Title
A Monte Carlo Simulation Study Examining Statistical Power in Latent Transition Analysis

Author
Baldwin, Erika Elaine

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

A Monte Carlo Simulation Study Examining Statistical Power in Latent Transition Analysis

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Education

by

Erika E. Baldwin

Dissertation Committee:

Professor Karen Nylund-Gibson

Professor John Yun

Professor Nancy Collins

June 2015

The dissertation of Erika E. Baldwin is approved.

John T. Yun

Nancy L. Collins

Karen Nylund-Gibson, Committee Chair

April 2015

A Monte Carlo Simulation Study Examining Statistical Power in Latent Transition Analysis

by

Erika E. Baldwin

For my grandfather, Hideaki Kato

Thank you for inspiring me, even in my dreams.

ACKNOWLEDGMENTS

Above and beyond, I wish to thank my parents, John and Keiko Baldwin. Your unconditional love, support, and pride have been my greatest motivation. Thank you for always believing in me and pushing me to reach my dreams. Okaasan, I am finally done! Thank you for teaching me patience and all of your sewer's secrets. Dad, you have always been my best example of a hard worker. Thank you for your love and for paying for my cell phone bill! I will always be your baby, Bubbs—but now I'll be your Dr. Bubbs!

Thank you Miyoko, for not only being my older sister but also for being my best friend. Your simple phone calls and silly texts helped ease some of the most difficult days in grad school. I love you! Thank you Mike for being such a great brother-in-law, and thank you both for giving me two handsome and goofy nephews. They bring a smile to my face every day! I am so lucky to have such an amazing, loving family!

Many individuals have supported me professionally along the way, and for that, I am eternally grateful. During my time at UC Santa Barbara, I must thank Patricia Marin. You have taught me so much about professionalism and research. I will never, ever forget to remove tracked changes before sending an email! Laila Diguilio (LD), what would I do without you in North Hall?! I am so thankful for your jokes, laughs, and good times, not to mention your support, advice and leadership. Edwin Hunt (Big E), your advice and stories have seasoned this long trek. To Pamela Scott, for being a great boss and even greater example of a strong, independent woman. At Hawaii Pacific University, Drs. Valentina Abordonado, Linda Wheeler, and Edwin Van Gorder: You believed in me from the start of my journey to becoming an educator. To my high school teachers Mr. Herman Leong and Mrs. Maryann Kurose. Mr. Leong, you encouraged me as a young girl to pursue my strengths in mathematics. Mrs. Kurose, you guided me and showed me what it meant to be an amazing teacher (and corrected my grammar along the way).

A very special shout-out to my friends, many of whom I am also lucky to call my colleagues. To everyone at the UCEC: thank you for the grooviest memories in graduate school! A very special thank-you goes to Mark Grimes, Veronica Fematt and Ryan Grimm. You three have been my support since day one, in every aspect of my life. I could not have persevered through the hardest points (both academically and personally) without you. To everyone in the Latent Variable Group, thank you for listening to me go on and on about my results every week. Who needs coffee when you have stats, amirite?! I wish to thank my best friends since high school, Christina Maximo and Breanne Harris. I am so lucky to have such beautiful and intelligent ladies in my life. And of course, thank you to Behzad Anbarani: you already know… I love you. I mean it! Your love and support means the world to me. <3

I think I forgot someone… oh wait—my committee! Drs. John Yun, Karen Nylund-Gibson, and Nancy Collins! Thank you all for your support, your inspiration, and the hours and hours you spent meeting, reading and revising my work. John, thank you for giving me a chance, for taking me under your wing, for being my mentor, for being my advisor, for being a funny advisor, for funding me, for feeding me...the list is endless! Karen, thank you for adopting me as your academic child, especially when you had twins and a four-year old of your own! I will always admire your stats wisdom and knowledge. Cool show! Nancy, thank you for being my QMSS advisor and for your constant enthusiasm and encouragement. I would also like to thank Dr. Maria Charles for serving on my qualifying examinations committee.

Thank you, UC Santa Barbara, for five great years. Olé!

VITA OF ERIKA E. BALDWIN
June 2015

EDUCATION

June 2015 Doctor of Philosophy in Education, University of California, Santa Barbara

March 2014 Master of Arts in Education, University of California, Santa Barbara

Aug. 2009 Master of Education in Secondary Education, Hawaii Pacific University

Dec. 2007 Bachelor of Arts in Mathematics, Hawaii Pacific University

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

EMPLOYMENT

2015-Present Senior Manager of Data Analytics, California State University Office of the Chancellor

2014-2015 Graduate Student Researcher, Office of Institutional Research, University of California, Santa Barbara

2014-2015 Graduate Student Association Representative, Chancellor's Outreach Advisory Board, University of California, Santa Barbara

2009-2015 Graduate Student Researcher, University of California Educational Evaluation Center

2011-2013 Graduate Student Apartments Resident Coordinator, Residential and Community Living, University of California, Santa Barbara

2011-2012 Instructor, Freshman Summer Start Program, University of California, Santa Barbara

2010-2010 Tutorial Services Assistant II, Mount San Antonio College

2008-2009 Middle School Mathematics Teacher, Pearl Harbor Christian Academy

2008-2008 High School Mathematics Teacher, Hawaii Department of Education

TEACHING CREDENTIALS

Hawai`i Teacher Standards Board Teaching License: Mathematics 7 – 12
California Single Subject Teaching Credential: Middle and High School Mathematics

CONFERENCE PRESENTATIONS

Baldwin, E. E. (October 2013). Examining Local Level Surveys for Larger Scale Evaluative Use: A Factor Analysis of a Student Survey on Mathematics Attitudes. Poster presentation at the annual meeting of the American Evaluation Association, Washington, DC.

Baldwin, E. E. (April 2013). Gender and Eighth Grade Algebra on Mathematics Achievement and Self-Concept: A Parallel Processes Growth Model. Paper presentation at the annual meeting of the American Educational Research Association, San Francisco, CA.

Baldwin, E. E., & Yun, J. T. (January 2012). Mathematics curricula and formative assessments: Toward an error-based approach to formative data use in mathematics. Paper presentation at the annual conference of the Hawai`i International Conference on Education, Honolulu, HI.

AWARDS & RECOGNITION

Selected Participant, AERA Institute on Statistical Analysis: Mathematics Education and Equity

Graduate Student Scholar, Institute for Training in Educational Evaluation, University of California Educational Evaluation Center

Mosaic–UCEC Innovative Evaluation Research Award, University of California Educational Evaluation Center

Department of Education Recruitment Fellowship, Gevirtz Graduate School of Education, University of California, Santa Barbara

Recognition of Excellence for Mathematics Content Knowledge Test, The Praxis Series

Aloha Chapter College Scholarship, Military Officers Association of America

President's Scholarship Recipient, Hawaii Pacific University

PROFESSIONAL ASSOCIATIONS

The Latent Variable Group, UC Santa Barbara

Women in Science and Engineering (WiSE), UC Santa Barbara

American Educational Research Association (AERA)

American Evaluation Association (AEA)

Society for Research on Educational Effectiveness (SREE)

Pi Lambda Theta Honor Society (Education)

Kappa Mu Epsilon Honor Society (Mathematics)

Alpha Chi Honor Society (Top 10% of college juniors and seniors)

Pi Gamma Mu Honor Society (Social sciences)

Phi Sigma Honor Society (Scholastic undergraduates)

ABSTRACT

A Monte Carlo Simulation Study Examining Statistical Power in Latent Transition Analysis

by

Erika E. Baldwin

Latent transition analysis (LTA) is a mixture modeling approach that is gaining popularity in social science, behavioral, and health research. LTA is a longitudinal method that can be used to investigate how individuals transition from one latent, or unobserved class, to another over time. Although LTA is gaining use in many disciplines, to date only two studies have examined the statistical power of this statistical approach. The present study aims to examine how sample size and model characteristics such as latent transition probabilities, model definition, item-response probabilities, and class size influence the statistical power of to detect effects in latent transition probabilities. Meta-analysis findings were used to guide conditions ultimately used in this Monte Carlo simulation study. All data were generated using Mplus (Muthén & Muthén, 1998-2014).

Results from this study revealed how larger sample sizes, larger transition probabilities and class sizes were more likely to have greater power. Results also highlighted the importance of a well-defined measurement model with high class separation and homogeneous classes and its influence on statistical power. Findings from this dissertation provide evidence on which conditions tend to have higher or lower power. Additionally, findings show how poor conditions can have model convergence issues and provide misleading results due to “artificially high” power values. This study also includes practical recommendations and suggestions for future directions.

TABLE OF CONTENTS

Chapter 1 Introduction

1.1 Overview of Mixture Modeling and Latent Transition Analysis (LTA)

Latent transition analysis (LTA) is a mixture modeling approach that can be used to examine the transition between existing heterogeneous subgroups within a homogeneous population (Nylund, Asparouhov, & Muthén, 2007). LTA is one approach among a set of methods termed mixture modeling. In mixture modeling, the overall distribution of one or more variables is composed of a mixture of a finite number of sub-distributions (Masyn, 2013). The key assumption in mixture modeling is that there is an underlying latent variable that divides the population into two or more mutually exclusive groups called latent classes (Collins & Lanza, 2010). Latent variables are unobservable constructs that are measured by observable variables, which are also called indicators. In psychology, constructs such as extraversion and self-image are latent variables. In education, researchers examine unobservable traits such as academic engagement and persistence. In economic research, quality of life is a latent variable, as it cannot be directly measured. Latent variables such as these are inferred from a number of other observable variables. For example, educational engagement can be measured through responses to survey questions such as “I am easily distracted when I study” or “I am enthusiastic about my studying.”

Mixture modeling has become increasingly popular in social, behavioral, and health sciences, as it allows researchers to examine typologies among individuals. This modeling technique has been used in a wide range of different applications, from making more accurate myocardial infarction diagnoses (Rindskopf & Rindskopf, 1986) to finding different classes of heavy drinking patterns among young adults (Lanza & Collins, 2006). There are a variety of models that fall under the umbrella of mixture modeling techniques, including latent class analysis (LCA), latent profile analysis (LPA), growth mixture modeling (GMM), and latent transition analysis (LTA). The present study will focus on LTA.

