the data was to understand how different sample sizes (achieved by subsetting observations from a larger population) and numbers of indicators influenced the class enumeration process. That is, the current study wanted to understand the point in which decreasing sample sizes and potentially limited indicators were not able to uncover all of the true heterogeneity (i.e., latent classes) in the population.

Given that applied researchers do not typically know the true number of latent classes when estimating an LCA model, a series of latent class solutions were explored in this Monte Carlo study. Specifically, 1-5 latent class models were fit to the data, per each condition, using the MIXTURE option in Mplus 7.1. Additionally, as mentioned above, latent class models with differing sample sizes and number of indicators were fit to the data, for each of the eight previously mentioned conditions. This process produced a total of 400 analysis conditions, which corresponded to 400 results files (i.e., 8 (data conditions) * 5 (latent classes) * 5 (sample sizes) * 2 (number of indicators) = 400 analyses; see Table 2). Next, these Mplus generated results files that included all model parameter estimates and fit index information for each replication of the analyses, were saved. These results files were then combined for a given condition, merging the fit information for the 1-5 latent classes. Specifically, the results files for each condition were merged and summarized separately[10]. This allowed the identification of how many latent classes each fit index identified, per condition. Sample analysis Mplus input files are provided in Appendices E-G.

Table 2
Data Analysis Conditions

	D	E	F
	Sample size, *n*	# of latent classes	# of indicators
Number of Conditions	5	5	2
Details	200,500,1,000, 5,000, 10,000	1-5 classes	7 indicators, 10 indicators

3.6 Expectation Maximization (EM) Algorithm

When analyzing the data, Mplus used the expectation-maximization (EM) algorithm to obtain maximum likelihood estimates of LCA parameters (Masyn, 2013). The first step in obtaining these maximum likelihood parameter estimates is to specify the likelihood function. Specifically, Masyn (2013) described this likelihood function as "the probability density of all of the data (the array of all values on all variables, latent and observed, in the model for all individuals in the sample) given a set of parameter values" (p. 561). A maximum likelihood solution consists of

[10] Merging files across replications and making comparisons across conditions required fixing the random seed value that Mplus uses to be equal across all data generation models. Specifically, this was done by using the "seed =" command and specifying the same unique value in all data generation input files (see Appendices A-D).

parameter values that maximize the likelihood function. Obtaining these estimates is done through an iterative process. The goal of this iterative process is to reach what is known as a *global* solution (or set of parameter values), however, mixture models are known to be susceptible to converging on *local* solutions (Nylund et al., 2007a). A global solution is attained when the peak (or the maximum) of the likelihood function is reached, whereas a local reaches a peak but it is not the highest of the entire likelihood function.

Reaching a local, and not global solution is problematic and concerning because the maximum likelihood estimates you obtain can be unstable and not trustworthy (Masyn, 2013). There are however some practical suggestions that can be employed to increase confidence that the estimation process converges on a global solution. First, the use of multiple sets of random start values for the model was used to help ensure finding a global solution (Muthén & Muthén, 1998-2011; Nylund et al., 2007a; Masyn, 2013). To help facilitate this, Mplus includes a random start value feature that generates different random start sets **(**Nylund et al., 2007a). Next, the log likelihood value estimates in the model output were examined to make sure that they were replicated. Mplus orders the log likelihood values from the random starts from highest to lowest in the output file. If the best log likelihood value is not replicated at least twice (preferably more), is it possible that a local, not global solution has been reached **(**Muthén & Muthén, 1998-2011). Lastly, a good set of model start values can improve the chances of locating the maximum of the likelihood function as opposed to the top of one of the smaller peaks (i.e., local maximum). The default random start setting was overridden in Mplus to increase the chances of locating a global solution. Specifically, the number of random starts for all LCA models in this simulation study was specified as “Starts=100 20.” The first number in this command refers to the number of initial iterations and the second number refers to the number of final iterations. In other words, the number of random sets of starting values was equal to 100 and the 20 best likelihood values obtained were used as starting values for an optimization that continued until the models converged. Thus, there was little chance that the results of this simulation study were based on local solutions. Additionally, to avoid warnings about log likelihoods not being replicated in the bootstrap draws for the BLRT fit index, the default start settings for the LRTSTARTS option were also overridden (i.e., LRTSTARTS = 0 0 150 40;).

Chapter 4

Results

4.1 General Overview

This chapter provides the results of the simulation study. The purpose of this simulation study was twofold. First, this study examined the performance of the most commonly used fit indices in selecting the “correct” latent class model while varying factors such as: the true number of latent classes, the size of the latent classes (i.e., class prevalence), the nature of the latent classes, the number of indicators, and sample size. Second, this dissertation explored a potential intersection between sample size, the number of observed indicators, and class enumeration in LCA models. Therefore, results of this study include detailed information about the performance of the ICs (i.e., AIC, CAIC, BIC, ABIC), the likelihood ratio tests (i.e., the adjusted LMR-LRT and BLRT), and the Bayesian fit indices (i.e., the BF and cmP) in identifying the LCA model with the correct number of classes. Additionally, this chapter includes findings related to the degree to which sample size, the number of indicators, and class enumeration intersect in LCA models.

4.2 Model Fitting and Checking

4.2.1 Model Convergence

The Monte Carlo capabilities in Mplus Version 7.1 (Muthén & Muthén, 1998-2013) were used to both generate and analyze data. Summary information was requested and provided across all completed replications. *P*-values for the adjusted LMR-LRT and BLRT were provided for each replication that converged by specifying TECH11 and TECH14 as an output option. A previous study stated that "nonconvergence of any given replication may occur because of singularity of the information matrix or an inadmissible solution that was approached as a result of negative variances" (Nylund, Asparouhov, & Muthén, 2007, p. 551). Moreover, non-convergence is often considered an indication of model misfit, and used as evidence that the model with one fewer classes is superior (Nylund et al., 2007a). However, one hundred percent of the replications converged across all conditions in this simulation study. This is not surprising since in general, nonconvergence rates in previous simulation studies were documented to be extremely low (i.e., less than 1% of the time) with unconditional LCA models (Nylund et al., 2007a). Additionally, this high convergence rate may be attributed to the fact that the latent classes in this simulation study were well defined, the specified indicator and population characteristics were relatively simple in nature (i.e., only binary indicators and unconditional LCA models were considered), and the default number of random starts within Mplus were overridden and increased during the data analysis stage.

4.2.2 Coverage

Coverage estimates were examined for each estimated parameter across all 500 replications, per condition in this simulation study. C*overage estimates* are summary statistics that convey the ability of the analysis models to accurately recover the specified true population parameter values (i.e., the parameters that the data was generated based on). These estimates are summarized by looking at average parameter values across 500 replications with 95% confidence intervals that contain the true population parameter values (Nylund et al., 2007a). For example, when data from a three-class LCA model are generated and analyzed with a three-class LCA model, it is expected that the estimated average parameter values (across 500 replications) will be close to the true population parameter values that were used to generate the data. If the estimated average parameter values in a correctly specified model are not close to the true population parameter values that were used to generate the data, the results of the simulation study have little meaning (Nylund et al., 2007a). For example, a coverage value of .94 for a given parameter would indicate that, across all 500 replications, 94% of the model estimates fall within a 95% confidence interval of the true population parameter value (Nylund et al., 2007a). Coverage estimates for each of the estimated model parameters across all conditions of this simulation study fell within the recommended range of .91 and .98[11] (Muthén & Muthén, 2002). These high coverage estimates are important because they indicate that when specified correctly, the analysis models are able to accurately recover the true population values.

4.3 Overview of Simulation Results

Tables 3-6 summarize the results of the current simulation study. Specifically, the values in Tables 3-6 represent percentage recovery rates for all fit indices considered across all possible conditions. In other words, the values indicated the percentage of times (out of 500 replications, per condition) that the fit indices

[11] It is important to note that when coverage is studied in the LCA context, the random starts option within Mplus is not used (i.e., starts=0). If random starts are not turned off, “label switching” of the latent classes may occur across each replication and distort coverage estimates.

recovered each latent class solution across the various LCA conditions. If all fit indices were able to correctly identify the true three or four-class solution, the percentages in columns identifying four latent classes in Tables 3-5 should result in the highest values, and the percentages in columns identifying three latent classes in Table 6 should result in the highest values. For example, when looking at the first row in Table 4, the AIC recovered the correct 4-class solution 82% of the time when sample size was 200, classes were unordered and equal, and when only 7 items were included in the estimation process. Figures 7-36 display various graphical representations of the recovery rates across all fit indices considered, both per condition and across conditions. For the purposes of this simulation study, a recovery rate of 95% or higher was considered an indication of high fit index performance. A 95% threshold value has also been used in previous simulation studies that examined fit index performance (Lubke & Neale, 2006; Nylund et al., 2007a). Therefore, to visually depict the level at which the fit indices should be performing to be considered "high performing" (i.e., 95% or higher), a red dotted line was included in Figures 7-36.

4.4 Four Ordered Latent Class Model

The recovery rates across all conditions where data was generated to have a total of four ordered latent classes (see Figure 2) are presented in Table 3. If all fit indices were able to correctly identify the true four-class solution, the percentages in columns identifying four latent classes in Table 3 should result in the highest values. However, in general, all fit indices did not perform well under this specific condition (see Table 3). In fact, across all fit indices considered, the true number of latent classes was vastly underestimated and in general, the percentages in the columns identifying three latent classes resulted in the highest values. A possible explanation for these findings is that class 1 and class 2 in Figure 2 did not have good separation, meaning that the distance between those two latent classes was relatively small, which made it difficult to detect the true latent class solution. Based upon these results, it appears that the results of this condition are not accurate and therefore will not be discussed further in detail since they do not represent the true ability of the fit indices considered.

It is interesting to note however that a recent LCA simulation study concluded that the AIC was able to correctly identify the correct latent class solution at higher rates than the other fit indices even when between-class separation was small (Morgan, 2012). The results in Table 3 below seem to indicate a similar trend with the AIC even though in general, recovery rates were extremely low. The extremely low recovery rates for this condition prompted the current study to consider another ordered latent class solution that was generated to have greater between class separation and one fewer latent class (i.e., the ordered three class solution). Results from this condition will be discussed in further detail in section 4.7.

4.5 Four Unordered Latent Class Model A

The recovery rates across all four unordered latent class A conditions (see Figure 3) are presented in Table 4. If all fit indices were able to correctly identify the true four-class solution, the percentages in columns identifying four latent classes in

Table 4 should result in the highest values. Figures 7-10 display a graphical representation of recovery rates for all fit indices considered across all four unordered latent class A conditions.

4.5.1 Information Criteria (ICs)

The AIC performed the worst among the ICs, with recovery rates well below the 95% benchmark across all unordered A conditions considered. Specifically, results indicated that when the AIC failed (i.e., does not identify the correct latent class solution), there was a tendency to overestimate the correct number of latent classes (see Table 4). The results in Table 4 indicate that the AIC index has a tendency to perform better as sample size decreases. The CAIC, BIC, and ABIC, all performed extremely well across all conditions however, in general, performance seemed to decrease as sample sized decreased (see Figures 11-16). When sample size was $n = 200$, class sizes were equal, and "poor" items were eliminated (i.e., there were only 7 indicators as opposed to all 10), the BIC performed best (see Table 4 and Figures 7, 8, and 11) of the ICs. On the other hand, when sample size was $n = 200$ and class sizes were unequal, the ABIC performed best (see Table 4, Figures 9 and 10), especially when "poor" items were eliminated (see Figure 13). These results indicate that as sample size decreased (i.e., when $n = 200$), the BIC and the CAIC were the most sensitive to unequal latent class sizes (see Figures 11 and 12) and the ABIC was most sensitive to the inclusion of "poor" indicators (see Figure 13).

4.5.2 Likelihood Ratio Tests: Adjusted LMR-LRT and BLRT

The BLRT outperformed the adjusted LMR-LRT across every four unordered latent class A condition (see Table 4 and Figures 7-10). In fact, recovery rates for the adjusted LMR-LRT index never reached the 95% threshold value across any of the conditions and tended to overestimate the number of latent classes (see Table 4). Recovery rates for the BLRT were extremely high across all conditions considered and interestingly, did not seem to be affected by a reduction in sample size, number of indicators included, or latent class size (see Figure 14). Specifically, recovery rates for the BLRT ranged from 89% to 96% across all conditions considered, however, most conditions yielded recovery rates between 94% and 96%. The recovery rate of 89% was for the condition where sample size was small (i.e., n = 200), latent class sizes were unequal, and all 10 indicators were included in the model. In other words, the recovery rates for the BLRT benefited and increased (from 89% to 94%) solely by removing "poor" indicators (see Figure 14).

4.5.3 Bayesian Indices: BF and cmP

The recovery rates for the Bayes Factor (BF) and Correct Model Probability (cmP) indices were high across most of the conditions considered (see Table 4 and Figures 7-10). Specifically, these indices were able to correctly identify the correct latent class solution 100% of the time when sample size was equal to 10,000, 5,000, and 1,000 (see Figures 15 and 16). Moreover, both the BF and cmP maintained high recovery rates (i.e., between 97% and 100%) when sample size dropped down to n = 500, regardless of the number of indicators included in the model or latent class size (see Figures 15 and 16). On the other hand, when sample size decreased to $n = 200$, recovery rates decreased as well. Specifically, only conditions with equal latent class sizes, and 7 indicators were able to still identify the true four-class solution near 95% of the time (see Figures 15 and 16). Not surprisingly, given that the BF and cmP indices are direct derivatives of the BIC, there seemed to be a similar performance trend in that the cmP and BF were also sensitive to unequal class sizes as sample size decreased (see Figures 11, 15, and 16). Lastly, when sample size was $n = 200$, the recovery rates for the BF index were higher than the cmP index, especially when classes sizes were unequal (see Figures 15 and 16).