LTA is a longitudinal extension of LCA. LCA is a quantitative approach that examines whether there exists unobservable groups, or classes, within a population, whereas LTA examines qualitative changes in latent class membership over time. LTA is a particularly important method of analysis, as results can be substantively used to directly treat individuals based on the class membership and understand experiences of individuals in each latent class, among other uses. LTA has been used in many important studies, including the following recently published studies:

- Discovering eating disorder phenotypes (Castellini et al., 2013)
- Determining the relationship between parenting type and adolescent drinking behavior (Abar, 2012)
- Examining changes in reading classification after an intervention (Catts, Tomblin, Compton, & Bridges, 2012)
- Understanding comorbidity among anxiety and depressive disorders (Spinhoven, de Rooij, Heiser, Willem, & Penninx, 2012)
- Modeling transitions to and from alcohol abuse and sexual activity among freshman college students (Palen, Smith, Caldwell, Mathews, & Vergnani, 2009)

LTA models class membership changes between one time point and another time point. It is specifically used to study the probability of an individual transitioning from one time point to another or from one state to another. This is dissimilar to the more conventional longitudinal method, latent growth curve modeling, which examines the rate at which some process changes over continuous time. LTA, on the other hand, approximates latent class membership at time $t + 1$, conditional on an individual's latent class membership at time t. This type of model is a called a first order Markov chain model because the distribution of a variable at time t is dependent only on the distribution of the previous state at time $t - 1$ and not dependent on any other times before that (e.g., $t - 2$, $t - 3$, etc.). For this reason, Markov models are often regarded as "memoryless" because it ignores what happened prior to $t - 1$. Markov chain models have been used in many studies, including research on consumer brand loyalty, meteorology, and voting behavior (Langeheine & van de Pol, 2002). A main assumption of Markov chain models is that change is operated over discrete time (Langeheine & van de Pol) whereas change in GCM is assessed over continuous time.

There are many advantages of using LTA over other longitudinal modeling approaches. LTA allows for estimation of measurement error and the use of multiple indicators (Velicer, Martin, & Collins, 1996). LTA also has the ability to model change in a discrete manner and provides an easier way to examine large contingency tables (Lanza & Collins, 2008). According to Velicer et al., LTA can be used to answer a number of research questions, including:

- How does LTA compare to other theoretical models that look at change over time?
- Does there exist treatment effects for different groups?
- How do different measures contribute to each latent status?
- What is the distribution of participants by latent status at each time point?

Despite the advantages and the extent to which LTA can help answer research questions, little research has been conducted to examine how this statistical method operates. Researchers often rely on thresholds and rules-of-thumb when applying statistical methods such as structural equation modeling or covariance structure models. These types of studies do not yet exist for LTA. Additionally, many studies have examined the performance of model fit indices in these models. For example, Sharma, Mukherjee, Kumar, and Dillon (2005) conducted a study to examine covariance structure modeling, where they assessed the effect of sample size, factor loadings, factor correlations, and number of indicators on whether both true and misspecified models were accepted or rejected based on goodness-of-fit cut-off values. Hu and Bentler (1998) also assessed fit indices in covariance structure modeling to see whether the indices were sensitive in models that were misspecified. Similarly, Beauducel and Wittmann (2005) assessed fit indices in misspecified models, yet this time in confirmatory factor analysis models. Despite the large number of studies in SEM and other latent variable models, few studies have looked at LTA models.

1.2 Monte Carlo Simulation Studies

In many cases, Monte Carlo studies are conducted to examine “best practices” and create “rules of thumb” for statistical models. Monte Carlo studies are simulation studies that are typically used to investigate the performance of statistical estimators under varying conditions. Sharma et al. (2005) and Beauducel and Wittmann (2005) both used simulations in the aforementioned studies. In Monte Carlo studies, data are generated under hypothesized modeling conditions, samples are drawn, models are estimated for each sample, and then standard errors and parameter values are averaged over these samples (Muthén & Muthén, 2002).

The advantage of using a simulation study is that the researcher has control over the conditions under which the simulation is conducted. In simulation studies, researchers can alter conditions such as sample size and number of factor indicators to compare results across models. For example, Nylund et al. (2007) used a Monte Carlo simulation to study the performance of information criterion and likelihood-based fit indices used in LCA, GMM and factor mixture models. Nylund et al. (2007) examined how indices perform under different modeling conditions to help determine the number of classes in correctly specific models. Sharma et al. (2005) and Beauducel and Wittmann (2005) used simulation studies to see how indices performed on models that purposefully had a specific misspecification. Simulation studies can also be used to replicate common measurement conditions and specifications to decide on sample size and to estimate power. For example, Myers, Ahn, & Jin (2011) used CFA model conditions commonly found in exercise and sport research to determine what minimum sample size was needed and what level of power researchers might expect under those conditions. Simulation studies such as these are beneficial to the overall field of latent variable modeling, as they help provide information such as thresholds, cut-off values, and rules of thumb. They also help determine which fit indices are the best, most consistent, and/or least sensitive.

Simulation studies are also useful in helping researchers determine statistical power. The ability to vary conditions in a simulation studies allows researchers to investigate the extent to which these various conditions affect statistical power. Power studies cans also help determine what sample size is necessary to detect adequate statistical power. If sample size guidelines are not developed, researchers run the risk of conducting studies that may not reveal significant relationships or changes between variables because the features of their design do not allow for adequate power to detect these effects. “A sample may be large enough for unbiased parameter estimates, unbiased standard errors, and good coverage, but it may not be large enough to detect an important effect in the model” (Muthén & Muthén, 2002). For example, the results of a Monte Carlo power study involving a multilevel structural equation model (Meuleman & Billiet, 2009) revealed that to detect effects greater than .50 at the between-group level, at least 60 groups are required. For adequate power to detect smaller effects, more than 100 groups are required.

Despite the strength and importance of Monte Carlo simulation studies in methodology research, very few simulation studies have been conducted under the LTA framework. In fact, the only zero LTA simulation studies have been conducted in the last 18 years. Collins and Wugalter (1992) used a simulation study to determine if adding additional indicators in a LTA model would provide better measurement or more sparse contingency tables. The study concluded that, under the imposed conditions, including more indicators improved standard errors even when the contingency tables were sparse. Collins and Tracy (1997) later conducted a similar study. Because few studies have looked at the effects of different conditions, there still remains a substantial gap of knowledge about LTA. To date, there is no known simulation study examining how varying conditions affects the statistical power of latent transition probabilities in a LTA framework.

1.2.1 Statistical power. As stated earlier, Monte Carlo simulation studies can help determine the level of statistical power in a parameter or model. In statistical hypothesis testing, there are two types of hypotheses:

1) the null hypothesis, or H_0, which states some population parameter that we assume to be true, and

2) the alternative hypothesis, or H_1, which is contradictory to the null hypothesis and which we test against the null hypothesis.

Statistical power refers to the probability of making a correct decision to reject a false null hypothesis. In other words, power is the probability that a test will detect an effect when there is in fact an effect. For example, consider a prescription drug Company A that has manufactured a new sleep-aid pill. This company wants to show that their drug is more effective than the current leading drug manufactured by Company B. To do so, the company collects data on those who take their pill as well as those who use their competitor's pill. This company wants to show that their consumers sleep more hours per night than their competitor's consumers. In this case, the null hypothesis would claim that there is no difference between the two companies. The alternative hypothesis would claim that Company A's pill provides more hours of sleep (μ_A) than Company B (μ_B). In statistical terms:

$$H_0: \mu_{A\,=}\,\mu_B$$

$$H_0: \mu_{A\,>}\,\mu_B$$

As seen in Table 1, there are four possible outcomes in a hypothesis based statistical test: two possible correct decisions and two possible types of error. A correct decision could occur if a true null hypothesis was not rejected or if a false null hypothesis was rejected. A Type I (α) error occurs when a true null hypothesis H_0 is rejected whereas a Type II error (β) occurs when a false H_0 is not rejected. Type I errors are often called false positives while Type II errors are called false negatives. Using the example from above, a Type I error would mean that there was in fact no difference between the two pills; however, sample data led researchers to reject the null hypothesis that the two pills provided equal amounts of sleep. A Type II error would occur if there was in fact a difference between the two pills; however, sample data led researchers to fail to reject the null hypothesis.

When the probability of detecting a Type II error is low in hypothesis based testing, statistical power is high. This means that there is high power to detect an effect when there really is an effect. A power value of .80 or higher is deemed adequate among researchers (Cohen, 1988; Muthén & Muthén, 2002). In other words, statistical power is considered high when there is a probability of 80% or higher to detect an effect when there is, in fact, an effect.

Table 1: *Four Possible Outcomes of Research*

	True State	
Decision	H_0 True	H_0 False
Do not reject H_0	Correct decision (1 - α)	Type II error (β)
Reject H_0	Type I error (α)	Correct decision (1 - β = Power)

Note. Adapted from "The Relation Among Fit Indexes, Power, and Sample Size in Structural Equation Modeling," by K. H. Kim, 2005, *Structural Equation Modeling, 12*(3), p. 368-390.

Muthén and Muthén (2002) show how Monte Carlo simulation studies can be used to determine necessary sample size and how to detect statistical power. To demonstrate Monte Carlo studies, this paper used two latent variable models, specifically a confirmatory factor analysis (CFA) and a growth model. For the CFA, they studied how non-normality and missing data affected the sample size necessary for adequate power of factor correlations. Non-normal data had a greater influence on detecting the statistical power of factor correlations than data that were missing completely at random. Findings suggested that regardless of normality, missing data increased the required sample size by 18 percent. When data were both non-normal and missing, the required sample size was increased 100 percent. In other words, the study found that when these two complications are present, a CFA study needs twice as many participants to provide adequate statistical power.

The second part of Muthén and Muthén's (2002) study examined power in a growth model. The simulation conditions for this study included missing data, regression coefficient size, and a covariate. Results indicated that reducing the regression coefficient from .2 to .1 (*d* = .63, *d* = .32, respectively) had the biggest influence on the necessary sample size. Regardless of whether the data were missing, when the regression coefficient was decreased from .2 to .1, the sample size needed for adequate power increased four times. The paper concluded that statistical power was highly conditional on the varying factors of each model and that these conditions vary between statistical methods.

Fan (2003) also used a simulation study to compare sample size requirements for power, although this time in latent growth curve modeling under the structural equation modeling (SEM) framework and in repeated-measures analysis of variance (ANOVA). Results from the simulations revealed that the SEM latent growth models had higher statistical power for detecting group differences than the repeated-measures ANOVA. The study also showed that to detect a small group difference, a sample size of $N > 500$ was typically necessary for a power value between .70 and .80. For a medium group difference, $100 \leq N \leq 200$ was needed. Another major finding from this study was that to yield adequate power, the sample size using the SEM approach could be two-thirds to one-half the amount of that using the repeated-measures approach.

Necessary sample size for adequate power is a key area of interest to researchers using latent variable models such as factor analysis, SEM, and LTA. Muthén and Muthén (2002) state that some claim a rule of thumb of five to ten observations per parameter, while others state no less than 100, and others recommend 50 observations per variable. However, as Muthén and Muthén point out, there is no guideline that can be applied to all models or modeling conditions. In fact, a Monte Carlo simulation study found that there is a strong interplay between sample size, the number of indicators in a model, and class enumeration in

LCA models (Morovati, 2014). Thus, simulation studies can be helpful in determining what sample size is required to provide adequate statistical power to decrease the probability of a Type II error within a particular type of statistical method or model. Simulation studies have been conducted using many statistical models such as latent growth curve modeling and factor analysis. However, as stated, extensive research has not been conducted to investigate sample size requirements and power in LTA.

1.3 The Present Study

It has been established that latent variable research and mixture modeling are useful and widely used statistical approaches in the social sciences. Additionally, analyses of longitudinal change are essential in many disciplines such as developmental, behavioral, social, and health research. Together, LTA is an increasingly popular and advantageous method to discern change over time. Despite LTA's functionality, to this date only two simulation studies have been conducted to examine power for this method. The goal of this study is to assess how varying sample size and other conditions affect statistical power in LTA. Because so few LTA simulations have been conducted, the scope of this study is to investigate models that are commonly found in literature and the levels of statistical power that these conditions produce. This study also aims to provide recommendations for LTA use. This study aims to find a minimum sample size that provides adequate power under the proposed varying conditions. The results of this study can contribute to the mixture modeling literature by providing sample size and modeling guidelines for LTA use in applied research.

1.4 Overview of Dissertation

This section will outline the chapters included in this dissertation. First, Chapter 2 will serve as a literature review of how LTA has been used in research. This chapter will also include fundamental information about the LTA model and its parameters. Chapter 3, the methods section, will summarize meta-analysis findings of recently published LTA studies. The results from this analysis guided the conditions that were ultimately used in this Monte Carlo simulation study. This chapter also outlines how data were generated and all of the various conditions that were imposed in the simulation study, as well as explain how potential class switching issues were addressed. Next, Chapter 4 will provide analyses of all simulation studies and summarize results. In this chapter, issues that arose in this study are explained, such as artificially high power and model non-convergence. Chapter 5 includes a discussion of all results, practical implications, limitations to this study and future directions for research. The Appendix includes sample Mplus syntax with annotated comments.

Chapter 2 Literature Review

2.1 Overview of LTA

Latent transition analysis (LTA) is a longitudinal extension of latent class analysis (LCA) and was first introduced in the 1950s by researcher Paul Lazarsfeld (Lazarsfeld & Henry, 1968). LTA was further developed by Goodman (1974) and Haberman (1979) when they provided more efficient maximum-likelihood estimation algorithms. LTA is a longitudinal approach that examines qualitative changes in latent statuses where the main objective is to examine how individuals transition between latent classes over time. LTA uses repeated measures data of the same people over time. For example, consider a sample of female high school seniors enrolled in a mentoring program to encourage interest in the sciences. In the beginning of the year, these students take a survey that measures attitudes toward science. A LCA of these data reveal four distinct latent classes. After one year of attending events and mentoring sessions, students take the same survey. At this point, after a year of program participation, an LTA can be conducted to see whether these young girls shifted from one class to another or whether they remained in their original latent class. This particular hypothetical LTA example could potentially reveal whether young girls were more likely to have an increased interest in science or more likely to want to major in a science as an undergraduate after participating in the mentorship program. Note that the data collected were repeated measures data at two time points, although LTA can be conducted on more than two time points.

2.2 General applied example

To help demonstrate how LTA has been used in applied research, this section will walk the reader through a published journal article that used LTA. Lee, Chassin, and Villalta

(2013) investigated whether individuals “matured out” of alcohol involvement from age 17 to age 40. Lee et al.’s study aimed to examine alcohol involvement due to both the short- and long-term risks of alcoholism. Their literature review revealed that alcohol involvement increases in late adolescence, peaks between ages 20–22, and then tends to decrease thereafter. However, literature also revealed that there tends to be four groups of alcohol users over time: one group that “matures out” of drinking habits over time, two groups of abstainers or low users, and one “chronic” group of alcohol users who tend to never mature out of drinking habits. Lee et al.’s rationale for their LTA study was that studies focus on one area of alcohol involvement but never all three areas of alcohol involvement: drinking frequency, binge drinking, and drinking consequences.

To examine all three areas in one study, Lee et al. (2013) used longitudinal data from individuals aged 17 for four waves of data collection until age 40 ($N = 844$). The study found four latent statuses: abstainers, low-risk drinkers, moderate-risk drinkers, and high-risk drinkers. Results from their LTA revealed that individuals tend to mature out of heavy or problematic drinking and continue to drink alcohol, but at lower levels. The study supported existing literature that claims that individuals tend to mature out of drinking between late adolescence and young adulthood. However, the use of LTA added to literature by modeling how these high-risk drinkers typically matured to the next lowest level, moderate-risk, and rarely to a low- or non-risk drinking status. This result was particularly interesting because it revealed that individuals who start off as high-risk drinkers at late adolescence rarely eliminated all risky drinking behavior.

2.3 LTA Model and Parameters

2.3.1 LTA model. The general LTA model is similar to its non-longitudinal counterpart, LCA. Figure 1 presents the path diagram representation of the general LCA model. This model represents latent class membership at one given time point. The model diagram for the general LTA model can be seen in Figure 2. In the LCA diagram (Figure 1), the observed variables, *u*, are in rectangles, while the unobserved, latent class variable, *C*, with *K* classes is in a circle. The observable variables u_{tj} in the LTA model, on the other hand, have two subscripts: one to represent time *t* and one for each outcome *j*. Latent class *C* at time *t* is regressed onto latent class *C* at time *t* - 1. In this LTA model, latent class C_2 is regressed onto C_1.

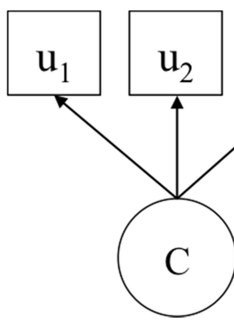


Figure 1. General *LCA model diagram with three indicator variables.*

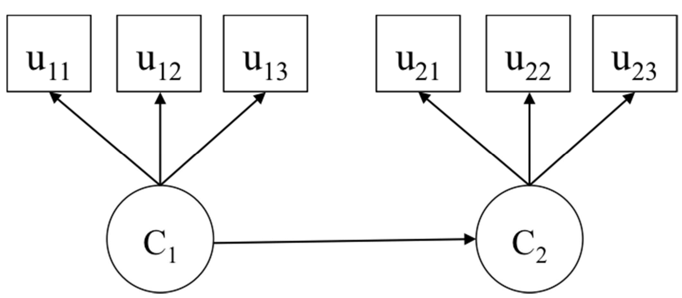


Figure 2. General LTA model diagram with two time points and three indicator variables.

2.3.2 Parameters. Three different parameters are estimated in LTA: latent status prevalences (δ), item-response probabilities (ρ), and transition probabilities (τ). First, latent status prevalence (δ_{s_t},) is the probability of being in latent status *s* at time *t.* If there are three classes at each of two time points, there are six δ's rather than nine because the last class is treated as a reference group. An individual can only be a member of one latent class at each time point and thus,

$$\sum_{s_t=1}^{S} \delta_{s_t} = 1,$$

where *s* is latent status at time *t.* Latent status prevalences for time $t = 1$ are often estimated independently. In this case, latent status prevalences for times $t \geq 2$ can be computed by:

$$\delta_{s_t} = \sum_{s_{t-1}=1}^{S} \delta_{s_{t-1}} \tau_{s_t|s_{t-1}} .$$

As seen above, the probability of being in latent status *s* at time *t* is a function of the probabilities of being in a latent status at time *t* - 1 and the conditional probability of transitioning from a latent status at time *t -1* to a latent status at time *t*. In a model with two time points, δ can be computed once latent status prevalances at $t = 1$ and the transition probabilities between $t = 1$ and $t = 2$ have been estimated.

Next, item response probabilities can be expressed by $\rho_{j,r_{j,t}|s_t}$, or the probability of response $r_{j,t}$ to observed variable *j*, conditioned on membership in latent status *s* at time *t*. There are R_j item response probabilities. An individual can only provide only one response to variable *j* at each time *t*, and thus,

$$\sum_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|s_t} = 1.$$

The number of item-response probabilities estimated can be computed with the following equation:

$$P_{\rho} = ST\sum_{j=1}^{J}(R_j - 1) \quad .$$

In a model with 3 latent statuses, two time points, and seven binary indicators, the number of item-response probabilities estimated would be equal to P_{ρ} = 3 x 2 x (1 + 1 + 1 + 1 +1 + 1 + 1) = 42. If item-response probabilities are constrained to be equal, this would reduce to P_{ρ} = 3 x 1 x (1 + 1 + 1 + 1 +1 + 1 + 1) = 21. Often in LTA, parameter restrictions are imposed such that item-response probabilities are set to be equal across times. This is often called measurement invariance. Latent statuses are assumed to be constant over time and do not change meaning.

Lastly, transition probabilities ($\tau_{s_{t+1}|s_t}$) represent the probability of transitioning from latent status *s* at time point *t* + 1 conditioned on membership in latent status *s* at time point *t*. Transition probabilities are often the most examined parameter, as they reveal how latent status membership changes over time. In the context when measurement invariance is assumed, these probabilities are called stability estimates, as they represent how stable—or unstable— latent status membership over time. Latent statuses are also considered recurrent or transient. Processes that remain at its state are considered recurrent, whereas processes that do not return to its state are considered transient (Marcoulides, Gottfried, Gottfried, & Oliver, 2008). Transition probabilities are often represented in a matrix. There are *T* - 1 matrices, where *T* is the total number of time points included in the study. A transition probability matrix of τ's is as follows,

$$\begin{bmatrix} \tau_{1_{t+1}|1_t} & \tau_{2_{t+1}|1_t} & \cdots & \tau_{S_{t+1}|1_t} \\ \tau_{1_{t+1}|2_t} & \tau_{2_{t+1}|1_t} & \cdots & \tau_{S_{t+1}|2_t} \\ \cdots & \cdots & \cdots & \cdots \\ \tau_{1_{t+1}|S_t} & \tau_{2_{t+1}|S_t} & \cdots & \tau_{S_{t+1}|S_t} \end{bmatrix}.$$

The rows represent the first time point whereas the columns represent later time point, t - 1. When measurement variance is assumed over time, the diagonals represent the probability of remaining in the same status. The off-diagonals represent the probability of being in a latent status conditional on being in a different latent status at the previous time point. If no parameter restrictions are imposed, the number of transition probabilities estimated can be computed with the following equation,

$$P_\tau = (T - 1)S(S - 1).$$

One is subtracted from S in the third part of the product because the last class is treated as a reference class. In a model with two time points and three latent statuses, the number of transition probabilities would be equal to P_τ = 1 x 3 x 2 = 6. Individuals may only belong to one latent status at each time and thus,

$$\sum_{s_{t+1}=1}^{S} \tau_{s_{t+1}|s_t} = 1 ,$$

meaning each row of the transition probability matrix sums to 1, with some rounding error. Taken together, the fundamental expression of LTA is expressed as,

$$P(Y = y) = \sum_{s_1=1}^{S} \cdots \sum_{s_T=1}^{S} \delta_{s_1} \tau_{s_2|s_1} \ldots \tau_{S_T|S_{T-1}} \prod_{t=1}^{2} \prod_{j=1}^{J} \prod_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|s_t}^{I(y_{j,t}=r_{j,t})} .$$

This equation shows that the probability of a particular vector of responses is a function of the three aforementioned probabilities: the probability of membership in each latent status at t = 1 (latent status prevalances, δ), the probability of being in a latent status at a later time conditional on the previous time (latent transition probabilities, τ), and the probability of each response at each time point conditional on latent status membership (item-response probabilities, ρ).