4.5.4 Summary of Four Unordered Latent Class Model A

Taken together, results from this condition indicated a relationship between sample size and fit index performance. Specifically, as sample size decreased, so did recovery rates for the most part (except for the BLRT, which remained relatively stable across all sample sizes and the AIC which counter-intuitively decreased in performance and sample size increased). Moreover, as sample size decreased, recovery rates were generally higher when the latent classes were equal in size compared to unequal.

Results also indicated sensitivity to the number of indicators included in the model. Specifically, as sample size decreased, conditions with only 7 indicators had higher recovery rates than conditions with all 10 indicators when class size was held constant. In other words, it looks like removing “poor” indicators in unordered latent class solutions increased the chances of recovering the true latent class solution, especially when sample size was small. Results from this unordered condition indicated that when sample size was small and latent classes were equal in size, the BIC performed the best among the ICs. On the other hand, when sample size was small and the latent class sizes were unequal, the ABIC performed the best. Moreover, results from this condition also indicated that the BLRT is the most stable fit index when taking factors such as sample size, class size, and number of indicators into account (see Figures 11-16). Lastly, results from this condition indicated that the BF index seems to perform marginally better than the cmP index in conditions where sample size is extremely small (i.e., n = 200; see Figures 15 and 16).

Table 3

Percentage of Times the Fit Indices Recovered Each Class in LCA with Four Ordered Classes

		Recovery Rates for Four Ordered Latent Class Conditions																															
		AIC				**CAIC**				**BIC**				**ABIC**				**LMR-LRT**				**BLRT**				**cmP**				**BF**			
		Classes				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*			
Model	*N*	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5
	200	6	74	**18**	2	98	2	**0**	0	91	9	**0**	0	9	79	**12**	0	55	40	**4**	1	24	72	**4**	0	91	9	**0**	0	85	15	**0**	0
	500	0	72	**24**	4	51	49	**0**	0	33	67	**0**	0	0	99	**0**	1	9	80	**10**	1	0	96	**3**	0	33	67	**0**	0	20	80	**0**	0
Equal (7 items)	1,000	0	69	**25**	6	1	99	**0**	0	0	100	**0**	0	0	100	**0**	0	1	83	**15**	1	0	93	**7**	1	0	100	**0**	0	0	100	**0**	0
	5,000	0	39	**44**	17	0	100	**0**	0	0	100	**0**	0	0	100	**0**	0	0	78	**20**	2	0	78	**21**	0	0	100	**0**	0	0	100	**0**	0
	10,000	0	11	**64**	25	0	100	**0**	0	0	100	**0**	0	0	99	**1**	0	0	60	**37**	3	0	47	**52**	1	0	100	**0**	0	0	100	**0**	0
	200	0	50	**32**	19	68	32	**0**	0	39	61	**0**	0	0	62	**30**	8	27	67	**5**	1	1	95	**3**	1	39	61	**0**	0	30	70	**0**	0
	500	0	39	**38**	24	0	100	**0**	0	0	100	**0**	0	0	96	**4**	0	2	91	**7**	0	0	90	**9**	1	0	100	**0**	0	0	100	**0**	0
Equal (10 items)	1,000	0	23	**39**	38	0	100	**0**	0	0	100	**0**	0	0	97	**3**	0	0	88	**11**	1	0	76	**22**	1	0	100	**0**	0	0	100	**0**	0
	5,000	0	0	**42**	58	0	100	**0**	0	0	98	**2**	0	0	43	**57**	0	0	12	**79**	8	0	1	**95**	5	0	98	**2**	0	0	97	**0**	0
	10,000	0	0	**37**	63	0	52	**48**	0	0	31	**69**	0	0	1	**99**	0	0	0	**90**	9	0	0	**95**	5	0	31	**69**	0	0	23	**77**	0
	200	21	64	**11**	4	100	0	**0**	0	99	1	**0**	0	29	61	**8**	2	72	26	**2**	0	52	46	**2**	0	99	1	**0**	0	98	2	**0**	0
	500	1	73	**21**	5	96	4	**0**	0	84	16	**0**	0	12	87	**1**	0	38	56	**6**	0	8	88	**4**	0	84	16	**0**	0	75	25	**0**	0
Unequal (7 items)	1,000	0	68	**25**	7	42	58	**0**	0	26	74	**0**	0	0	100	**0**	0	7	81	**11**	1	0	92	**8**	0	26	74	**0**	0	17	83	**0**	0
	5,000	0	46	**39**	15	0	100	**0**	0	0	100	**0**	0	0	100	**0**	0	0	80	**19**	1	0	87	**13**	0	0	100	**0**	0	0	100	**0**	0
	10,000	0	29	**53**	18	0	100	**0**	0	0	100	**0**	0	0	100	**0**	0	0	72	**26**	2	0	75	**25**	1	0	100	**0**	0	0	100	**0**	0
	200	2	52	**29**	18	96	4	**0**	0	82	18	**0**	0	3	65	**23**	10	55	42	**3**	0	10	86	**4**	0	82	18	**0**	0	74	26	**0**	0
	500	0	44	**32**	23	30	70	**0**	0	15	85	**0**	0	0	97	**3**	0	12	78	**9**	1	0	91	**9**	0	15	85	**0**	0	8	92	**0**	0
Unequal (10 items)	1,000	0	34	**37**	29	0	100	**0**	0	0	100	**0**	0	0	99	**1**	0	1	89	**9**	1	0	87	**11**	1	0	100	**0**	0	0	100	**0**	0
	5,000	0	0	**40**	60	0	100	**0**	0	0	100	**0**	0	0	90	**10**	0	0	39	**55**	6	0	14	**83**	3	0	100	**0**	0	0	100	**0**	0
	10,000	0	0	**37**	63	0	98	**2**	0	0	95	**5**	0	0	35	**65**	0	0	5	**83**	12	0	0	**96**	4	0	95	**5**	0	0	92	**8**	0

Table 4

Percentage of Times the Fit Indices Recovered Each Class in LCA with Four Unordered Classes A

		Recovery Rates for Four Unordered Latent Class A Conditions																															
		AIC				**CAIC**				**BIC**				**ABIC**				**LMR-LRT**				**BLRT**				**cmP**				**BF**			
		Classes				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*			
Model	*N*	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5
Equal (7 items)	200	0	0	**82**	18	0	19	**81**	0	0	6	**94**	0	0	0	**89**	11	7	10	**56**	27	0	0	**96**	4	0	0	**94**	6	0	3	**97**	0
	500	0	0	**71**	29	0	0	**100**	0	0	0	**100**	0	0	0	**99**	1	0	1	**70**	29	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0
	1,000	0	0	**67**	33	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**76**	24	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
	5,000	0	0	**63**	37	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**79**	21	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
	10,000	0	0	**63**	37	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**81**	19	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0
Equal (10 items)	200	0	0	**50**	50	0	42	**58**	0	0	18	**82**	0	0	0	**63**	37	12	16	**60**	12	0	0	**96**	4	0	18	**82**	0	0	10	**90**	0
	500	0	0	**44**	56	0	0	**100**	0	0	0	**100**	0	0	0	**98**	2	1	0	**85**	13	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
	1,000	0	0	**42**	58	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**85**	14	0	0	**96**	4	0	0	**100**	0	0	0	**100**	0
	5,000	0	0	**31**	69	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**85**	15	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0
	10,000	0	0	**35**	65	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**84**	16	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
Unequal (7 items)	200	0	1	**85**	15	0	75	**24**	0	0	50	**50**	0	0	1	**90**	9	8	23	**48**	22	0	3	**94**	3	0	50	**50**	0	0	33	**67**	0
	500	0	0	**77**	23	0	3	**97**	0	0	0	**100**	0	0	0	**99**	1	1	2	**66**	31	0	0	**96**	4	0	0	**100**	0	0	0	**100**	0
	1,000	0	0	**68**	32	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**71**	28	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
	5,000	0	0	**67**	33	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**80**	20	0	0	**96**	4	0	0	**100**	0	0	0	**100**	0
	10,000	0	0	**61**	39	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**81**	19	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0
Unequal (10 items)	200	0	0	**50**	50	2	89	**9**	0	0	73	**27**	0	0	1	**67**	32	13	38	**41**	7	0	5	**89**	6	0	73	**27**	0	0	60	**40**	0
	500	0	0	**47**	53	0	9	**91**	0	0	3	**97**	0	0	0	**97**	3	2	4	**84**	10	0	0	**95**	5	0	3	**97**	0	0	2	**98**	0
	1,000	0	0	**43**	57	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	1	0	**86**	14	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0
	5,000	0	0	**43**	57	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**87**	13	0	0	**96**	4	0	0	**100**	0	0	0	**100**	0
	10,000	0	0	**40**	60	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**87**	13	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0

Figure 7. **Recovery rates for four equal unordered classes A (7 items)**

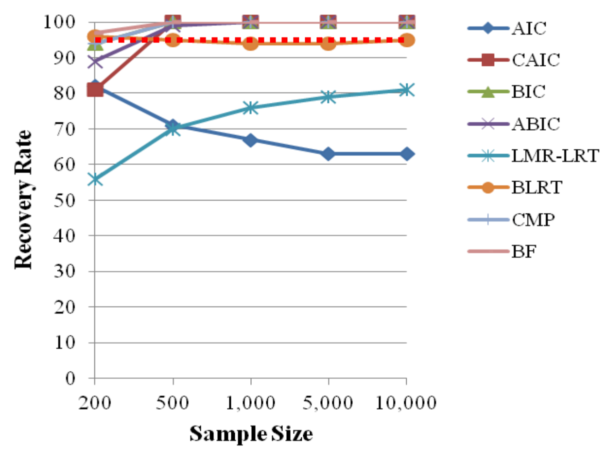


Figure 8. **Recovery rates for four equal unordered classes A (10 items)**

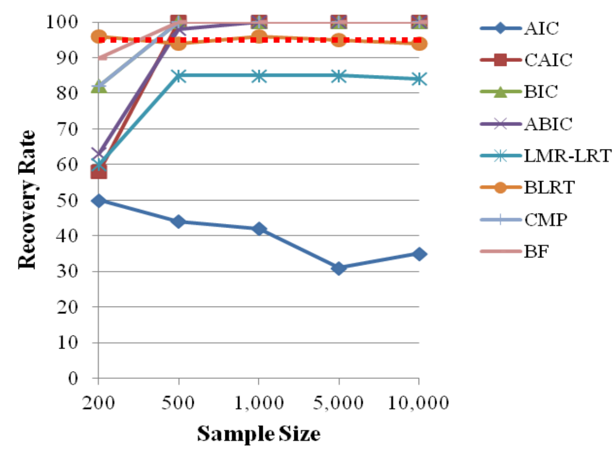


Figure 9. **Recovery rates for four unequal unordered classes A (7 items)**

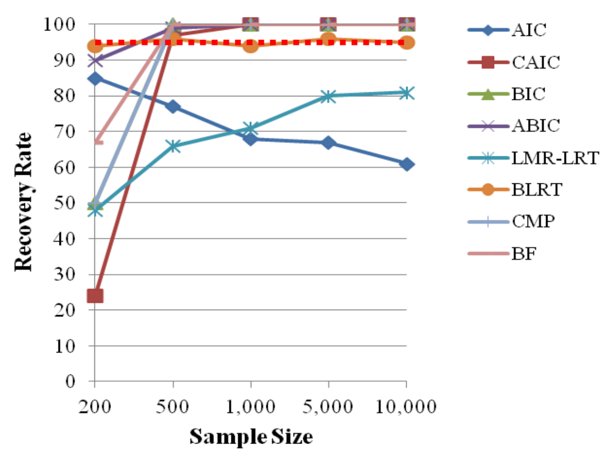


Figure 10. **Recovery rates for four unequal unordered classes A (10 items)**

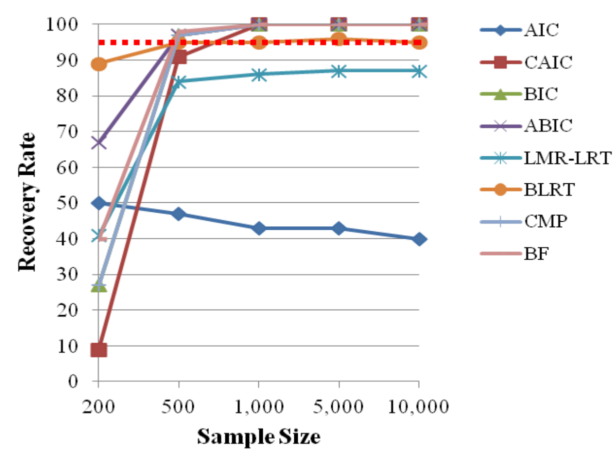


Figure 11. **Recovery rates for BIC across conditions for four unordered classes A**

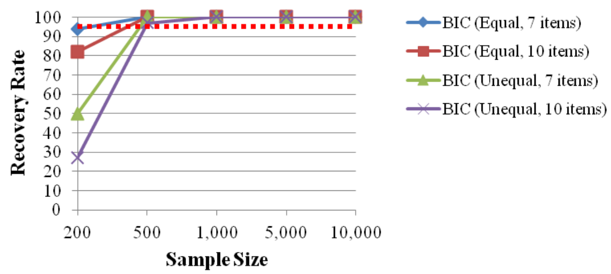


Figure 12. **Recovery rates for CAIC across conditions for four unordered classes A**

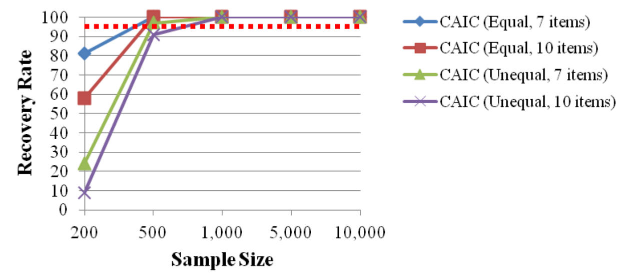


Figure 13. **Recovery rates for ABIC across conditions for four unordered classes A**

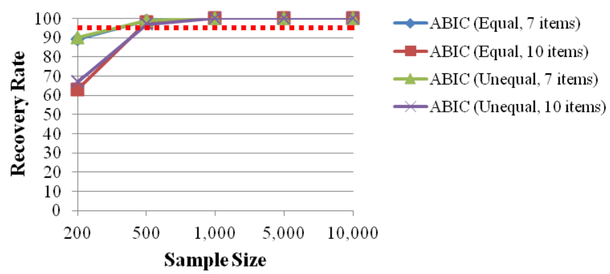


Figure 14. **Recovery rates for BLRT across conditions for four unordered classes A**

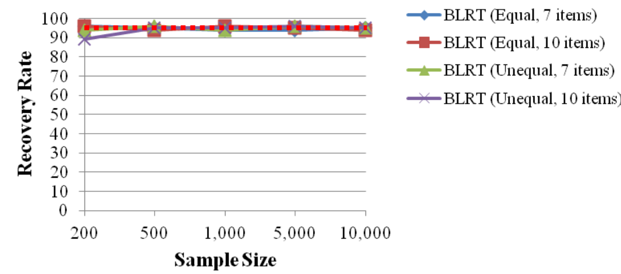


Figure 15. **Recovery rates for cmP across conditions for four unordered classes A**

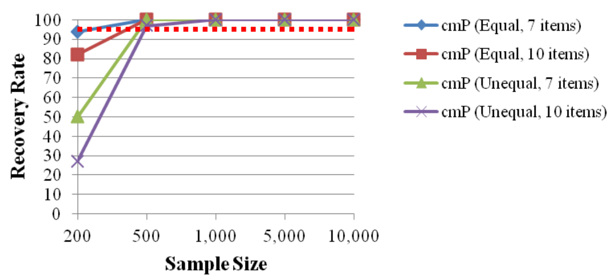
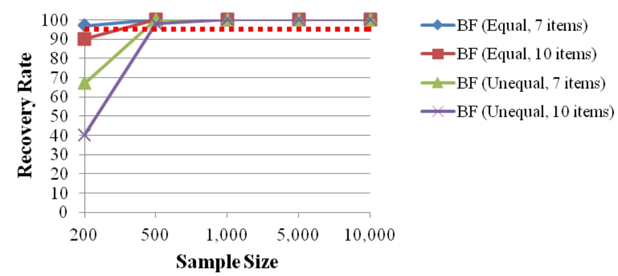


Figure 16. **Recovery rates for BF across conditions for four unordered classes A**



4.6 Four Unordered Latent Class Model B

The recovery rates across all four unordered latent class B conditions (see Figure 4) are presented in Table 5. If all fit indices were able to correctly identify the true four-class solution, the percentages in columns identifying four latent classes in Table 5 should result in the highest values. Figures 17-20 display a graphical representation of recovery rates for all fit indices across all four unordered latent class B conditions. In general, as sample size decreased, the fit indices yielded lower recovery rates in the unordered latent class B conditions when compared to the unordered latent class A conditions. One highly plausible explanation for this finding is that the latent classes were generated to have greater between class separation in the unordered latent class A conditions when compared to the unordered latent class B conditions (see Figures 3 and 4), which allows for easier detection of the true latent class solution.

4.6.1 Information Criteria (ICs)

The AIC performed the worst among the ICs, with recovery rates well below the 95% benchmark across all unordered latent class B conditions considered (see Table 5 and Figures 17-20). Similar to the four unordered class A condition, recovery rates for this condition show that the AIC is prone to overestimating the correct number of latent classes (see Table 5). Additionally, recovery rates for the AIC tended to increase as sample size decreased (see Table 5). When classes sizes were equal, the CAIC, BIC, and ABIC, all performed extremely and equally well (i.e., 100% recovery rates) across all conditions with sample sizes of 1,000, 5,000 and

10,000 (see Figures 21-23). However, as sample size decreased to *n* = 500, recovery rates for the BIC and CAIC began to decrease as well, but still remained high at 99% for the ABIC, regardless of the number of indicators included in the model (See Figures 21-23). Additionally, as sample size decreased to *n* = 500, the BIC and CAIC displayed the greatest sensitivity to unequal latent classes (see Figures 21 and 22). When sample size decreased even further to *n* = 200, none of the ICs met the 95% threshold value, however the ABIC performed the best among the ICs (i.e., recovery rates fell between 74% and 85%) even with this reduction in sample size (see Figures 21-23). Moreover, the CAIC, BIC and ABIC showed sensitivity to "poor" indicators (see Figures 21 and 22) because when holding class size constant, recovery rates were marginally lower when all 10 indicators were included in the model as opposed to just 7. The ABIC did not show the same level of sensitivity to both a reduction in sample size and latent class size when compared to the BIC and CAIC (see Figures 21-23). In general, when taking all factors into account, the ABIC performed the best among the ICs under the four unordered latent class B condition (especially as sample size decreased).

4.6.2 Likelihood Ratio Tests: Adjusted LMR-LRT and BLRT

The BLRT once again outperformed the adjusted LMR-LRT across every four unordered latent class B condition (see Table 5 and Figures 17-20). In fact, recovery rates for the adjusted LMR-LRT index never reached the 95% threshold value and showed a tendency to overestimate the correct number of latent classes when sample size was large and a tendency to underestimate the correct number of latent classes as sample size decreased (see Table 5). Recovery rates for the BLRT were high across all conditions (i.e., between 93% and 96%) except when sample size decreased to *n* = 200 (in which case recovery rates for the BLRT fell between 65% and 85%; see Figure 24). Although not as drastic as the ICs, the BLRT also showed sensitivity to sample size, unequal classes, and the inclusion of "poor" indicators (see Figure 24).

4.6.3 Bayesian Indices: BF and cmP

The recovery rates for the Bayes Factor (BF) and Correct Model Probability (cmP) indices were highest across conditions with large sample sizes (see Table 5 and Figures 25-26). Specifically, these indices yielded high recovery rates (i.e., between 97% and 100%) when sample size was equal to 10,000, 5,000, or 1,000 regardless of whether the number of indicators included and latent class size (see Table 5 and Figures 25-26). When sample size dropped down to *n* = 500, recovery rates remained highest (i.e., between 96% and 99%) when classes were equal and only 7 indicators were included in the model (see Figures 25 and 26). Additionally, when sample size was *n* = 500, the cmP and BF showed sensitivity to unequal latent classes (i.e., recovery rates dropped and fell in between 42% and 67%), especially when all 10 indicators were included in the model. When sample size dropped down to *n* = 200, recovery rates were extremely low across the board for both the BF and cmP (i.e., between 3% and 21%; see Figures 25 and 26). Lastly, as sample size decreased, the recovery rates for the BF were slightly higher when compared to the cmP (see Table 5).

4.6.4 Summary of Four Unordered Latent Class Model B

Taken together, results from this condition also indicated a relationship between sample size and fit index performance. Specifically, as sample size decreased, so did the recovery rates (except for the AIC). Results also indicated that as sample size decreased, recovery rates were generally higher when latent class sizes were equal compared to unequal. Additionally, results indicated sensitivity to the number of indicators included in the model. Specifically, when holding class size constant and as sample size decreased, conditions with only 7 indicators had higher recovery rates than conditions with all 10 indicators. In other words, it appears that removing "poor" indicators in unordered latent class solutions increased the chances of recovering the true latent class solution, especially when sample size was small. Lastly, when taking all fit indices into account, results from this condition indicated that the ABIC and BLRT are the least sensitive to (although still affected by) small sample sizes (see Figures 21-26).

Table 5

Percentage of Times the Fit Indices Recovered Each Class in LCA with Four Unordered Classes B

		Recovery Rates for Four Unordered Latent Class B Conditions																															
		AIC				**CAIC**				**BIC**				**ABIC**				**LMR-LRT**				**BLRT**				**cmP**				**BF**			
		Classes				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*			
Model	*N*	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5	≥2	3	**4**	5
Equal (7 items)	200	0	1	**79**	20	75	22	**3**	0	45	46	**9**	0	0	2	**85**	13	42	27	**26**	5	0	8	**85**	6	45	46	**9**	0	27	52	**21**	0
	500	0	0	**77**	23	0	11	**89**	0	0	4	**96**	0	0	0	**99**	1	12	6	**70**	12	0	0	**96**	4	0	4	**96**	0	0	1	**99**	0
	1,000	0	0	**77**	23	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	3	0	**83**	14	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0
	5,000	0	0	**71**	29	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**82**	18	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0
	10,000	0	0	**69**	31	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**82**	18	0	0	**95**	5	0	0	**100**	0	0	0	**100**	0
Equal (10 items)	200	0	1	**66**	33	92	8	**0**	0	67	30	**3**	0	0	2	**74**	24	53	28	**17**	2	0	12	**82**	6	67	30	**3**	0	50	44	**6**	0
	500	0	0	**60**	40	2	26	**72**	0	0	12	**88**	0	0	0	**99**	1	12	9	**69**	9	0	0	**96**	4	0	12	**88**	0	0	7	**93**	0
	1,000	0	0	**55**	45	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	1	0	**87**	12	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
	5,000	0	0	**46**	54	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**84**	16	0	0	**93**	7	0	0	**100**	0	0	0	**100**	0
	10,000	0	0	**43**	57	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**86**	14	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
Unequal (7 items)	200	0	14	**75**	11	72	28	**0**	0	40	57	**3**	0	0	17	**76**	7	42	35	**19**	4	0	28	**70**	2	40	57	**3**	0	24	69	**8**	0
	500	0	0	**82**	18	0	66	**34**	0	0	45	**55**	0	0	1	**98**	1	17	20	**53**	10	0	0	**95**	5	0	45	**55**	0	0	33	**67**	0
	1,000	0	0	**77**	23	0	6	**94**	0	0	3	**97**	0	0	0	**100**	0	7	2	**78**	13	0	0	**94**	6	0	3	**97**	0	0	1	**99**	0
	5,000	0	0	**70**	30	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**83**	17	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
	10,000	0	0	**73**	27	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**80**	20	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
Unequal (10 items)	200	0	8	**67**	25	94	6	**0**	0	75	24	**2**	0	0	11	**75**	14	57	30	**12**	1	0	31	**65**	3	75	24	**2**	0	60	38	**2**	0
	500	0	0	**58**	42	3	77	**20**	0	0	58	**42**	0	0	2	**97**	1	25	26	**43**	6	0	1	**94**	6	0	58	**42**	0	0	46	**53**	0
	1,000	0	0	**54**	46	0	8	**92**	0	0	3	**97**	0	0	0	**100**	0	11	1	**80**	9	0	0	**96**	4	0	3	**97**	0	0	0	**100**	0
	5,000	0	0	**51**	49	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**86**	14	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0
	10,000	0	0	**46**	54	0	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**87**	13	0	0	**94**	6	0	0	**100**	0	0	0	**100**	0

Figure 17. **Recovery rates for four equal unordered classes B (7 items)**

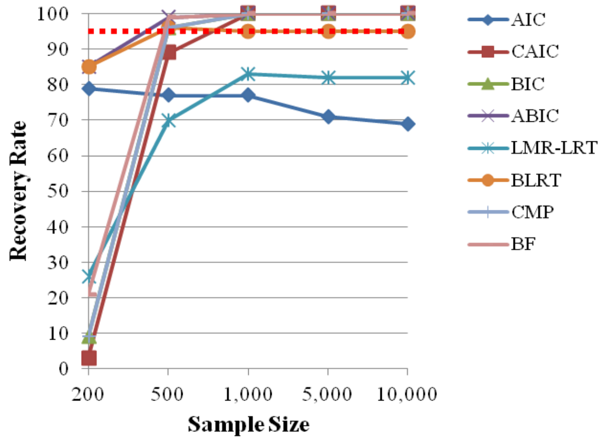


Figure 18. **Recovery rates for four equal unordered classes B (10 items)**

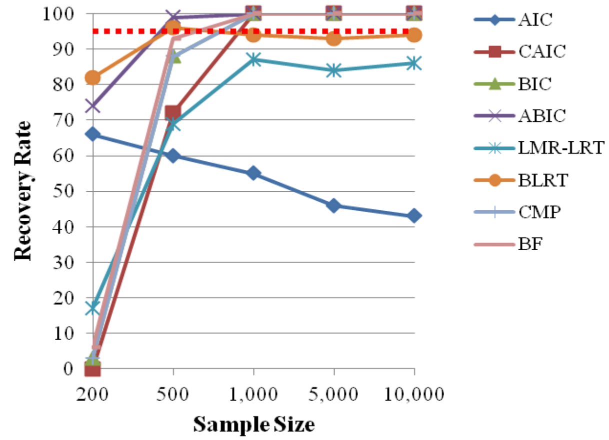


Figure 19. **Recovery rates for four unequal unordered classes B (7 items)**

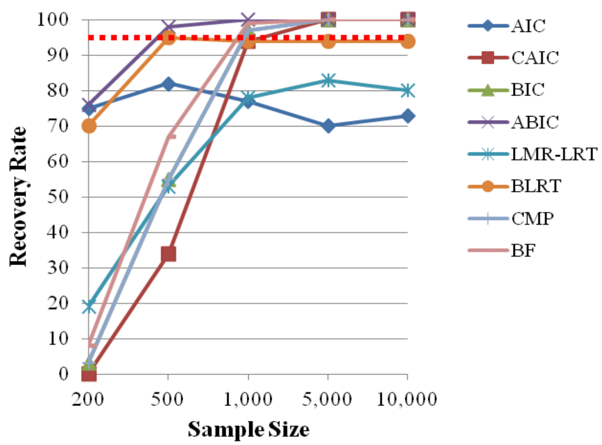


Figure 20. **Recovery rates for four unequal unordered classes B (7 items)**

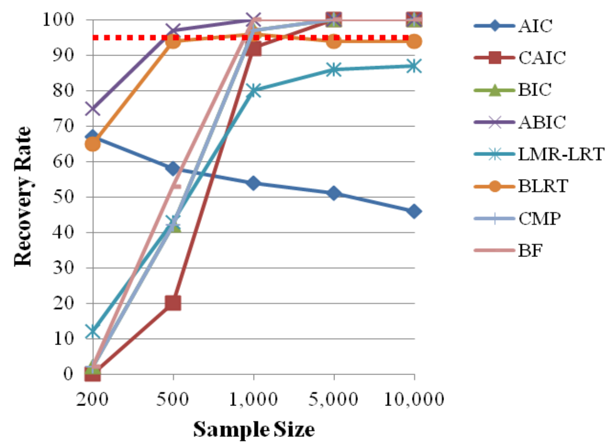


Figure 21. **Recovery rates for BIC across conditions for four unordered classes B**