When there are only two time points, the equation reduces to:

$$P(Y = y) = \sum_{s_1=1}^{S} \cdots \sum_{s_2=1}^{S} \delta_{s_1} \tau_{s_2|s_1} \prod_{t=1}^{2} \prod_{j=1}^{J} \prod_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|s_t}^{I(y_{j,t}=r_{j,t})}$$

.

2.3.3 Measurement invariance. The major assumption of measurement invariance is that measurement parameters, the item response probabilities defined earlier, are held equal across the entire measurement model. Full measurement invariance denotes that the conditional item probabilities are the same across all time points. In other words, at all time points, there are equal numbers of classes and each class is the same over time. Classes do not change meaning or interpretation. Measurement invariance must be tested. Measurement invariance should be tested because the assumption of invariance introduces bias, although many LTA studies assume measurement invariance because less measurement parameters must be estimated and interpretation is less complicated.

2.3.4 Applied example with parameters. To help exemplify the LTA parameters described above, this section will walk through an applied LTA article. Marcoulides et al. (2008) used LTA to examine academic intrinsic motivation from childhood to adolescence. This study utilized data from the Fullerton Longitudinal Study (FLS), which collected a wide range of developmental data from age one to 17 years old. In particular, this study examined academic intrinsic motivation using Likert scale measures from the Children’s Academic

Intrinsic Motivation Inventory (CAIMI). The study used five time points: ages 9, 10, 13, 16 and 17.

First, the researchers evaluated one-, two-, three-, and four-class models at all five time points. To handle missing data, the models were based on full information maximum likelihood estimation. It also used random start values to ensure that models converged on global solutions. The Bayesian Information Criterion index was used to evaluate the most appropriate class model. At all five time points, the three-class model had the best BIC fit values relative to the other models. Researchers labeled these three classes Intermediate, At-Risk, and Gifted. Item-response probabilities (ρ) were used to develop these classes and their researcher-defined titles, although Marcoulides et al. (2008) did not include item-response probabilities for each item and class.

The second of the three LTA parameters is exemplified here in each individual's latent status prevalence (δ). At age 9 (t = 1), the majority of students were in the Gifted class while 36% of students were in the Intermediate Group and only 7% in the At-Risk group. As students got older, less students were in the Gifted group and more students were in the At-Risk group. The Intermediate group was the most consistent over time, with class membership percentages between 20-25% over the next four time points between ages 10 and 17. From age 9 to 17, the Gifted group dropped from 57% to 19% while the At-Risk group increased from 7% to 59%. These latent status prevalences showed that over time, students were less likely to be in the Gifted group and more likely to be in the At-Risk group.

Latent transition probability (τ) matrices between each time point helped show how students transition to or stayed in a class over time. This study not only looked transition probability matrices between each consecutive time point, it also examined what are called higher order lag models. Higher order lag models allow researchers to see the latent transition probability at one time point from any other time point in the model, not just the time point immediately prior. Marcouides et al. (2008) were also interested in the transition from childhood to late adolescence, and thus examined latent transition probabilities: ages 9 to 13, 9 to 16, 9 to 17, 10 to 16, 10 to 17, and from ages 13 to 17.

In general, this study found that transition between the three instrinsic motivation classes mostly occurred during childhood. As students get older, they are more likely to stay within the same latent class. This study revealed important findings that at-risk students were highly likely to stay at-risk over time and that by mid-adolescence, it was unlikely a child would transition into the gifted class. From a practical standpoint, researchers argued the need for early motivational intervention. In summary, we see here the interplay of the three LTA parameters and how they are interpreted in an applied study.

2.3.5 Sparseness. Contingency tables in LTA are large because data are measured at two or more time points. This sparseness of cells may lead to identification problems. For example, an LTA with eight binary items at two time points would have a contingency table with $W = 2^{(8)(2)} = 65{,}536$ cells. That means that there are 65,536 possible response patterns. Because of the large number of cells, LTA models tend to have a large amount of degrees of freedom. Degrees of freedom in LTA can be computed by:

$$df = Wi - P_\delta - P_\rho - P_\tau - 1.$$

Despite this advantage, individual cells can be sparse, leading to estimation issues. Models may fail to converge. Another issue involving sparseness the distribution of the G^2 statistic (a likelihood ratio test to test for goodness of fit) is no longer well represented by the chi-square distribution and thus, *p*-values are inaccurate. Nonetheless, Collins and Flaherty (2002) state that parameter estimation using EM is robust even when cells are sparse.

The sparseness of cells in LTA is an important issue that may lead to low power in LTA studies. The following is a numerical demonstration of how sparseness can occur in LTA based on varying latent transition probabilities and the distribution of the sample across latent classes.

First, consider an LTA study with a fairly large sample size of N = 3,000. In this hypothetical study, assume that data were collected at two time points and that equal class/statuses sizes emerge at time t = 1 and measurement invariance was assumed. Further, assume that there was a fairly stable amount of transition from time t =1 to t = 2, specifically, about three-quarters of participants did not transition to a different class. The remainder of individuals did transition, which resulted in a transition probability matrix as follows (see Table 2 below). Table 2 shows how the sample size would be using the above transition probabilities and an initial class size of n = 1,000 for all k classes at t = 1. To most, this contingency table may not seem to suffer from sparseness, as the largest cell has 800 individuals and the smallest still has 100.

Table 2: *Transition Probabilities for Hypothetical LTA with N = 3,000, Moderate Stability, and Even Class Sizes*

		t = 2		
		Class 1	Class 2	Class 3
t = 1	Class 1 (n = 1,000)	0.70 (n_{11} = 700)	0.20 (n_{12} = 200)	0.10 (n_{13} = 100)
	Class 2 (n = 1,000)	0.10 (n_{21} = 100)	0.80 (n_{22} = 800)	0.10 (n_{23} = 100)
	Class 3 (n = 1,000)	0.10 (n_{31} = 100)	0.15 (n_{32} = 150)	0.75 (n_{33} = 750)

Now consider the same scenario, yet in a study where participants are highly likely to stay in their initial latent class. Again, this hypothetical study has N = 3,000, equal class sizes at time point 1, and assumes measurement invariance. This study results in the transition probability matrix seen in Table 3:

Table 3: *Transition Probabilities and Sample Size Distribution for Hypothetical LTA with N = 3,000, High Stability and Even Class Sizes*

		t = 2		
		Class 1	Class 2	Class 3
t = 1	Class 1 (n = 1,000)	0.97 (n_{11} = 970)	0.20 (n_{12} = 20)	0.03 (n_{13} = 300)
	Class 2 (n = 1,000)	0.04 (n_{21} = 40)	0.95 (n_{22} = 950)	0.01 (n_{23} = 10)
	Class 3 (n = 1,000)	0.05 (n_{31} = 50)	0.05 (n_{32} = 900)	0.90 (n_{33} = 900)

Unlike the first scenario, the transition probabilities along the diagonal are higher, resulting in smaller probabilities in the cells outside of the diagonal. The smallest cell size in this case is 10. Even with a study with 3,000 participants, some researchers may deem a cell size of 10 to be too small, as this number is close to but not violating Cochran's (1954) suggestion that no more than 20% of cells in a chi-square test contingency table have cell sizes less than 5.

A sample size of N = 3,000 may not be feasible to most social science researchers. It is likely that a dataset only contain about 300 participants. Consider the same two previously used scenarios with a decreased sample size of 300. Using the same moderately stable transition probabilities in Table 2 results in the following sample distribution for a study with N = 300. Again, assume that there is measurement invariance across time points and that class sizes are even at time t = 1.

Table 4: *Sample Size Distribution for Hypothetical LTA with N = 300, Moderate Stability, and Even Class Sizes*

		t = 2		
		Class 1	Class 2	Class 3
t = 1	Class 1 (n = 100)	0.70 (n_{11} = 70)	0.20 (n_{12} = 20)	0.10 (n_{13} = 10)
	Class 2 (n = 100)	0.10 (n_{21} = 10)	0.80 (n_{22} = 80)	0.10 (n_{23} = 10)
	Class 3 (n = 100)	0.10 (n_{31} = 10)	0.15 (n_{32} = 15)	0.75 (n_{33} = 75)

The smallest sample size in this scenario is 10, which may not seem problematic. However, now examine how a study with high stability/values along the diagonal, using transition probabilities in the second example seen in Table 5:

Table 5: *Sample Size Distribution for Hypothetical LTA with N = 300, High Stability, and Even Class Sizes*

		t = 2		
		Class 1	Class 2	Class 3
t = 1	Class 1 (*n* = 100)	0.97 (n_{11} = 97)	0.20 (n_{12} = 2)	0.03 (n_{13} = 30)
	Class 2 (*n* = 100)	0.04 (n_{21} = 4)	0.95 (n_{22} = 95)	0.01 (n_{23} = 1)
	Class 3 (*n* = 100)	0.05 (n_{31} = 5)	0.05 (n_{32} = 90)	0.90 (n_{33} = 90)

The cell sizes decrease dramatically. Although the cell counts in the diagonal are high, only one participant transitioned from class 2 to class 3. Counts are less than 10 in six of the nine cells. These cells would be even smaller if class sizes were not even at time *t* = 1. In this situation, all conditions are kept the same; however, as seen in the right-hand marginal, class sizes at *t* = 1 vary. The smallest cell count now is less than 1.

Table 6: *Sample Size Distribution for Hypothetical LTA with N = 300, High Stability, and Uneven Class Sizes*

		t = 2		
		1	2	3
t = 1	1 (*n* = 150)	0.97 (n_{11} = 145.5)	0.20 (n_{12} = 3)	0.03 (n_{13} = 4.5)
	2 (*n* = 80)	0.04 (n_{21} = 3.2)	0.95 (n_{22} = 76)	0.01 (n_{23} = .8)
	3 (*n* = 70)	0.05 (n_{31} = 3.5)	0.05 (n_{32} = 3.5)	0.90 (n_{33} = 63)

In applied research, it may be likely that off-diagonal transition probabilities are low, suggesting that a small percentage of individuals transition from one class to another over time. There are no current rules of thumb for cell count size in LTA contingency tables. This study aims not to examine a minimum cut-off for cell count. Rather, this study aims to investigate how the interplay of model characteristics such as high and moderate stability with sample size affects the overall statistical power of a study. A meta-analysis was conducted to explore commonly found transition probabilities. Results from this meta-analysis can be found in the Method section of this dissertation.

2.4 Areas for Continued Work

As stated earlier, LTA is gaining greater use yet only two simulation studies have been conducted examining best practices and sample size requirements for this statistical model. Simulation studies looking at power have shown how sample size, along with other varied conditions, has effects on the overall statistical power of a model. However, to date, a power simulation study has not yet been conducted for the LTA model. This gap in research calls for further investigation on how conditions such as sample size, the size of latent transition probabilities, and sparseness of cells affect power. This dissertation aims to find whether there is a point at which an LTA no longer has the power to detect an effect if sample size is decreased. In other words, if two LTA models have the same number of time points, the same number of classes at each time point, identical transitional probabilities, equal sample distributions of across classes, yet different sample sizes, will both models produce adequate power?