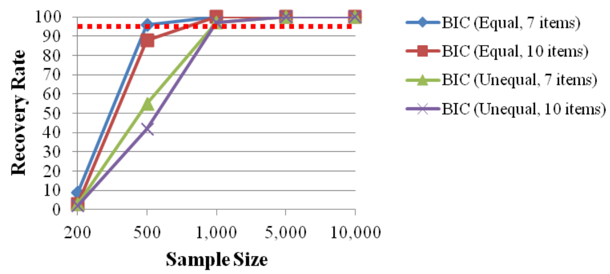


Figure 22. **Recovery rates for CAIC across conditions for four unordered classes B**

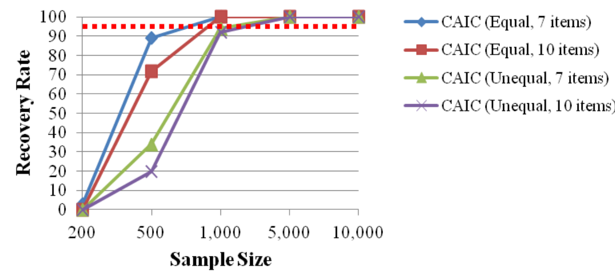


Figure 23. **Recovery rates for ABIC across conditions for four unordered classes B**

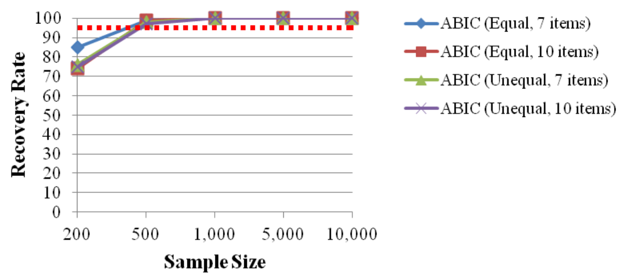


Figure 24. **Recovery rates for BLRT across conditions for four unordered classes B**

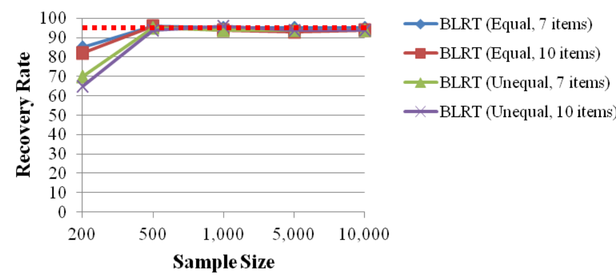


Figure 25. **Recovery rates for cmP across conditions for four unordered classes B**

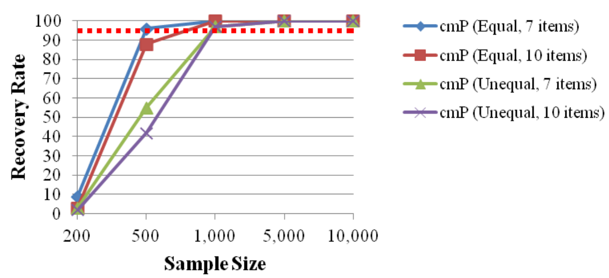
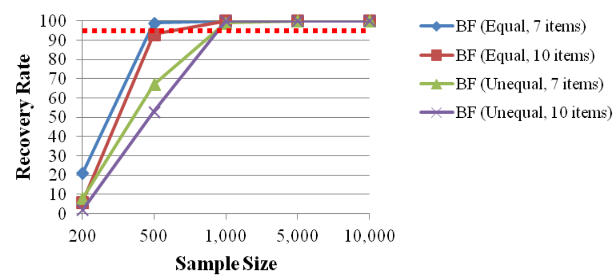


Figure 26. **Recovery rates for BF across conditions for four unordered classes B**



4.7 Three Ordered Latent Class Model

The recovery rates across all conditions where data was generated to have three ordered latent classes (see Figure 5) are presented in Table 6. As previously mentioned, the extremely low recovery rates for the four ordered latent class condition prompted the current simulation study to consider another ordered latent class solution that was generated to have greater between-class separation (i.e., the current three ordered class condition). If all fit indices were able to correctly identify the true three-class solution, the percentages in columns identifying three latent classes in Table 6 should result in the highest values. Figures 27-30 display a graphical representation of recovery rates for all fit indices considered across all conditions where data was generated to have three ordered latent classes. In general, recovery rates for the ordered three-class conditions were extremely better than the recovery rates for the ordered four-class conditions (see Tables 3 and 6). This is most likely due to the ordered three-class condition being generated to have greater between-class separation and therefore tended to underestimate the correct number of latent classes at a lower rate.

4.7.1 Information Criteria (ICs)

The AIC performed equally poorly across all three ordered latent class conditions (see Table 6 and Figures 27-30). Similar to results from the four unordered class conditions, the AIC tended to overestimate the true number of latent classes (see Table 6). Additionally, recovery rates for the AIC tended to increase as sample size decreased. When class sizes were equal (regardless of the number of indicators included in the model) or unequal with all 10 indicators included, the BIC and CAIC performed extremely well (i.e., recovery rates between 95% and 100%) when sample size was 10,000, 5,000, 1,000 and 500 (see Figures 31 and 32). However, under the same conditions, the BIC and CAIC displayed sensitivity to unequal latent classes that only included 7 indicators (i.e., recovery rates fell to 47% for the BIC and 26% for the CAIC; see Figures 31 and 32). In other words, when ordered classes are unequal and sample size is small (i.e., n = 500), the BIC and CAIC perform a lot better (i.e., recovery rates between 96% and 99%) when all 10 indicators are included in the model as opposed to only 7. When sample size dropped even further to n = 200, the BIC and CAIC maintained high recovery rates of 99% and 94% respectively, when classes were equal and all 10 indicators were included in the model (see Figures 31 and 32). When sample size was n = 200 and class sizes were unequal, the recovery rates for the BIC and CAIC decreased dramatically (i.e., between 0% and 41%) and yielded a preference for including all 10 indicators in the model (see Figures 31 and 32). The ABIC yielded the highest recovery rates among the ICs when taking all factors into account (see Figure 33). Specifically, regardless of sample size, the number of indicators included in the model, or latent class size, the ABIC performed extremely well (between 97% and 100%) when sample size was equal to 10,000, 5,000, 1,000, or 500 (see Figure 33). As sample size decreased to n = 200, recovery rates for the ABIC decreased as well (i.e., fell between 69% and 84%) and performed best with a recovery rate of 84% when classes were equal, and when the number of indicators included was reduced to 7.

4.7.2 Likelihood Ratio Tests: Adjusted LMR-LRT and BLRT

The BLRT outperformed the adjusted LMR-LRT across all ordered three class conditions (see Table 6 and Figures 27-30). In fact, recovery rates for the adjusted LMR-LRT index never reached the 95% threshold value and showed a tendency to overestimate the correct number of latent classes (see Table 6). The BLRT yielded high recovery rates between 92% and 97% across all sample sizes and conditions with one exception. Specifically, when class sizes were unequal, sample size was *n* = 200, and only 7 items were included in the model, the recovery rate for the BLRT dropped down to 62% (see Figure 34). This finding is most likely due to the fact that when classes are ordered in nature, fit index performance is generally higher when all 10 indicators are included in the model as opposed to only 7 (see Table 6).

4.7.3 Bayesian Indices: BF and cmP

Recovery rates for the Bayes Factor (BF) and Correct Model Probability (cmP) were extremely high across all modeling conditions considered (i.e., between 99% and 100%) when sample size was equal to 1,000, 5,000 or 10,000, regardless of latent class size and the number of indicators included in the model (see Figures 35 and 36). When sample size decreased to *n* = 500, recovery rates for the cmP and BF largely decreased (i.e., to 47% and 62%, respectively) only for conditions with unequal class sizes and only 7 items (see Figures 35 and 36). In other words, when sample size was *n* = 500 and class sizes were unequal, the cmP and BF highly benefited from solely leaving in all available indicators (i.e., recovery rates (i.e., recovery rates increased and feel between 99% and 100%). When sample size decreased even more to *n* = 200, recovery rates for the cmP and BF were above the 95% benchmark only when class sizes were equal and all 10 indicators were included in the model (see Figures 35 and 36). These results indicate that as sample size decreased with ordered class solutions, the cmP and BF maintained the highest performance when all potential indicators were included in the model and class sizes were relatively equal. In general, as sample size decreased, the BF seemed to marginally outperform the cmP.

4.7.4 Summary of Three Ordered Latent Class Model

Taken together, results from this condition also indicate a relationship between sample size and fit index performance. Specifically, as sample size decreased, so did the recovery rates (except for the AIC). Results also indicated that as sample size decreased, recovery rates were generally higher when latent class sizes were equal compared to unequal. Moreover, results indicated sensitivity to the number of indicators included ordered latent class models. Specifically, as sample size decreased with ordered class solutions, conditions with all 10 indicators generally had higher recovery rates than conditions with only 7 indicators when class size was held constant. In other words, when latent classes are ordered, removing potentially redundant indicators tends to decrease the chances of recovering the true latent class solution, especially when sample size is small. Lastly, when classes are ordered and sample size was small, the ABIC and BLRT had the highest recovery rates across the board (see Figures 31-36).

4.8 Sample Size, Number of Indicators, and Class Enumeration

In addition to examining fit index performance, a second goal of this simulation study was to understand how sample size, number of indicators, and class enumeration intersect in LCA models. As expected, results from this simulation study show a relatively strong interplay between sample size, number of indicators, and class enumeration. Specifically, as sample size decreased, the majority of the fit indices decreased in their ability to recover the correct latent class solution (see Figure 11-16, 21-26, and 31-36). Additionally, when the population sample was subsetted to n = 200 (or n = 500 in some cases), the majority of the fit indices showed a sensitivity to the number of indicators that were included in the model. In other words, when sample size decreased, the number of indicators included in the LCA model became more important. Specifically, when sample size was small, latent classes were unordered, and class size was held constant (i.e., recovery rates for conditions with 7 and 10 indicators were compared separately for equal class conditions and unequal class conditions), the fit indices performed best when "poor" indicators were eliminated (see Figures 11-16 and 21-26). On the other hand, when sample size was small, latent classes were ordered, and class size was held constant, the fit indices performed best when all 10 indicators were included in the model (i.e., when all available information was included in the model; see Figures 31-36). Lastly, as sample size decreased and the number of indicators included was held constant, the fit indices showed generally higher recovery rates when latent classes were equal in nature as opposed to unequal in nature. This is most likely due to the fact that as the analysis sample size decreased, the smallest latent class in the unequal conditions (i.e., the latent class comprised of solely 10% of the overall population) was underrepresented. In fact, when looking at Tables 4-6, there seems to be a tendency for the fit indices to underestimate the correct number of latent classes in the unequal latent class conditions, as sample size decreases.

Table 6

Percentage of Times the Fit Indices Recovered Each Class in LCA with Three Ordered Classes

		Recovery Rates for Three Ordered Latent Conditions																															
		AIC				**CAIC**				**BIC**				**ABIC**				**LMR-LRT**															
		Classes				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*				*Classes*			
Model	*N*	≥2	**3**	4	5	≥2	**3**	4	5	≥2	**3**	4	5	≥2	**3**	4	5	≥2	**3**	4	5	≥2	**3**	4	5	≥2	**3**	4	5	≥2	**3**	4	5
Equal (7 items)	200	1	**77**	18	4	81	**19**	0	0	59	**41**	0	0	1	**84**	14	2	27	**69**	5	0	4	**92**	4	0	59	**41**	0	0	44	**56**	0	0
	500	0	**77**	19	4	3	**97**	0	0	1	**99**	0	0	0	**98**	2	0	0	**87**	11	1	0	**94**	6	0	1	**99**	0	0	0	**100**	0	0
	1,000	0	**74**	21	5	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**88**	11	1	0	**95**	5	0	0	**100**	0	0	0	**100**	0	0
	5,000	0	**64**	28	8	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**76**	20	4	0	**94**	6	0	0	**100**	0	0	0	**100**	0	0
	10,000	0	**63**	29	8	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**80**	18	1	0	**93**	7	0	0	**100**	0	0	0	**100**	0	0
Equal (10 items)	200	0	**61**	25	14	6	**94**	0	0	1	**99**	0	0	0	**73**	20	7	5	**86**	9	0	0	**97**	3	0	1	**99**	0	0	0	**100**	0	0
	500	0	**53**	32	15	0	**100**	0	0	0	**100**	0	0	0	**99**	1	0	1	**86**	12	1	0	**95**	5	0	0	**100**	0	0	0	**100**	0	0
	1,000	0	**52**	33	15	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**88**	10	1	0	**97**	3	0	0	**100**	0	0	0	**100**	0	0
	5,000	0	**40**	38	22	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**87**	11	2	0	**96**	4	0	0	**100**	0	0	0	**100**	0	0
	10,000	0	**36**	35	29	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**85**	12	2	0	**92**	8	0	0	**100**	0	0	0	**100**	0	0
Unequal (7 items)	200	11	**66**	19	3	100	**0**	0	0	96	**4**	0	0	17	**69**	12	2	54	**41**	5	0	36	**62**	2	0	96	**4**	0	0	92	**8**	0	0
	500	0	**76**	19	4	74	**26**	0	0	53	**47**	0	0	2	**97**	1	0	9	**80**	10	1	0	**95**	5	0	53	**47**	0	0	38	**62**	0	0
	1,000	0	**70**	25	5	6	**94**	0	0	1	**99**	0	0	0	**100**	0	0	1	**83**	15	1	0	**94**	6	0	1	**99**	0	0	1	**99**	0	0
	5,000	0	**68**	25	7	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**81**	17	2	0	**94**	6	0	0	**100**	0	0	0	**100**	0	0
	10,000	0	**67**	26	8	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**82**	17	1	0	**94**	6	0	0	**100**	0	0	0	**100**	0	0
Unequal (10 items)	200	0	**57**	28	15	81	**19**	0	0	59	**41**	0	0	0	**72**	21	7	24	**67**	8	0	3	**92**	5	0	59	**41**	0	0	48	**52**	0	0
	500	0	**55**	29	16	4	**96**	0	0	1	**99**	0	0	0	**99**	1	0	1	**88**	10	1	0	**95**	5	0	1	**99**	0	0	0	**100**	0	0
	1,000	0	**51**	29	19	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**88**	11	1	0	**95**	5	0	0	**100**	0	0	0	**100**	0	0
	5,000	0	**42**	34	24	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**86**	13	1	0	**94**	6	0	0	**100**	0	0	0	**100**	0	0
	10,000	0	**39**	38	23	0	**100**	0	0	0	**100**	0	0	0	**100**	0	0	0	**86**	12	1	0	**93**	7	0	0	**100**	0	0	0	**100**	0	0

Figure 27. **Recovery rates for three equal ordered classes (7 items)**

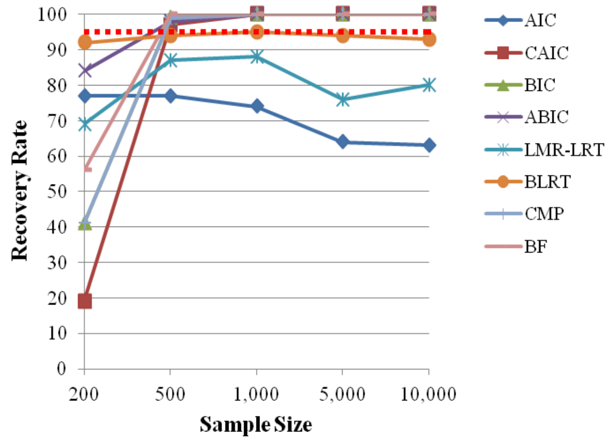


Figure 28. **Recovery rates for three equal ordered classes (10 items)**

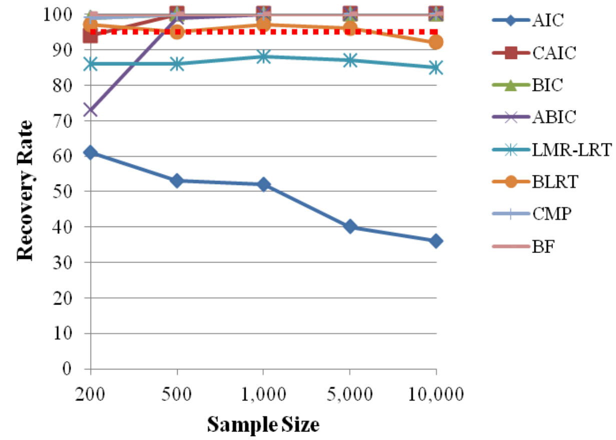


Figure 29. **Recovery rates for three unequal ordered classes (7 items)**

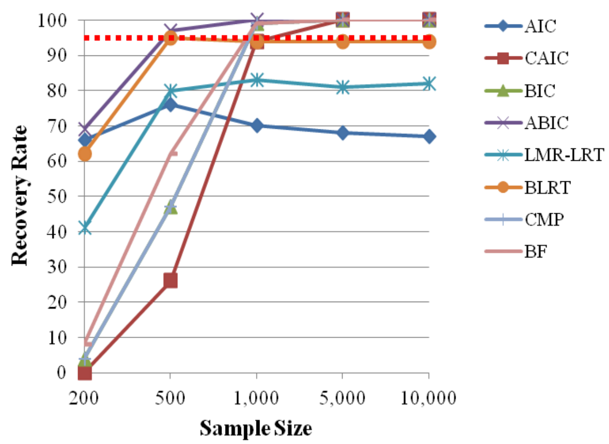


Figure 30. **Recovery rates for three unequal ordered classes (10 items)**

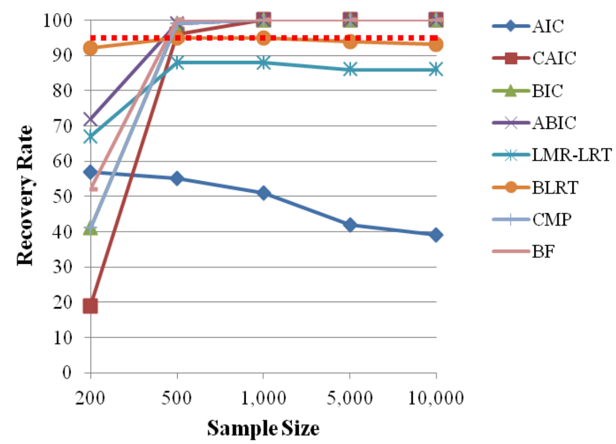


Figure 31. **Recovery rates for BIC across conditions for three ordered classes**

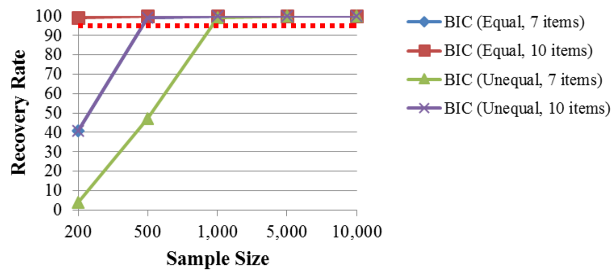


Figure 32. **Recovery rates for CAIC across conditions for three ordered classes**

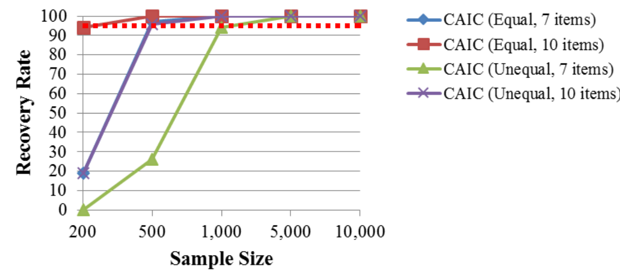


Figure 33. **Recovery rates for ABIC across conditions for three ordered classes**

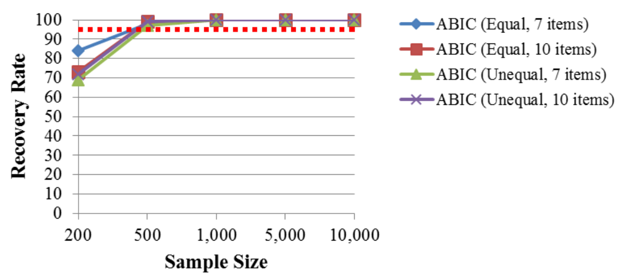


Figure 34. **Recovery rates for BLRT across conditions for three ordered classes**

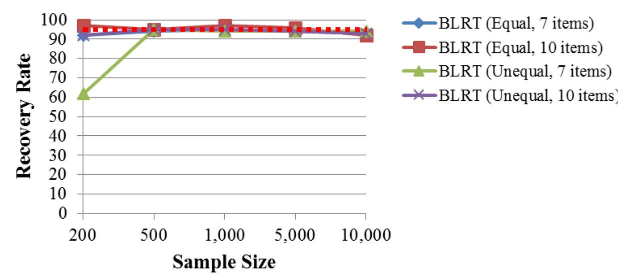


Figure 35. **Recovery rates for cmP across conditions for three ordered classes**

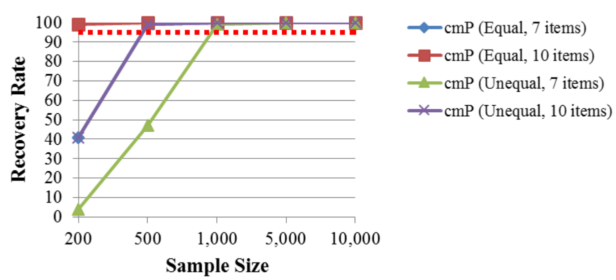
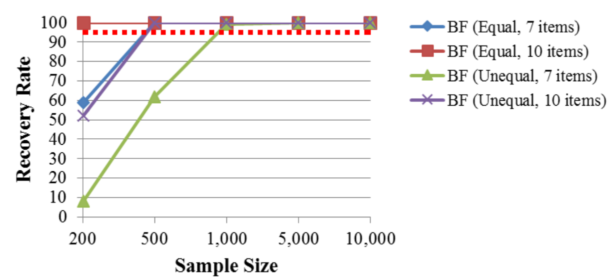


Figure 36. **Recovery rates for BF across conditions for three ordered classes**



Chapter 5

Discussion

5.1 Overview

This study investigated the performance of the most commonly used fit indices in selecting the “correct” latent class model while varying factors such as: the true number of latent classes, the size of the latent classes (i.e., class prevalence), the nature of the latent classes, the number of indicators, and sample size. Additionally, this study aimed to empirically examine the intersection of sample size, the number of observed indicators, and class enumeration in LCA models. This discussion section will first address general trends and patterns that were gleamed from examining recovery rates across the various LCA conditions explored in this study. While identifying these general trends, this chapter will also provide general recommendations for applied researchers that use LCA models in their research. Next, this discussion section will summarize results related to the performance of the ICs (i.e., the AIC, CAIC, BIC, and ABIC), the likelihood ratio tests (i.e., the adjusted LMR-LRT and BLRT), and the newer Bayesian indices (i.e., the BF and cmP), as well as provide specific recommendations for their use in practice. Finally, limitations of this simulation study will be discussed and future directions will be suggested for forthcoming Monte Carlo simulation research.

5.2. General Trends and Practical Recommendations

Taken together, results from this simulation study indicated that there is a strong relationship between sample size and fit index performance in LCA models.

As expected, as sample size decreased, fit index performance tended to decrease across all conditions considered (see Figures 7-36). This finding is consistent with results from previous simulation studies (Yang, 2006; Tein, Coxe, & Cham, 2013; Tueller & Lubke, 2010). It should be noted that all conditions explored in this simulation study were generated with an original population sample size of 10,000 observations. Therefore, analysis conditions that reduced sample size to n = 5,000, n = 1,000, n = 500, and n = 200, aimed to understand what happens to fit index performance when researchers do not have access to 100% of a population (which is often the case in practice). Specifically, this study wanted to know how the performance of the most commonly used fit indices and thus potential LCA findings as a whole are affected if researchers are only able to access 50% (i.e., n = 5,000), 10% (i.e., n = 1,000), 5% (i.e., n = 500) or 2% (i.e., n = 200) of a population. This approach of subsetting random observations during the analysis stage, rather than generating data with truly different population sample sizes, was a novel way to understand the influence of sample size on fit index performance. With this in mind, findings of this simulation study are optimistic in that all fit indices considered (except the AIC and adjusted LMR-LRT) performed well (i.e., reached or nearly reached a 95% recovery rate) when at least 1,000 observations were included in the analysis (see Tables 4-6). This finding is important for applied researchers that want to use LCA because they can aim to for a sample size of n = 1000 to ensure high fit index performance.