Because so few simulation studies have been conducted in this area, there are many areas of study for the LTA model. Many other simulation studies look at how fit indices perform in a given model and whether these indices are better or worse when sample size changes or whether the number of indicators increase or decrease. These types of studies in LTA would also provide insight into how LTA models perform in practice. The scope of this current study is to investigate power when varying sample size, measurement models, logit thresholds, transition probabilities, and class size. The hope is that the results of this simulation will be the first step in establishing modeling guidelines for the specification and use of LTA models and a better understanding of the power of latent transition probabilities.

Chapter 3 Method

3.1 Empirical Conditions

For this study, a meta-analysis was conducted to examine the characteristics of recent LTA studies. In Monte Carlo simulation studies, population values are often chosen based from theory or previous research. Muthén and Muthén (2002) recommend using values from previously conducted studies. To do so, four online social science databases were used: Education Resources Information Center (ERIC), EBSCOhost, PsycINFO and PsycARTICLES. Only recent, full-text articles in peer-reviewed journals from 2008–2014 were included in this meta-analysis.

Using the above criteria, a keyword search of the phrase “latent transition analysis” resulted in a total of 92 unique articles across the four databases. Of these 92 articles, 38 were removed from the analysis because 1) the search phrase was mentioned in article but not used as a method of analysis, or 2) the search phrase appeared in the reference section (i.e., the search phrase was part of a journal article title cited in the study). Two articles were eliminated from the meta-analysis because they were commentaries and one article was omitted because it was a comparison of different longitudinal approaches. In the final analysis, 54 articles were examined.

A quick overview of these articles revealed a wide range of fields that utilize LTA. Because LTA is an approach examining longitudinal change, many articles examined a treatment effect over time. For example, a number of articles examined clinical eating disorder classifications after counseling and treatment, reading ability after an intervention, or substance abuse after rehabilitation. Other studies did not include a treatment or clinical trial and rather focused on how individuals transitions over time, or more specifically, over ages. For example, Quaiser-Pohl, Rohe, and Amberger (2010) examined mental-rotation ability beginning from age 10 to age 17. This study found a three-class solution: intermediate, at-risk, and gifted. Results revealed that between ages 10 and 13, individuals were more likely to transition between classes. However, from ages 13 to 17, transition probabilities were highly stable. Other LTA topics included attitudes among foster care youth, depressive subtypes, and civic involvement.

To gather characteristics of common LTA studies to use as attributes for the simulation study, sample size, model characteristics, and transition probability matrices were compiled and are explained in the subsections below.

3.1.1 Sample size. First, sample sizes of these 54 articles were examined. The sample size of these studies ranged from $N = 94$ to $N = 11{,}750$ ($M = 1493.07$, $SD = 2083.60$). As seen in Figure 3, more than half of the journal articles involved studies with samples less than 1,000. A third of the total number of articles had sample sizes between 200 and 500. Only two studies had sample sizes larger than 5,000, both of which used nationally administered datasets. Figure 4 presents a closer look at the distribution of articles with $N <$ 1,000. This quick overview of study sample sizes shows that 16 of the 54 published LTA articles used sample sizes less than 500.

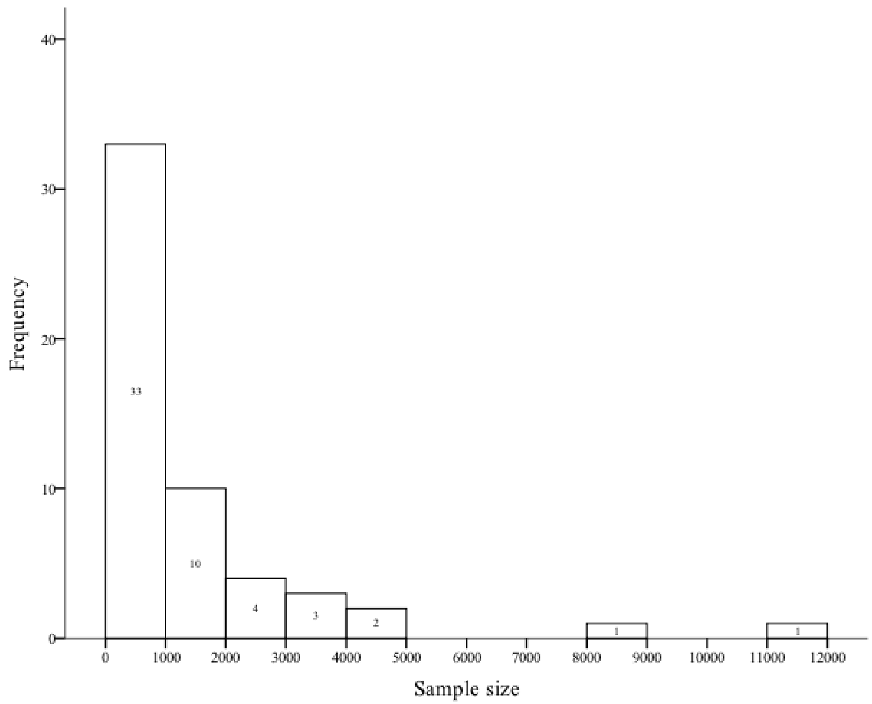


Figure 3. Histogram of sample sizes included in meta-analysis articles.

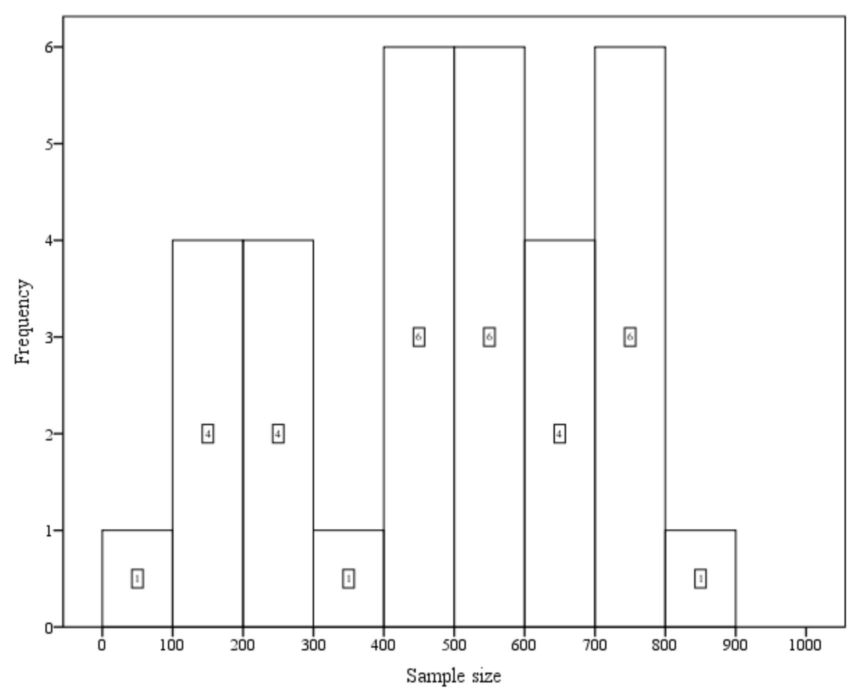


Figure 4. Histogram showing studies with sample sizes less than N = 1,000 included in meta-analysis articles.

3.1.2 Model Characteristics. Next, model characteristics were examined. As seen in Figure 5, studies ranged from two to nine time points (*M* = 2.65, *SD* = 1.20). The majority of articles included two time points. In 93% of the articles, there was the same number of classes at each time point. Additionally, class sizes were typically uneven at time $t = 1$ and measurement invariance was assumed.

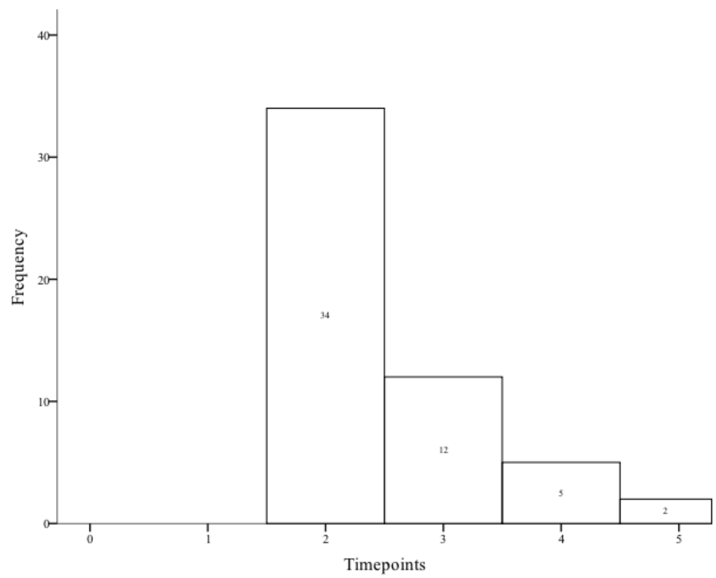


Figure 5. Histogram of number of time points included in meta-analysis articles.

3.1.3 Transition probabilities. Next, latent transition probabilities were examined. A wide range of probabilities emerged from these studies. Latent transition matrices ranged from being highly stable (values close to 1.00 on the diagonal) to moderately stable (values close to .70 on the diagonal and values near .30 on the off-diagonal). More than half of the studies reported at least one stability estimate near .90. The lowest value on the diagonals was near .50 for some articles, although some articles reported diagonal values close to .20. In one study, latent transition stability was very low, with values between .00 and .50. In the off-diagonals, the lowest transition probabilities for most studies were near .00 and the highest off-diagonal values were between .20 and .40. In some cases, off-diagonals exceeded .50 and were even as high as .94, exhibiting high transition between classes over time.

3.2 LTA by Topic

To gain a better understanding of the size of latent transition probabilities we see in the literature, articles were further examined by discipline. Most of the identified articles fell under three areas: substance abuse (19 articles, 35%), Education (9 articles, 17%), and eating disorders (6 articles, 11%). The remainder of articles included in the meta-analysis ranged from topics such as family structure and intervention, foster care, and social and civic engagement. A further look at these articles by the most common topics revealed patterns in latent transition probabilities and model characteristics summarized below and displayed in Table 7.

3.2.1 LTA models in Education research. Nine of the papers identified in the meta-analysis were related in the field of education. More specifically, four articles were about mathematics ability, four articles were on reading ability, and one article examined intrinsic motivation. The mathematics and reading articles examined kindergarten through early adolescence, low-performing students or late-emerging readers. In most studies, there were three or four classes at each time point, with the first class (class 1) representing low performance and the last class (class 3 or 4) representing gifted students or high performing students. When organized this way, transition probability values above or to the right of the diagonal represent transitioning into a higher performing class. For example, Ding, Richardson, and Schnell (2013) examined word literary from kindergarten to second grade. This study identified a three-class solution: low achievers (class 1), slow achievers (class 2), and typical achievers (class 3). Because of the classes were listed in the matrices in this hierarchy, the diagonal values represented stability and values to the right of the diagonal represented movement into a higher-achieving class over time.