As analysis sample size decreased even further (i.e., *n* = 500 and *n* = 200), the relative size of the latent classes became more important. Specifically, when sample size decreased, fit indices generally performed better when latent classes were relatively equal in size compared to unequal (see Figures 11-16, 21-26, and 31-36). This finding could be attributed to the fact that when the analysis sample size is small (e.g., *n* = 500 or *n* = 200) and classes are unequal, within-class sample size is also influenced resulting in potentially small latent classes that are underpowered and therefore have a lower chance of being detected. In fact, when looking at the recovery rates in rows that indicate a sample size of *n* = 200 in Tables 4-6, we can see a tendency for the fit indices to underestimate the correct latent class solution. This result is in line with previous research findings related to sample size and statistical power. Specifically, Muthén & Muthén (2002) found that as sample size decreased, power to detect the true model decreased as well, in both growth models and confirmatory factor analysis (CFA) models. Similarly, Lubke and Neale (2006) found that increasing the within-class sample size increased the likelihood of recovering the correct number of latent classes in LPA models. Building off the findings from Lubke and Neale (2006), Masyn (2013) suggested that small sample sizes might be underpowered to detect smaller latent classes in the LCA context as well. The current simulation study was able to provide empirical support for this suggested positive relationship between sample size (specifically, within-class sample size) and power to detect the correct latent class solution in LCA models. Therefore, researchers interested in studying small subpopulations within a larger population (e.g., students that are victimized in school or minors that exhibit various symptoms of psychopathology) via LCA should consider over-sampling certain segments of their target population (e.g., those that experience extremely high levels of victimization or exhibit many/all symptoms of psychopathology) to ensure there is enough statistical power to detect these potentially small latent classes. Additionally, researchers should be aware that the fit indices may often times underestimate the correct latent class solution if these small and often underpowered subpopulations exist, therefore, substantive theory should also be heavily relied on during the class enumeration process. For example, if the fit indices point to either a three or four class solution, and substantive theory can justify adding an additional fourth class, researchers should allow substantive theory to drive the decision about which final LCA model they should retain. Given this known relationship between small sample sizes, latent class size, and the underestimation of latent classes, it is suspected that existing LCA studies frequently do not detect the correct number of latent classes due to low statistical power.

Another trend the results yielded was that as analysis sample size decreased, the number of indicators included in the model became more important. In general, when class size was held constant, and when latent classes were unordered, removing "poor" indicators increased the likelihood of recovering the correct latent class solution when sample size was small (see Figures 11-16 and 21-26). For example, Table 4 indicated that when classes were unequal and sample size was n = 200, the BIC identified the correct latent class solution 27% of the time when all 10 indicators were included in the model. However, when “poor” indicators were eliminated (i.e., only 7 indicators were included in the model), the BIC was able to correctly identify the correct model 50% of the time. This increase in fit index performance when “poor” indicators were eliminated was consistent across most unordered latent class conditions (see Tables 4-5).

On the other hand, when class size was held constant, and the latent classes were ordered, keeping all potential indicators in the model (regardless of whether or not they were potentially redundant or contributed to distinguishing latent classes) increased the likelihood of recovering the correct latent class solution when sample size was small (see Figures 31-36). For example, Table 6 indicated that when classes were unequal and sample size was n = 200, the BIC identified the correct latent class solution 41% of the time when all 10 indicators were included in the model. However, when potentially redundant indicators were eliminated (i.e., only 7 indicators were included in the model), the BIC was only able to correctly identify the correct model 4% of the time (see Table 6). This decrease in fit index performance when potentially redundant indicators were eliminated was consistent across most ordered latent class conditions (see Tables 6). Current “best practices” related to class enumeration suggest always removing “poor” indicators (i.e., indicators that do not separate classes well or that have low within-class homogeneity) when fitting LCA models (Masyn, 2013). Therefore, these results are important because they empirically support the current suggestion of removing indicators that have low between class separation and low within-class homogeneity during the model fitting process with unordered latent class solutions. Additionally, researchers often decide to remove redundant indicators (either by evaluating the content of the indicators, or using item correlations and deleting redundant items) in practice. Therefore, these results are also important because researchers should consider including all available indicators when estimating ordered LCA models (as long as the model is identified), even if the indicators are potentially redundant. This finding is most likely due to the fact that as sample size decreases, including all available indicators in ordered LCA models provides more information and helps the estimation process.

Taken together, the general trends highlighted in this section provide evidence for a relatively strong interplay between sample size, number of indicators, and class enumeration in LCA models. Specifically, when sample size is large (i.e., between 1,000 and 10,000), the inclusion of "poor" or potentially redundant indicators in LCA models is less important. However, as sample size decreased, removing "poor" indicators in unordered latent class solutions and including all available indicators in ordered latent class solutions, as well as the relative size of the latent classes (i.e., equal vs. unequal classes) became rather important in ensuring high performance of the fit indices.

5.3 Information Criteria and Practical Recommendations

The results of this simulation study found the AIC to consistently perform the worst when compared the rest of the ICs (see Tables 4-6). Specifically, results indicated that the AIC tended to overestimate the correct number of latent classes across all of the conditions considered. This finding is in line with a host of other simulation studies that focused on fit index performance in mixture models (Celeux & Soromenho, 1996; Koehler & Murphree, 1988; Nylund, Asparouhov, & Muthén, 2007; Tein, Coxe, & Cham, 2013). Additionally, the AIC tends to overestimate the number of latent classes the most when sample size was very large (i.e., n = 10,000). This finding is also consistent with previous research (Woodroofe, 1982). Similarly, Nylund et al. (2007a) found that AIC accuracy decreased as sample size increased and suggested that this was due to the fact that the AIC has no adjustment for sample size. Results of the current simulation study also support this negative relationship between sample size and AIC accuracy.

The CAIC, on the other hand, performed much better across all conditions when compared to the AIC. As previously mentioned, this is likely due to the CAIC's adjustment for the number of parameters and sample size. These findings are also consistent with previous research that found the CAIC to perform well across multiple conditions (Peugh & Fan, 2013). Furthermore, there is evidence to support the CAIC's ability to perform well when sample size is relatively large. Specifically, this study was able to replicate the findings in Nylund et al. (2007a), that the CAIC performs well when sample size is equal to at least n = 1,000. Additionally, findings from this study indicate that the CAIC tends to perform better when latent classes are relatively equal instead of unequal in nature.

In line with previous research, the BIC was found to be a top performer among the ICs (Jedidi, Jagpal, & DeSarbo, 1997; Magidson & Vermunt, 2004; Morgan, 2012; Nylund et al., 2007a; Peugh & Fan, 2013; Roeder & Wasserman, 1997; Tein, Coxe, & Cham, 2013; Tueller & Lubke, 2010). However, in this study, the recovery rates for the BIC tended to decrease when sample size dropped down to n = 500. For this reason, this study found the ABIC to be the best performer among the ICs overall, especially when sample size was small, which is in line with previous findings from Yang (2006). Other simulation studies have also found the ABIC to be the best performer overall (Peugh & Fan, 2013; Tein, Coxe, & Cham, 2013; Tofighi & Enders, 2007).

Taken together, when sample size was large (i.e., between 1,000 and 10,000), the CAIC, BIC, and ABIC were all high performing and applied researchers should therefore trust them when fitting LCA models that resemble the conditions explored in this study. On the other hand, when sample size was smaller than n = 1,000, the ABIC should be trusted the most by applied researchers since simulation results indicated that ABIC is the least sensitive out of the ICs to a reduction in sample size. Findings of this study also suggest that the AIC should not be trusted at all in an LCA context. Lastly, results indicate that when latent classes are unordered and sample size is small, researchers should remove "poor" indicators from their models to ensure the highest performance of the ICs. Additionally, when latent classes are ordered, and sample size is small, applied researchers should leave in potentially redundant indicators in their models to ensure the highest performance of the ICs. When taking all of the conditions explored in this simulation study into account, results seem to indicate that the ABIC is the best performer among the ICs overall and should therefore be trusted the most in practice.

5.4 Likelihood Ratio Tests and Practical Recommendations

In line with previous findings, the BLRT consistently outperformed the adjusted LMR-LRT (Nylund et al., 2007a; Tein, Coxe, & Cham, 2013) in the current study. Contrary to this finding, a previous study explored a range of LPA models and found that the adjusted LMR-LRT performed extremely well in conditions where the latent classes were well separated (Lubke & Muthén, 2007). Therefore, perhaps all conditions explored in this study needed greater between-class separation in order for recovery rates of the adjusted LMR-LRT to increase. Results also indicated that when the adjusted LMR-LRT index failed, it displayed a tendency to overestimate the number of latent classes. These findings are in line with previous simulation studies that compared the adjusted LMR-LRT and BLRT indices (Nylund et al., 2007a). This study also found the BLRT to be one of the least sensitive to smaller sample sizes (see Figures 14, 24, and 34).

In fact, results indicated that the BLRT could be trusted regardless of the nature of the latent classes, the number of indicators included in the model, or the relative size of the latent classes if sample size was at least $n = 500$. As sample size decreased even further (i.e., $n = 200$), the BLRT seemed to be influenced more in certain modeling contexts by the number of indicators included in the model and the relative size of the latent classes. Specifically, when latent classes were unordered and sample size was $n = 200$, the BLRT performed best when classes were relatively equal in nature (see Figure 24). When sample size was $n = 200$ and latent classes were unordered and unequal, the BLRT performed best when "poor" indicators were removed from the model (see Figure 24). On the other hand, when sample size was *n* = 200 and latent classes were ordered and unequal, the BLRT performed best when potentially redundant indicators were left in the model (see Figure 34). Taken together, this implies that researchers should generally trust the BLRT more than the adjusted LMR-LRT in applied LCA research. It is important to note that Mplus is the only software package that currently includes the BLRT index in statistical output. Additionally, when sample size is small (i.e., approximately *n* = 200), researchers should consider removing "poor" indicators in unordered latent class solutions and leaving in all available indicators in ordered latent class solutions to ensure the highest performance of the likelihood ratio tests.

5.5 Bayesian Fit Indices and Practical Recommendations

No previous research has examined the performance of the cmP and BF indices for use in class enumeration with LCA models. Findings from this study indicated that the cmP and BF tended to agree on the true latent class model across all conditions considered, especially when sample size was large (see Tables 4-6). Specifically, when sample size was between *n* = 1,000 and *n* = 10,000 there was evidence that both the cmP and BF performed extremely well regardless of the nature of the latent classes, the number of indicators included in the model, or the relative size of the latent classes. As sample size decreased, the nature of the latent classes, the number of indicators included, and the relative size of the latent classes influenced recovery rates more for the cmP and BF. In fact, when sample size dropped below *n* = 1,000, the BF performed better than the cmP (see Tables 4-6). Specifically, when sample size was equal to n = 500 and n = 200, recovery rates for the BF were generally higher than recovery rates for the cmP. Moreover, when sample size decreased to n = 500 and classes were unordered in nature, the cmP and BF tended to perform better when classes were equal in size compared to unequal, and when "poor" indicators were removed (see Figures 25 and 26). On the other hand, when sample size was n = 500, and classes were ordered in nature, the cmP and BF performed best when potentially redundant indicators were left in the model, regardless of class size (see Figures 35 and 36). When sample size decreased even further to n = 200 and classes were unordered, the cmP and BF performed best when classes were equal and "poor" indicators were eliminated (see Figures 15 and 16). Finally, when sample size was n = 200 and classes were ordered, the cmP and BF performed best when classes were equal and all available indicators were included in the model (see Figures 35 and 36).

Taken together, recovery rates for the cmP and BF indices suggest that researchers using LCA with small samples should trust the BF more than the cmP. However, in most cases, these indices should agree on the same final latent class solution. The BF tends to indicate the correct latent class solution as a higher rate than the cmP. Also, not surprisingly, given that the cmP and BF are derivatives of the BIC, recovery rates for the cmP, BF, and BIC were consistent with each other across all conditions explored. Therefore, as suspected, researchers are likely to see high agreement between the cmP, BF, and BIC in terms of class enumeration. Lastly, similar to all the fit indices considered in this simulation study, applied researchers should remove all “poor” indicators from unordered latent class solutions, and leave in all available indicators (even if they are potentially redundant) in ordered latent class solutions to ensure highest performance of the Bayesian indices.

5.6 Key Findings and Recommendations

For summary sake, the following set of general recommendations can be made based on the results of this dissertation. First, as sample size decreased, fit index performance tended to decrease across all conditions considered. Nonetheless, all fit indices considered (except the AIC and adjusted LMR-LRT) performed well (i.e., reached or nearly reached a 95% recovery rate) when at least 1,000 observations were included in the analysis. Based on this finding, it is recommended that researchers wanting to use LCA should ideally aim to for a sample size of n = 1000 to ensure high fit index performance.

Second, as analysis sample size decreased, the number of indicators included in the model became more important. Specifically, when class size was held constant, and when latent classes were *unordered*, removing “poor” indicators increased the likelihood of recovering the correct latent class solution when sample size was small. Based on this finding, it is recommended that researchers remove indicators that have low between class separation and low within-class homogeneity during the model fitting process with unordered latent class solutions. On the other hand, when class size was held constant, and the latent classes were *ordered*, keeping all potential indicators in the model (regardless of whether or not they were potentially redundant or contributed to distinguishing latent classes) increased the likelihood of recovering the correct latent class solution when sample size was small. Therefore, based on this finding, it is recommended that researchers include all available indicators when estimating ordered LCA models (as long as the model is identified), even if the indicators are potentially redundant.

Third, when sample size was large (i.e., between 1,000 and 10,000), the CAIC, BIC, and ABIC were all high performing and should therefore be trusted in practice when the latent classes resemble the conditions explored in this study. On the other hand, when sample size is smaller than *n* = 1,000, the ABIC should be trusted most since results of this study indicated that ABIC is the least sensitive out of the ICs to a reduction in sample size.

Fourth, findings from this study suggest that researchers should generally trust the BLRT more than the adjusted LMR-LRT in applied LCA research. Specifically, simulation results indicated that the BLRT could be trusted regardless of the nature of the latent classes, the number of indicators included in the model, or the relative size of the latent classes if sample size was at least *n* = 500. Moreover, if sample size is close to *n* = 200 and latent classes are unordered, it is recommended that researchers remove "poor" indicators to ensure the highest performance of the BLRT. On the other hand, if sample size is close to n = 200 and latent classes are ordered, it is recommended that researchers leave all available indicators in the LCA model to ensure the highest performance of the BLRT.

Next, it is recommended that researchers using LCA with small samples trust the BF more than the cmP. It is suspected that in most cases, the BF and cmP will agree on the same final latent class solution. However, results from this study imply that the BF tends to indicate the correct latent class solution as a higher rate than the cmP. Additionally, since the cmP and BF are derivatives of the BIC, researchers are likely to see high agreement between the cmP, BF, and BIC during the class enumeration process.

Lastly, due to the fact that this study used simulated data, the class enumeration process in this study was solely guided by statistical fit. However, in practice, it is recommended that researchers also allow substantive theory to heavily guide the LCA model fitting process. Although, it is important to understand how various fit indices perform when deciding the correct latent class solution, substantive theory should be used as a deciding factor (Muthén, 2003). The existing simulation literature focusing on mixture models is starting to find some general trends in the performance of these commonly used fit indices, however, we still do not fully know which ones work perfectly in every context. Therefore, fit indices should be used as a tool to help decide on the number of classes to retain, but they shouldn't be used blindly and without incorporating substantive theory. Based on the findings of this study, as well as previous simulation studies, a table summarizing the practical recommendations researchers should keep in mind when fitting LCA models was created and can be found in Appendix I.

5.7 Limitations and Future Directions

As with any Monte Carlo simulation study, the findings of this study are only generalizable to the LCA conditions generated in this study. Therefore, even though the conditions explored in this study were created based on an investigation of existing empirical research, the findings may not be directly applicable to all LCA research studies. This study found some sensitivity in the performance of the fit indices when sample size was n = 500 and even more sensitivity when sample size was n = 200. Therefore, more research needs to look at samples between n = 200 and n = 500 to further understand the optimal sample size needed for the fit indices to operate well. This is especially important since smaller samples tend to be used frequently in practice. In fact, the aforementioned meta-analysis of empirical LCA research conducted in this dissertation (refer back to Chapter 3.2 for more detail) found that approximately 25% of the existing LCA research used samples of n = 500 and below.

Also, the current simulation study only considered categorical indicators. While it is suspected that the results of this study generalize to a continuous setting (i.e., LPA models), future simulation research should also consider varying the class specific parameters that are estimated when latent classes are based on continuous indicators (i.e., means and variances). Moreover, the current simulation study only varied the number of indicators to be equal to 7 or 10. Therefore, future studies should vary the number of indicators more so that applied researchers could fully understand how the number of indicators included in the model influences class enumeration. All LCA models in this study were unconditional in nature (i.e., did not include any covariates or distal outcomes) therefore, future research should examine how these fit indices operate among conditional LCA models. Additionally, specific parameter estimates (i.e., conditional item probabilities) were not directly examined in this study therefore future research could examine the influence of sample size and number of indicators included on estimated parameter bias in LCA models.

Furthermore, LCA modeling is an exploratory process therefore the results of this dissertation may overly simplify a rather complex model fitting procedure. Specifically, researchers often go through an iterative process of fitting LCA models, potentially removing or changing indicators for various reasons, and refitting models. In fact, the LCA modeling process is likely different for every researcher. Therefore, given that there is no fixed procedure for fitting LCA models, this simulation study by definition does not fully mimic the true modeling process.

Next, all models in this study were generated from latent class analysis models, thus the data was generated to have latent subgroups. However, researchers in applied settings have an additional challenge of deciding if a latent class model is even appropriate for the data they have. Therefore, more research could focus on what happens to the performance of the fit indices when data is generated to have no true underlying latent class structure but then analyzed using latent class analysis. In other words, would there be any indication via the fit indices that LCA is the incorrect model when analyzing data that was generated by a factor model? One critique of LCA models is related to ordered latent class solutions, which were found in approximately 25% of the studies included in the meta-analysis in Chapter 3.2. Some would argue that ordered latent class solutions are not appropriate because perhaps latent classes are being forced onto a truly continuous distribution. Future research can address this critique by generating data as a one-factor model, analyzing it with latent class analysis, and examining the performance of the fit indices in detecting the correct underlying structure of the data. In general, although latent class analysis has its limitations (as with any other statistical model), it should be viewed as one type of statistical tool that can help researchers gain perspective into their data. Researchers should still keep in mind that there might be other alternative models that can provide equally accurate perspectives and potentially different insights into their data.

Latent class analysis has provided a lot of insight into social science phenomena and is a commonly used research tool. This study is part of a larger body of work that examines the use and best practices of mixture models. Specifically, this dissertation begins to answer some questions about how various model characteristics can influence the class enumeration process in LCA models however many unanswered questions still exist. Nevertheless, the results of this dissertation contribute to a broader understanding of how various factors such as sample size, and the number of indicators included intersect and influence class enumeration in LCA models by providing insight into the performance of commonly used fit indices. Additionally, results of this study highlight the importance of taking other factors into account such as class prevalence and the nature of the latent classes (i.e., ordered vs. unordered classes) when making methodological decisions during the LCA modeling process.

References

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

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

Bauer, D. J., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods, 8,* 384-393.

Bettencourt, A. F., & Farrell, A. D. (2013). Individual and contextual factors associated with patterns of aggression and peer victimization during middle school. *Journal of Youth and Adolescence*, *42,* 285-302.

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

Cavanaugh, C. E., Messing, J. T., Petras, H., Fowler, B., La Flair, L., Kub, J., Agnew, J., Fitzgerald, S., Bolyard, R., & Campbell, J. C. (2012). Patterns of violence against women: A latent class analysis. *Psychological Trauma: Theory, Research, Practice and Policy*, *4,* 169-176.

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

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-360). New York: Plenum Press.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences.* Hoboken, NJ: John Wiley & Sons, Inc.

Eddy, K. T., Le Grange, D., Crosby, R. D., Hoste, R. R., Doyle, A. C., Smyth, A., & Herzog, D. B. (2010). Diagnostic classification of eating disorders in children and adolescents: How does the DSM-IV-TR compare to empirically derived categories? *Journal of the American Academy of Child and Adolescent Psychiatry, 49*, 277-293.

Elia, J., Arcos-Burgos, M., Bolton, K. L., Ambrosini, P. J., Berrettini, W., & Muenke, M. (2009). ADHD latent class clusters: DSM-IV subtypes and comorbidity. *Psychiatry Research, 170*, 192-198.

Grant, J. D., Scherrer, J. F., Neuman, R. J., Todorov, A. A., Price, R. K., & Bucholz, K. K. (2006). A comparison of the latent class structure of cannabis problems among adult men and women who have used cannabis repeatedly. *Addiction, 101*, 1133-1142.

Heinen, T. (1996). *Latent class and discrete latent trait models*. Thousand Oaks, CA: Sage Publications.

Jedidi, K., Jagpal, H., & DeSarbo, W. S. (1997). Finite mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16, 39-59.*

Koehler, A. B., & Murphree, E. H. (1998). A comparison of the Akaike and Schwarz criteria for selecting model order. *Applied Statistics, 37,* 187-195.

Lazarfeld, P., & Henry, N. (1968). *Latent structure analysis*: New York: Houghton-Mifflin.

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

Lubke, G. H. & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*, 21-39.

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

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research, 41*, 499-532.

Lubke, G., & Neale, M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research, 43*, 592-620.

Magidson, J., & Vermunt, J. (2004). Latent class models. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 175-198). Newbury Park, CA: Sage.

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods: Statistical analyses* (Vol. 2, pp. 551-611). New York, NY: Oxford University Press.

McCutcheon, A. C. (1987). *Latent class analysis*. Beverly Hills, CA: Sage.

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

Morgan, G. (2012). Mixed mode latent class clustering: An examination of fit index performance for identifying latent classes. (Unpublished doctoral dissertation). University of South Carolina, Columbia.

Muthén, B., & Muthén, L. K. (1998-2013). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882-891.

Muthén, L., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 4*, 599-620.

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

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*, 139-157.

Nylund, K., Asparouhov, T., & Muthén, B. O. (2007a). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 553-569.

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

Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo Simulation. *Structural Equation Modeling: A Multidisciplinary Journal, 20*, 616-639.

Roeder, K., & Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association, 92*, 894-902.

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

Shin, S. H., Hong, H. G., & Hazen, A. L. (2010). Childhood sexual abuse and adolescent substance abuse: A latent class analysis. *Drug and Alcohol Dependence*, *109*, 226-235.

StataCorp. (2013). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

Stormont, M., Herman, K. C., Reinke, W. M., David, K. B., & Goel, N. (2013). Latent profile analysis of teacher perceptions of parent contact and comfort. *School Psychology Quarterly, 28*, 195-209.

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 20*, 640-657.

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.

Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling, 17*, 165-192.

Von Stumm, S., Chung, A., & Furnham, A. (2011). Creative ability, creative ideation, and latent classes of creative achievement: What is the role of personality? *Psychology of Aesthetics, Creativity, and the Arts, 5*, 107-114.

Wasserman, J. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44*, 92-107.

Woodroofe, M. (1983). On model selection and the arc sine laws. *The Annals of Statistics, 10*, 1182-1194.

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

Appendix A: Sample Mplus data generation input file for a four ordered class model with sample size of 10,000, 10 binary indicators, and class prevalences of .10 (π_1), .20 (π_2), .20 (π_3), and .50 (π_4) (see Figure 2 for item probablity plot of this model).

```
Title: Example Input File #1

Montecarlo:
        Names are u1-u10;
        Generate = u1-u10 (1);
        Categorical=u1-u10;
        Genclasses = c (4);
        Classes=c (1);
        Nobs=10000;
        Nrep=500;
        Repsave=ALL;
        SAVE=4class_unequal_ord*.DAT;
        Seed=46304630;

Model Population: (See Appendix H for how to solve for latent class proportions)

%OVERALL%
 C1=10%, C2=20%, C3=20%, C4=50%
 [c#1*-1.6094379124];
 [c#2*-0.9162907319];
 [c#3*-0.9162907319];

FOUR ORDERED LATENT CLASSES
%c#1% class 1: item probs of [.90, .92, .93, .95, .93, .90, .91, .92, .90, .93] for all U1-U10
 [u1$1*-2.20 u2$1*-2.44 u3$1*-2.59 u4$1*-2.94 u5$1*-2.59
 u6$1*-2.20 u7$1*-2.31 u8$1*-2.44 u9$1*-2.20 u10$1*-2.59];

%c#2% class 2: item probs of [.70, .75, .73, .76, .74, .69, .73, .72, .70, .72] for all U1-U10
 [u1$1*-.85 u2$1*-1.10 u3$1*-.99 u4$1*-1.15 u5$1*-1.05
 u6$1*-.80 u7$1*-.99 u8$1*-.95 u9$1*-.85 u10$1*-.95];

%c#3% class 3: item probs of [.40, .39, .42, .40, .39, .44, .43, .41, .42, .40] for all U1-U10
 [u1$1*.41 u2$1*.45 u3$1*.32 u4$1*.41 u5$1*.45
 u6$1*.24 u7$1*.28 u8$1*.36 u9$1*.32 u10$1*.41];

%c#4% class 4: item probs of [.07, .08, .07, .09, .07, .1, .1, .08, .09, .07] for all U1-U10
 [u1$1*2.59 u2$1*2.44 u3$1*2.59 u4$1*2.31 u5$1*2.59
 u6$1*2.20 u7$1*2.20 u8$1*2.44 u9$1*2.31 u10$1*2.59];

Analysis:
        Type=mixture;
        Starts=100 20;
        Processor=4 (starts);
```

Appendix B: Sample Mplus data generation input file for a four class model (unordered A) with sample size of 10,000, 10 binary indicators, and class prevalences of .10 (π_1), .20 (π_2), .20 (π_3), and .50 (π_4) (see Figure 3 for item probablity plot of this model).

```
Title: Example Input File #2

Montecarlo:
        Names are u1-u10;
        Generate = u1-u10 (1);
        Categorical=u1-u10;
        Genclasses = c (4);
        Classes=c (1);
        Nobs=10000;
        Nrep=500;
        Repsave=ALL;
        SAVE=4class_unequal_unord1*.DAT;
        Seed=46304630;

Model Population:

%OVERALL%
 C1=10%, C2=20%, C3=20%, C4=50%
 [c#1*-1.6094379124];
 [c#2*-0.9162907319];
 [c#3*-0.9162907319];

FOUR UNORDERED LATENT CLASSES A
%c#1% class 1: item probs of [.82, .83, .65, .85, .82, .84, .69, .83, .83, .67] for all U1-U10
 [u1$1*-1.52 u2$1*-1.59 u3$1*-.62 u4$1*-1.73 u5$1*-1.52
  u6$1*-1.66 u7$1*-.80 u8$1*-1.59 u9$1*-1.59 u10$1*-.71];

%c#2% class 2: item probs of [.80, .80, .55, .80, .80, .10, .50, .10, .10, .40] for all U1-U10
 [u1$1*-1.39 u2$1*-1.39 u3$1*-.21 u4$1*-1.39 u5$1*-1.39
 u6$1*2.20 u7$1*-.00 u8$1*2.20 u9$1*2.20 u10$1*.41];

%c#3% class 3: item probs of [.10, .10, .40, .10, .10, .80, .55, .80, .80, .55] for all U1-U10
 [u1$1*2.20 u2$1*2.20 u3$1*.41 u4$1*2.20 u5$1*2.20
 u6$1*-1.39 u7$1*-.21 u8$1*-1.39 u9$1*-1.39 u10$1*-.21];

%c#4% class 4: item probs of [.07, .08, .30, .09, .07, .1, .33, .08, .09, .35] for all U1-U10
 [u1$1*2.59 u2$1*2.44 u3$1*.85 u4$1*2.31 u5$1*2.59
 u6$1*2.20 u7$1*.71 u8$1*2.44 u9$1*2.31 u10$1*.62];

Analysis:
        Type=mixture;
        Starts=100 20;
        Processor=4 (starts);
```