A further look at the latent transition probabilities of all education related LTA studies showed that almost all transition probability matrices had fairly high stability rates of .70 or higher, meaning 70% or more of individuals stayed in their latent class over time. If students did transition, it was usually to the next highest class, as one would expect. In other words, off-diagonal values were highest when immediately to the right of the diagonal value. Some students transition down to a lower achieving class. However, this was usually with a transition probability of .10 or less. For all papers in this group, the latent class sizes were uneven. In most cases, the lowest performing class had the highest *n*. Additionally, all nine Education articles assumed measurement invariance across all time points. The results from the Education articles helped form the uneven transition probability matrix used in this study’s Monte Carlo simulation (see Table 10 in the next subsection).

3.2.2 LTA models in eating disorder research. Six of the meta-analysis articles examined latent transitions of individuals with eating disorders. Similar to the education studies, most of the latent classes found in the eating disorder studies were arranged order of severity (e.g., class 1 is asymptomatic while class 4 is the most severe case of eating disorders). For example, Cain, Epler, Steinley, and Sher’s (2012) study examined latent transitions between three classes: no obvious pathological eating-related concerns, limiting attempts with overeating, and pervasive bulimic-like concerns. In many studies, the first and second class had high stability while classes three and four had moderate stability. Because most of these articles examined transition after some treatment, this pattern of latent transition implies that individuals with less severe eating disorders are not likely to “get better" or "worse" after treatment. However, it does imply that those with more severe cases of an eating disorder are likely to transition into a less severe class after treatment. Similar to educational studies, class sizes in eating disorder LTA studies were uneven and the least severe class usually had the greatest *n.* All of the eating disorder LTA studies assumed measurement invariance.

3.2.3 LTA models in substance abuse research. There were 19 substance abuse studies found to use LTA in the meta-analysis. These articles varied across substance type, including alcohol, drug, and cigarette use. Substance abuse articles could be further divided into two sections: studies involving some sort of treatment or studies that examined use over time without a specific treatment. Again, similar to the education and eating disorder articles discussed earlier, the majority of substance abuse transition probability matrices were ordered by severity of substance use. Studies that did not involve a treatment, but rather looked at use across age, consistent found the following transition patterns over time:

- Nonusers tended to stay nonusers
- Heavy users tended to stay heavy users
- Moderate users shifted up or down
- Most transitions were one level up or down
- As one got older, stability increased

Two of the 19 substance abuse LTA studies did not assume measurement invariance. Both studies aimed to study the relationship between two different measures over two time points. For example, Abar (2012) examined the relationship between parenting types at time 1 and student alcohol-related behavior at time 2. This study found that students with pro-alcohol parents were more likely to high risk or extreme drinkers during their first year of college. Stapleton, Turrisi, Cleveland, Ray, & Lu (2014) looked at the relationship between alcohol decision-making patterns prior to college (time 1) and their patterns of alcohol use after entering college (time 2). This study revealed interesting patterns between the two measures. For example, given membership in the anti-drinker decision-making profile prior to the college, there was a probability of 100% of having a low drinking pattern. Furthermore, given membership in the risky decision-making profile, there was a probability of 55% of having a high drinking pattern.

Although these classes and transition probabilities emerged in published Education, eating disorder, and substance abuse studies, to date it is unknown how the measurement model influences the power of the latent transition probabilities. For example, it is unknown whether there is greater power for transition probabilities in or out of classes that are more distinct from others. This study will be the first to examine this and examine the extent to which measurement characteristics such as class separation and homogeneity affect power.

Table 7: *Summary of Meta-Analysis Results by Topic*

Topic		
Education	Eating disorders	Substance abuse
• Classes organized by achievement level • Reading or mathematics achievement • Childhood or early adolescence • Highly stable diagonals (.70 or above) • If individuals did transition, it was usually to the next highest/better class • Some transition down a class, but low transition probability (.10 or less) • Class sizes uneven • Lowest level class usually had highest *n*	• Classes organized by severity • Less severe classes had high stability • More severe classes had moderate stability • More likely to move down a class after treatment • Class sizes uneven • Lowest level class usually had highest *n*	• Classes organized by severity of substance use • Nonusers stayed nonusers • Heavy users stayed heavy users • Moderate users shifted up or down a class • Most transitions were one level up or down • As one got older, stability increased • Class sizes uneven

3.3 Summary of Meta-Analyses

A summary of LTA studies examined in this meta-analysis can be seen in Table 8. In general, most studies had sample sizes less than 1,000, two time points, uneven class sizes at $t = 1$, and assumed measurement invariance. Latent transition probabilities ranged across all articles. However, there were some patterns when examined by topic. These findings and patterns were used to create the simulation conditions used in this dissertation.

Table 8: *Summary of Meta-Analysis Results*

Sample size	Model Characteristics	Latent transition probabilities
• Most $N < 1{,}000$, but were as small as $N = 94$ and as large as $N \approx 11{,}000$	• Two time points • Uneven class sizes at $t = 1$ • Measurement invariance	• Wide range of transition probability patterns • See results by topic

3.4 Data Generation

For the current simulation study, data will be generated based on common characteristics of LTA studies found in the aforementioned meta-analysis to examine the statistical power of latent transition probabilities under various conditions and measurement models. The statistical software package Mplus (Muthén & Muthén, 1998-2014) will be used to conduct the simulation studies. For Monte Carlo simulation studies, Mplus has the capability to include both normal and non-normal data, missing data, clustering and mixture modeling. In this study, all generated data will include a total number of 1,000 replications and five binary indicators of latent class at of the two time points. Sample Mplus syntax file with annotated comments are in the Appendix.

The studies included in the meta-analysis ranged in the number of time points and the number of classes that emerged. However, because there are only two other existing power studies on LTA, the scope of this simulation study is to provide a foundational examination of how power is related to model conditions. Thus, this study will include four-class models measured at two time points and will assume measurement invariance for the latent classes across time. Latent classes will be defined by five items. Although studies in the meta-analysis covered a large range in the number of items used, we chose a parsimonious model as a starting point. A study with too few items may not be enough to reveal meaningful latent classes, yet a study with too many items may greatly negatively affect the level of power. The 4-class to 4-class model, and the number of time points, and the use of five items are the only three non-varying conditions. The conditions that will be varied include sample sizes, the value of logit thresholds, measurement models, latent transition probabilities, and class sizes.

3.4.1 Sample sizes. Eleven sample sizes will be used for each model in this simulation study (*N* =100, 250, 500, 1,000, 1,250, 1,500, 5,000, 6,000, 7,000, 8,000, 10,000). Initially, only six values of *N* were considered. However, as explained in the results section, issues regarding stability of power estimates required additional sample sizes to examine how models performed in intermediary values of *N*s.

3.4.2 Measurement models and item-response logit thresholds. Item homogeneity and class separation are two desirable attributes when selecting an LCA model (Collins & Lanza, 2010). Item homogeneity refers to item-response probabilities that are near 0 and 1. It is referred to as homogeneity because all members of that class have similar probabilities of endorsing an item. For example, an item-response probability close to 1indicates that there is nearly a 100% chance of endorsing that item, given membership in that class. Similarly, an item-response probability near 0 indicates that it is highly unlikely that members of that class would endorse that item.

High class separation occurs when each class has a distinct combination of item-response probabilities. In other words, classes should not look too similar to each other. For this simulation study, variance in homogeneity is reflected in item-response thresholds, while class separation is reflected in the definition of the measurement models. Because item-responses are on a probability scale, they range from 0 to 1. For Mplus simulation syntax, probabilities were converted to logits. Logits can be calculated from probabilities using the following formula,

$$\text{logit}(p) = \frac{1}{1+\exp(p)}.$$

Negative logit thresholds represent probabilities greater than .50 while positive logits reflect probabilities less than .50. In other words, low logits translate to high probabilities while high logits translate to low probabilities. This study will examine how low, moderate, and high threshold values affect the power of latent transition probabilities.

Two measurement models specifications were used in this simulation to investigate differences in power between LCA models that are well-defined versus those that are poorly defined. For the purposes of this study, a well-defined model is one in which there is high class separation. As seen in Figures 7 –10, there are high probabilities of endorsing all items given membership in class 3, while there are low probabilities of endorsing all items given membership in class 1. Given membership in class 4, there are higher probabilities of endorsing the first two items and lower probabilities for the last two items. Class 2 is the opposite of class 4, in that if an individual is classified in class 2, there are lower probabilities of endorsing the first two items and higher probabilities for the last two items. Logit threshold values for the well-defined model range from ±1, ±2, ±3, and ±5 which correspond to conditional probabilities of .27/.73, .12/.88, .05/.95, and .01/.99, respectively. This range was considered so we could examine the impact of homogeneity in item probabilities on statistical power. The exact threshold values used in each model type can be seen in Table 9.

Table 9: *Logit Thresholds for Each Model Type Used in This Study*

		Well-Defined Model				Poorly Defined Model		
		Thresholds ±1	Thresholds ±2	Thresholds ±3	Thresholds ±5	Moderate	Moderate (Revised)	High
Class 1	Item 1	1	2	3	5	-0.4	-3	-2.5
	Item 2	1	2	3	5	-1	0	-5
	Item 3	1	2	3	5	0	0	-1
	Item 4	1	2	3	5	0.4	0.4	2.5
	Item 5	1	2	3	5	1	0.85	2.5
Class 2	Item 1	1	2	3	5	1	0.85	1.5
	Item 2	1	2	3	5	1	3	1.5
	Item 3	-1	-2	-3	-5	0.4	0.4	0
	Item 4	-1	-2	-3	-5	0	0	-1
	Item 5	-1	-2	-3	-5	-1	-1	-5
Class 3	Item 1	-1	-2	-3	-5	1.5	1.3	5
	Item 2	-1	-2	-3	-5	1	0.4	1.5
	Item 3	-1	-2	-3	-5	1	0.4	1
	Item 4	-1	-2	-3	-5	0	3	-1
	Item 5	-1	-2	-3	-5	0	0	0
Class 4	Item 1	-1	-2	-3	-5	0	0	-1.5
	Item 2	-1	-2	-3	-5	1.5	0.4	5
	Item 3	1	2	3	5	-1	3	-5
	Item 4	1	2	3	5	1.5	1.3	5
	Item 5	1	2	3	5	1.5	1.3	5

The poorly defined model (see Figures 11–13) was included in this study to see how poor measurement models impact the statistical power to detect latent transition probabilities. The poorly defined models have low class separation and non-homogeneous classes. This model has more "noise," meaning the latent classes have many items that have conditional item probabilities near .50. This implies that there is a 50% chance of someone in in that class having endorsing that item, which means that the predictability of a person's response to that item in that class is equivalent to the odds of flipping a coin. Threshold values for the poorly defined model range from moderate to high. As explained in Chapter 4, an additional moderate (revised) thresholds model was added due to replication convergence issues. The exact threshold values used in each model type can also be seen in the sample Mplus output in the Appendix.

It is important to examine the difference between these two measurement models since it is more realistic to see published studies with the poorly defined model than the well-defined model. The poorly defined model includes classes that have many overlaps in the item probabilities among the four classes. This is seen in published LTA articles and often is supported by theory. For example, Peterson et al.'s (2011) study on eating disorders revealed a three-class LCA model: binge eating/purging, binge eating, and low-BMI. The binge eating/purging and binge eating classes look very similar on all of the five items except for the item measuring compensatory behaviors (i.e., actions that "un-do" binge eating, such as self-induced vomiting or over-exercising after binge eating). Theoretically, this is the key item that distinguishes members of the two classes and is a major distinction when it comes to classification and treatment. Although this model may seem to be "poorly defined" with respect to measurement characteristics because of the overlap of classes, it makes sense theoretically.

There are a number of potential issues that can emerge from this realistic measurement model. In the Peterson et al. (2011) study, there was one key variable that differentiated two of the three classes. Without this variable, the two classes would look almost exactly the same. However, adding more distinguishing items is sometimes difficult for researchers with small sample sizes. A measurement model with overlapping classes may have lower entropy but again, might make strong theoretical and practical sense. There might be lower power to detect the latent transition probabilities and a larger sample size might be necessary to reveal an adequate level of power in modeling conditions where the classes have a lot of overlap. Although the scope of this study does not examine the intersection of entropy with the measurement model, sample size, and power, it is important to acknowledge that they are not independent ideas. Thus, in a related sense this dissertation aims to take the first look at how the measurement model affects power.

Figure 7. *Well-Defined Model with Thresholds ±1.*

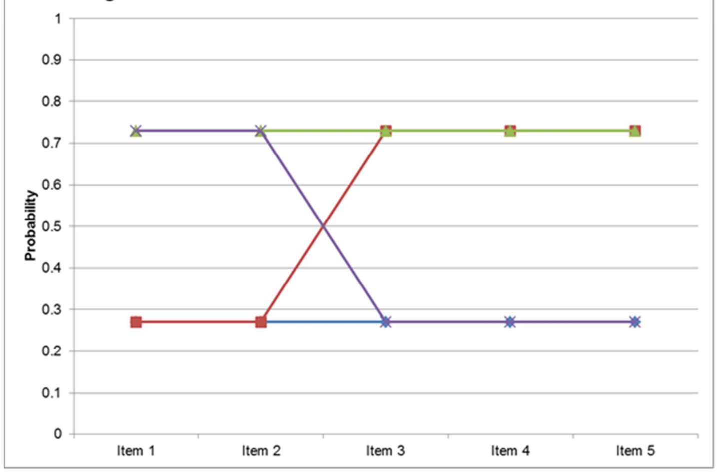


Figure 8. *Well-Defined Model with Thresholds ±2.*

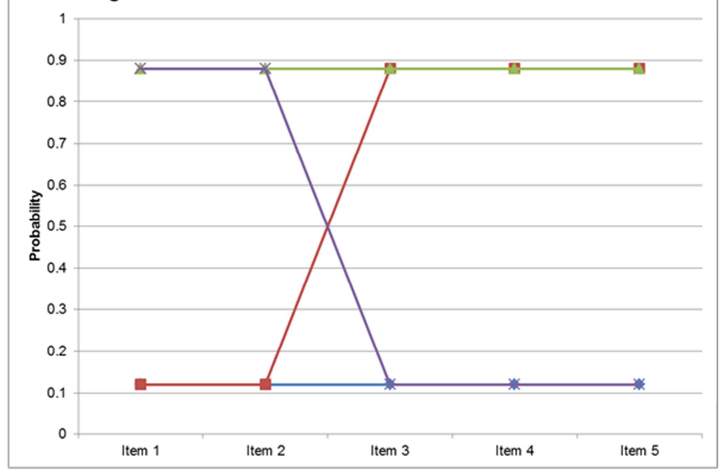


Figure 9. *Well-Defined Model with Thresholds ±3.*

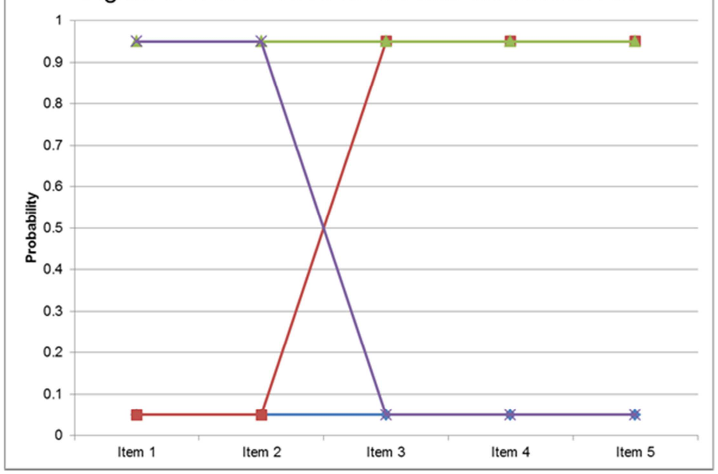


Figure 10. *Well-Defined Model with Thresholds ±5.*

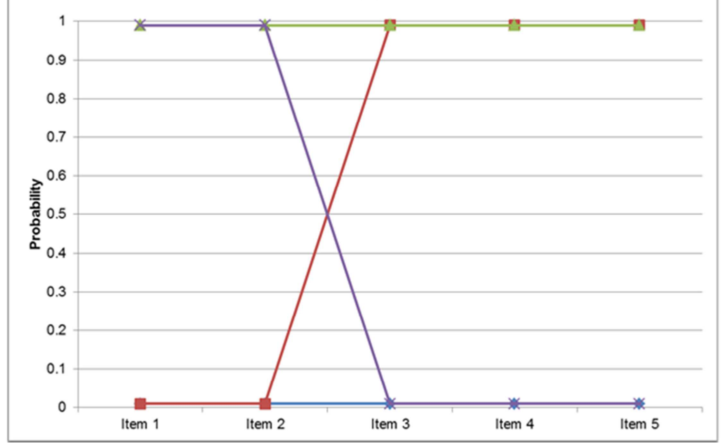


Figure 11. *Poorly Defined Model with Moderate Thresholds.*

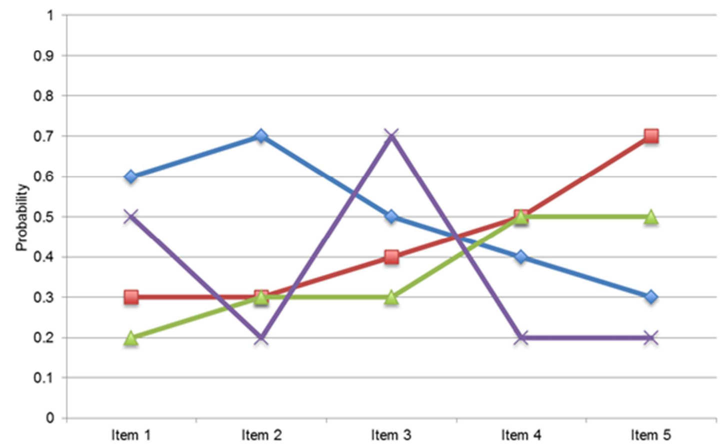


Figure 12. *Poorly Defined Model with High Thresholds.*

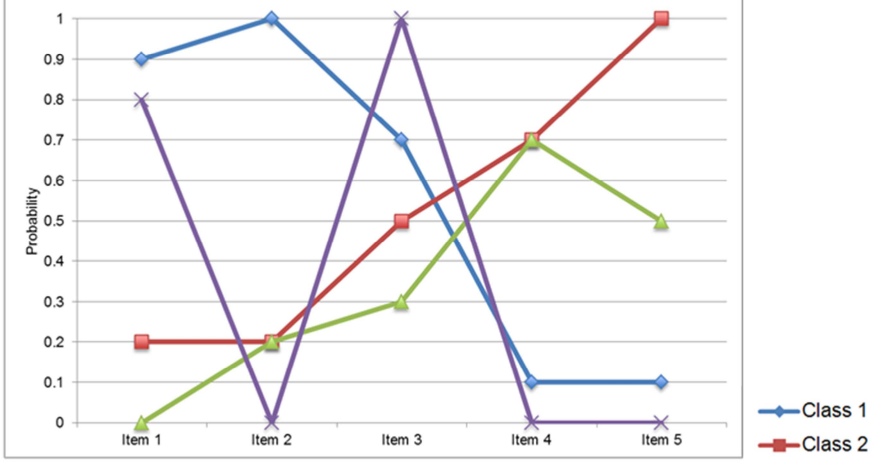
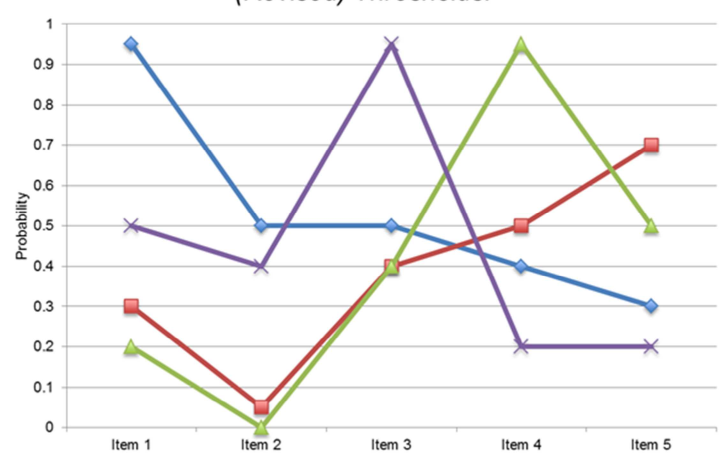


Figure 13. *Poorly Defined Model with Moderate (Revised) Thresholds.*



3.4.3 Latent transition probability matrices. Two transition probability matrices are used in this study. First, models with a completely equal transition probability matrix were examined, presented in Table 10. Henceforth this matrix will be referred to as "Even Transition Probabilities." The inclusion of this even transition probability matrix allowed us to see how varying other conditions impact the statistical power of latent transition probabilities, while holding transition probabilities constant.

Table 10: *Even Transition Probabilities Matrix*

		Time 2			
		Class 1	Class 2	Class 3	Class 4
Time 1	Class 1	.25	.25	.25	.25
	Class 2	.25	.25	.25	.25
	Class 3	.25	.25	.25	.25
	Class 4	.25	.25	.25	.25

A second of matrix of transition probabilities was considered which reflect the meta-analysis findings. None of the published LTA studies had completely equal transition probabilities and thus, a more realistic and representative matrix was necessary. This "Uneven Transition Probabilities" model, presented in Table 11, is based off of meta-analysis findings. Similar to the eating disorder, education, and substance abuse articles identified in the meta-analysis, the diagonals of the transition matrix are fairly stable with values greater than .70. Class 4 has the highest stability (.95). Moreover, the values were specified so that if an individual was to transition, they would transition to a latent class immediately above or below its original class, which can be seen in values immediately to the left or right of a diagonal transition probability, a pattern found in the meta-analysis as well.