Appendix C: Sample Mplus data generation input file for a four class model (unordered B) with sample size of 10,000, 10 binary indicators, and class prevalences of .10 (π_1), .20 (π_2), .20 (π_3), and .50 (π_4) (see Figure 4 for item probablity plot of this model).

```
Title: Example Input File #3

Montecarlo:
        Names are u1-u10;
        Generate = u1-u10 (1);
        Categorical=u1-u10;
        Genclasses = c (4);
        Classes=c (1);
        Nobs=10000;
        Nrep=500;
        Repsave=ALL;
        SAVE=4class_unequal_unord2*.DAT;
        Seed=46304630;

Model Population:

%OVERALL%
  C1=10%, C2=20%, C3=20%, C4=50%
  [c#1*-1.6094379124];
  [c#2*-0.9162907319];
  [c#3*-0.9162907319];

FOUR UNORDERED LATENT CLASSES B
%c#1% class 1: item probs of [.60, .30, .75, .70, .30, .83, .02, .10, .20, .90] for all U1-U10
 [u1$1*-.41 u2$1*.85 u3$1*-1.10 u4$1*-.85 u5$1*.85
 u6$1*-1.59 u7$1*3.89 u8$1*2.20 u9$1*1.39 u10$1*-2.20];

%c#2% class 2: item probs of [.05, .80, .20, .35, .75, .85, .01, .05, .01, .70] for all U1-U10
 [u1$1*2.94 u2$1*-1.39 u3$1*1.39 u4$1*.62 u5$1*-1.10
 u6$1*-1.73 u7$1*4.60 u8$1*2.94 u9$1*4.60 u10$1*-.85];

%c#3% class 3: item probs of [.15, .30, .30, .20, .22, .75, .70, .73, .18, .90] for all U1-U10
 [u1$1*1.73 u2$1*.85 u3$1*.85 u4$1*1.39 u5$1*1.27
 u6$1*-1.10 u7$1*-.85 u8$1*-.99 u9$1*1.52 u10$1*-2.20];

%c#4% class 4: item probs of [.15, .10, .05, .10, .07, .78, .02, .03, .09, .80] for all U1-U10
 [u1$1*1.73 u2$1*2.20 u3$1*2.94 u4$1*2.20 u5$1*2.59
 u6$1*-1.27 u7$1*3.89 u8$1*3.48 u9$1*2.31 u10$1*-1.39];

Analysis:
        Type=mixture;
        Starts=100 20;
        Processor=4 (starts);
```

Appendix D: Sample Mplus data generation input file for a three ordered class model with sample size of 10,000, 10 binary indicators, and class prevalences of .10 (π_1), .30 (π_2), and .60 (π_3) (see Figure 5 for item probablity plot of this model).

```
Title: Example Input File #4

Montecarlo:
        Names are u1-u10;
        Generate = u1-u10 (1);
        Categorical=u1-u10;
        Genclasses = c (3);
        Classes=c (1);
        Nobs=10000;
        Nrep=500;
        Repsave=ALL;
        SAVE=3class_unequal_ord_*.DAT;
        Seed=46304630;

Model Population:

%OVERALL%
  C1=10%, C2=30%, C3=60%
  [c#1*-1.7957674906];
  [c#2*-0.6931471806];

THREE ORDERED LATENT CLASSES
%c#1% class 1: item probs of [.90, .92, .93, .95, .93, .90, .91, .92, .90, .93] for all U1-U10
 [u1$1*-2.20 u2$1*-2.44 u3$1*-2.59 u4$1*-2.94 u5$1*-2.59
 u6$1*-2.20 u7$1*-2.31 u8$1*-2.44 u9$1*-2.20 u10$1*-2.59];

%c#2% class 3: item probs of [.50, .49, .52, .50, .49, .54, .53, .51, .52, .50] for all U1-U10
 [u1$1*0 u2$1*.04 u3$1*-.08 u4$1*0 u5$1*.04
 u6$1*-.16 u7$1*-.12 u8$1*-.04 u9$1*-.08 u10$1*0];

%c#3% class 4: item probs of [.07, .08, .07, .09, .07, .1, .1, .08, .09, .07] for all U1-U10
 [u1$1*2.59 u2$1*2.44 u3$1*2.59 u4$1*2.31 u5$1*2.59
 u6$1*2.20 u7$1*2.20 u8$1*2.44 u9$1*2.31 u10$1*2.59];

Analysis:
        Type=mixture;
        Starts=100 20;
        Processor=4 (starts);
```

Appendix E: Sample Mplus analysis input file for population condition with four equal unordered classes A, and analysis sample size of 10,000, 10 binary indicators, and 4 latent classes.

```
Title: Analysis 4class_equal_unord1_N10000_U10

Data:
    File is 4class_equal_unord1_list.dat;
    Type=Montecarlo;

Variable:
    Names are
    U1 U2 U3 U4 U5 U6 U7 U8 U9 U10 C id;
    USEVARIABLES = U1 U2 U3 U4 U5 U6 U7 U8 U9 U10;
    CATEGORICAL = U1 U2 U3 U4 U5 U6 U7 U8 U9 U10;
    Missing are all (-9999);
    Classes=C (4);
    IDVARIABLE=id;

Analysis:
    Type = mixture;
    Starts=100 20;
    LRTSTARTS=0 0 150 40;
    Processor=16(starts);

Output:
    tech9 tech11 tech14;

Savedata:
    RESULTS=4class_equal_unord1_N10000_U10_C4.dat;
```

Appendix F: Sample Mplus analysis input file for population condition with four equal ordered classes and with analysis sample size of 5,000, 7 binary indicators, and 4 latent classes.

```
Title: Analysis 4class_equal_ord_N5000_U7

Data:
    File is 4class_equal_ord_list.dat;
    Type=Montecarlo;

Variable:
    Names are
    U1 U2 U3 U4 U5 U6 U7 U8 U9 U10 C id;
    USEVARIABLES = U2 U3 U4 U6 U7 U8 U9;
    CATEGORICAL = U2 U3 U4 U6 U7 U8 U9;
    Missing are all (-9999);
    Classes=C (4);
    IDVARIABLE=id;
    Useobs = id LE 5000;

Analysis:
    Type = mixture;
    Starts=100 20;
    LRTSTARTS=0 0 150 40;
    Processor=16 (starts);

Output:
    tech9 tech11 tech14;

 Savedata:
    RESULTS=4class_equal_ord_N5000_U7_C4.dat;
```

Appendix G: Sample Mplus analysis input file for population condition with three unequal ordered classes and with analysis sample size of 200, 7 binary indicators, and 3 latent classes.

```
Title: Analysis 3class_unequal_ord_N200_U7

Data:
    File is 3class_unequal_ord_list.dat;
    Type=Montecarlo;

Variable:
    Names are
    U1 U2 U3 U4 U5 U6 U7 U8 U9 U10 C id;
    USEVARIABLES = U2 U3 U4 U6 U7 U8 U9;
    CATEGORICAL = U2 U3 U4 U6 U7 U8 U9;
    Missing are all (-9999);
    Classes=C (3);
    IDVARIABLE=id;
    Useobs = id LE 200;

Analysis:
    Type = mixture;
    Starts=100 20;
    LRTSTARTS=0 0 150 40;
    Processor=16 (starts);

Output:
    tech9 tech11 tech14;

Savedata:
    RESULTS=3class_unequal_ord_N200_U7_C3.dat;
```

Appendix H: How to solve for the threshold values that are used to specify the latent class proportions (i.e., class sizes for C1-C4) in Appendix A.

Example: C1 = 10%, C2 = 20%, C3 = 20%, C4 = 50%

Step 1- Use the following equation:

$$Class\ Proportions = \frac{exp(\mu_C)}{\sum_{C=1}^{C} exp(\mu_C)}$$

Therefore, for 4 latent classes, the equation above expands to:

$$\frac{exp(\mu_4)}{exp(\mu_1) + exp(\mu_2) + exp(\mu_3) + exp(\mu_4)}$$

Step 2- Pick a reference class. For this example, the largest class will be the reference class (i.e., C4 = 50%).

$$P\ (C = 4) = .50 = \frac{exp(0)}{exp(\mu_1) + exp(\mu_2) + exp(\mu_3) + exp(0)} =$$

$$.50 = \frac{1}{exp(\mu_1) + exp(\mu_2) + exp(\mu_3) + 1} =$$

$$\frac{1}{2} = \frac{1}{exp(\mu_1) + exp(\mu_2) + exp(\mu_3) + 1} =$$

Step 3- Cross Multiply.

$$2 = exp(\mu_1) + exp(\mu_2) + exp(\mu_3) + 1 =$$

$$1 = exp(\mu_1) + exp(\mu_2) + exp(\mu_3)$$

Step 4- To solve for the remaining class specific logit values, plug this new value of 1 above into the original equation in Step 1.

Solving for μ_1:

$$.10 = \frac{exp(\mu_1)}{exp(\mu_1) + exp(\mu_2) + exp(\mu_3) + 1} =$$

$$.10 = \frac{exp(\mu_1)}{1 + 1} = \frac{exp(\mu_1)}{2} =$$

$$.2 = exp(\mu_1) =$$

$$\ln(.2) = \mu_1 = \mathbf{-1.6094379124}$$

Solving for μ_2:

$$.20 = \frac{exp(\mu_2)}{exp(\mu_1) + exp(\mu_2) + exp(\mu_3) + 1} =$$

$$.20 = \frac{exp(\mu_2)}{1 + 1} = \frac{exp(\mu_2)}{2} =$$

$$.4 = exp(\mu_2) =$$

$$\ln(.4) = \mu_2 = \mathbf{-0.9162907319}$$

Solving for μ_3:

$$.20 = \frac{exp(\mu_3)}{exp(\mu_1) + exp(\mu_2) + exp(\mu_3) + 1} =$$

$$.20 = \frac{exp(\mu_3)}{1 + 1} = \frac{exp(\mu_3)}{2} =$$

$$.4 = exp(\mu_3) =$$

$$\ln(.4) = \mu_3 = \mathbf{-0.9162907319}$$

Appendix I: Summary of findings, recommendations for researchers, and directions for future research.

LCA Model Details	Finding from Current Study	Recommendation for Researchers Based on Finding from Current Study	Previous Studies that Validate Current Findings	Directions for Future Research
All LCA models across all conditions explored in this study	AIC consistently underestimated the correct number of latent classes	Researchers should not trust the AIC in practice	Celeux & Soromenho, 1996; Koehler & Murphree, 1988; Nylund, Asparouhov, & Muthén, 2007; Tein, Coxe, & Cham, 2013	N/A; existing literature has reached a consensus regarding the poor performance of the AIC in mixture models.
	BLRT outperformed the adjusted LMR-LRT	Researchers should trust the BLRT over the adjusted LMR-LRT.	Nylund et al., 2007a; Tein, Coxe, & Cham, 2013	Future research should further examine how the adjusted LMR-LRT is influenced by class separation
All LCA conditions explored in this study with a sample size ≥ n = 1,000	When sample size was at least 1,000, fit indices were generally high performing (with the exception of the AIC and LMR-LRT) across all conditions considered	Researchers can trust all fit indices (except the AIC and LMR-LRT) when sample size is at least 1,000 and conditions resemble those explored in this study	N/A; The specific combination of fit indices and the conditions explored in this study are not identical to previous studies	Future research can extend upon these findings and attempt to validate them by exploring conditional LCA models as well as LPA models
All LCA conditions explored in this study, especially those with a sample size < n = 1,000	As sample size decreased, fit index performance decreased	LCA researchers should aim for a sample size of n = 1,000	Nylund et al., 2007a; Yang, 2006; Tein, Coxe, & Cham, 2013; Tueller & Lubke, 2010	Future research should explore sample sizes between 500-1,000, between 200-500, and < 200
	ABIC and BLRT were the least sensitive out of the ICs to a reduction in sample size.	Researchers should trust ABIC and BLRT most if sample size is below 1,000	Nylund et al., 2007a; Peugh & Fan, 2013; Tein, Coxe, & Cham, 2013; Tofighi & Enders, 2007	Future research should explore if the ABIC and BLRT remain high performing in conditional LCA models
	As sample size decreased, the number of indicators included became more important. Specifically,	In order to ensure the highest fit index performance, researchers should remove indicators	N/A; no previous simulation study has compared ordered vs. unordered latent classes	Future research should look at other configurations of ordered and unordered latent class solutions.

when class size was held constant, and when latent classes were *unordered*, removing "poor" indicators increased the likelihood of recovering the correct latent class solution. When class size was held constant, and the latent classes were *ordered*, keeping all potential indicators in the model increased the likelihood of recovering the correct latent class solution.	that have low between class separation and low within-class homogeneity during the model fitting process with unordered latent class solutions. Also, researchers should include all available indicators when estimating ordered LCA models, even if the indicators are potentially redundant.		Future research should also consider varying the number of indicators to be equal to values other than 7 and 10. Lastly, future research should consider generating data with no underlying latent classes (i.e., a one-factor model) and analyze it using LCA.
The BF outperformed the cmP, especially when sample size was small	Researchers using LCA with small samples trust the BF more than the cmP	N/A; no other simulation study has previously examined the performance of the BF and cmP among LCA models	Future research should attempt to validate these findings of the BF and cmP with other mixture models and across other empirical conditions