Table 11: *Uneven Transition Probabilities Matrix*

		Time 2			
		Class 1	Class 2	Class 3	Class 4
Time 1	Class 1	.80	.15	.04	.01
	Class 2	.01	.85	.12	.02
	Class 3	.01	.07	.70	.22
	Class 4	.01	.02	.02	.95

3.4.4 Class sizes. Two different sets of class sizes at $t = 1$ were included in this study. First, completely equal class sizes were specified, so that we could examine constant while while varying other conditions. Each class had 25% of the overall sample size. The second set of class sizes were uneven. These values were based on the meta-analysis findings that showed that, for most eating disorder and education-related LTA studies, the first listed class had the highest percentage of the sample and the last class had the smallest percentage of the sample. Class percentages for both the even and uneven class sizes models can be seen in Table 12 below.

Table 12: *Class Percentages for Even and Uneven Class Sizes Models.*

Class at $t = 1$	Even Class Sizes Model	Uneven Class Sizes Model
Class 1	25%	50%
Class 2	25%	30%
Class 3	25%	15%
Class 4	25%	5%

3.5 Statistical Power in Mplus

Power values for each parameter are provided in Mplus output in column seven of the Model Results section when using the Monte Carlo facilities of the program. This column is labeled “% Sig Coeff” and represents the proportion of replications for which the null hypothesis that a parameter is equal to zero is rejected for each parameter at the .05 level. In a two-tailed test, the critical value is 1.96. The statistical test, or *z*-score, for each replication is the ratio of the parameter estimate ($\hat{\pi}$) to its standard error:

$$z = \frac{\hat{\pi} - 0}{s.e.(\hat{\pi})}.$$

The statistical power for the latent transition probabilities can be seen in the subsection labeled Categorical Latent Variables. A value of .80 or higher is considered adequate power (Cohen, 1988; Muthén & Muthén, 2002).

3.6 Expectation Maximization (EM) Algorithm

Monte Carlo simulation studies in Mplus use the Expectation Maximization (EM) Algorithm. Finding a mathematic solution in mixture modeling is difficult because the sample distribution is comprised of many sub-distributions (Jung & Wickrama, 2008). Ideally, in mixture modeling, researchers want global solutions. Global solutions are the set of parameters with the largest log likelihood out of all possible values. The likelihood function is the probability of an array of data given a set of parameter (Masyn, 2013). Local solutions are the solutions on which the estimation algorithm converges that is a local maxima, but not the global maximum.

To help explain the idea of global and local solutions of a likelihood function, Masyn (2013) uses the idea of a hiker (which represents the estimation algorithm) climbing a mountain (the likelihood function) with an ultimate goal of reaching the highest peak of the entire mountain range (the global solution). In order to do so, the hiker chooses a starting point (the initial starting point of the parameter estimates) and continues to hike until it is known that a peak has been found (convergence criterion has been met). The hiker keeps hiking and finds more peaks, but must eventually stop hiking when supplies have run out (maximum number of iterations have passed). It is also possible that during the entire trek, the hiker ran out of supplies (the maximum number of iterations has been exceeded) before finding a peak (failed to converge).

To extend this idea, a local maximum in LCA would represent a hiker who a found a peak (local maximum) then ran out of supplies, not knowing that if he/she had continued the trek, the highest peak (global maximum) was ahead. In the same manner, converging on a local maximum in LCA means that a solution has been found among a range of values, whereas if the range of values had been different, or larger, a better solution could have been found. Going back to the analogy of the hiker, if the hiker had started at a different point at the base of the mountain, the hiker could have found the highest peak before running out of supplies.

To help avoid reaching a local maximum, the researcher can indicate different start values in Mplus syntax, in essence setting out multiple hikers to find the global solution. The syntax,

```
TYPE = MIXTURE;
```

calls for random sets of start values to generate in Mplus using the default values. Ten iterations of 10 random sets of starting values are carried out. From this, the ending values with the highest loglikelihood are used in the last stage of optimization. For a more thorough investigation, Muthén (2008) recommends the following syntax when examining two classes:

```
STARTS = 50 5;
```
[1].

[1] If the "starts" syntax is not included in the input file, the Mplus default for mixture models is `STARTS = 10 2.`

The first number represents the number of random start values and the second number represents the number of final optimizations. Final optimizations optimize the specified number of best sets identified by the highest loglikelihood values after the first round of optimizations has been conducted. When there are more than two classes, Muthén (2008) recommends using the following start values because for more complex models, a more thorough investigation of solutions is necessary:

```
STARTS = 500 10;
STITERATIONS = 20;
```
[2]

In addition to using different start values, researchers can attempt to avoid specifying larger models and instead, aim for parsimony, as more parsimonious models are less complex. *Well-identified* models will arrive at a solution at any start value, *under-identified* models will find one global maximum and many local maxima, yet different start values may arrive at different solutions, and an *unidentified* model will find no unique solution (Collins & Lanza, 2010). Explained in further detail in Chapter 4, the poorly defined model had difficulty converging. Muthén's (2008) recommendations were considered when investigating this issue. Both well-defined and poorly defined models were compared across different sample sizes, looking at the Mplus default (`STARTS = 10 2`) and `STARTS = 500 10`. Power values and coverage values differed slightly between the two start values for both models. In both models, power values were slightly higher when start values were increased, yet differed by a maximum of .08. Additionally, coverage was slightly greater when starts were increased, yet differed only by a maximum of .07. Thus, the power values

[2] The Mplus default is `STITERATIONS = 10`.

in the remainder of this study represent the default Mplus start values and reflect a more conservative estimate of power and coverage.

3.7 Class Switching

Class switching is a common issue that may occur in LCA/LTA and LCA/LTA simulation studies. When an LCA or LTA model is run in Mplus, the latent classes are given an arbitrary class label. This is done for each run of an LCA model, meaning in one run the order of the classes may be different than in the second run with the exact model fit or meaningful difference in the modeling parameters. Similarly, in an LCA or LTA simulation study, class labels are at each permutation of a simulation and not over all permutations. Because of this, class labels may differ at each step of parameter estimation. For example, in a four-class model, the data generation values may appear at class 1 in the first permutation yet may appear at class 2 in the next permutation. This is an issue when aggregating parameter estimates over all simulation replications.

For the present study, parameter estimates were examined after simulations were complete. Mplus provides parameter estimates, standard errors, and fit statistics in a .csv file with the following command:

```
results = filename.csv;
```

Parameter estimates were examined for multiple simulation models and there was no occurrence of class switching thus no further action was necessary to correct for class switching.

3.8 Analysis Procedures

When determining what sample size is appropriate for adequate statistical power, Muthén and Muthén (2002) suggest three conditions regarding parameter estimate bias, standard error bias and coverage. The follow subsections will walk the reader through how to assess whether these conditions are met.

3.8.1 Parameter estimate bias. Parameter estimate bias should not exceed 10% for any parameter in the model (Muthén & Muthén, 2002). Parameter estimate bias can be calculated by finding the percent difference between the population value and the average parameter estimate over all replications. These two values can be found in the first and second column of Mplus output in the section labeled Model Results. The formula for calculating parameter estimate bias is expressed as:

$$\text{parameter estimate bias} = \frac{avg - population}{population}.$$

3.8.2 Standard error bias. Standard error bias should not exceed 10% for any parameter in the model (Muthén & Muthén, 2002). Furthermore, standard error bias should not exceed 5% for the parameter that is being examined for power. Standard error bias tends to be sensitive because standard errors are often overestimated or underestimated, which in turn affects confidence intervals and coverage. Standard error bias can be calculated by taking the percent difference between the population standard error and the average of the estimated standard errors for each parameter estimate over all replications. These values can be found in columns 3 and 4 of the Mplus output in the section labeled Model Results. The formula for calculating standard error bias is expressed as:

$$\text{standard error bias} = \frac{S.E.Average - std.dev.}{std.dev.}.$$

When the number of replications is large, the standard deviation of each parameter estimate over all replications is considered to be the population standard error and thus, this value is used for the population value.

3.8.3 Coverage. Lastly, coverage should be greater than .91 (Muthén & Muthén, 2002). Coverage is the proportion of replications for which the 95% confidence interval contains the true parameter value. In other words, at least 91% of replications should have true parameter values within the 95% confidence interval. Coverage values can be found in column 6 of Mplus output in the section labeled Model Results.

Considering these three conditions are met in this study, power values close greater than or equal to 0.80 will be considered adequate. In the next section, power values for all models are organized in Tables 16–19 and Tables 21–24. Power curves are also included in Figures 14–41 to visually show how power varies by sample size under each set of conditions.

Chapter 4 Results

4.1 General Overview

This chapter provides results from all 308 Monte Carlo simulations conducted as part of this study. The purpose of this dissertation was to examine the statistical power to detect latent transition probabilities under various conditions. These five various conditions included sample sizes, measurement models, latent transition probability matrices, class sizes, and threshold values. The following model types organize results into four sections:

1) Well-Defined Model with Even Transition Probabilities
2) Well-Defined Model with Uneven Transition Probabilities
3) Poorly Defined Model with Even Transition Probabilities
4) Poorly Defined Model with Uneven Transition Probabilities

Within each of these sections are two subsections summarizing results for even and uneven class sizes. These subsections also describe results for various sample sizes and threshold values. Recall that a .80 value of higher is considered adequate statistical power. Table 13 includes a summary of the percentage of latent transition probabilities that met or exceeded the .80 cutoff for each model type. All of the power values for each model can be seen in Tables 16–19 and Tables 21–24 and graphically in Figures 14–41. Before going into these sections, two unusual patterns in power curves are highlighted and are explained, including why a model may have artificially high power and when a model may not converge due to model characteristics such as small sample or low threshold values.

Table 13: *Percentage of Transition Probabilities that met .80 Cutoff*

Model Definition	Transition Probabilities	Class Sizes	Thresholds	*N* at which power stabilizes	*N*										
					100	250	500	1000	1250	1500	5000	6000	7000	8000	10000
Well-Defined	Even Transition Probabilities	Even Class Sizes	Thresholds ±1	1500	None	None	None	None	None	None	None	None	None	None	None
			Thresholds ±2	100	None	.92	All	All	All	All	All	All	All	All	All
			Thresholds ±3	100	.17	All	All	All	All	All	All	All	All	All	All
			Thresholds ±5	100	All	All	All	All	All	All	All	All	All	All	All
		Uneven Class Sizes	Thresholds ±1	5000	None	None	None	None	None	None	.42	.58	.67	.75	.75
			Thresholds ±2	250	None	.50	.75	.75	.83	All	All	All	All	All	All
			Thresholds ±3	100	.42	.75	.75	All	All	All	All	All	All	All	All
			Thresholds ±5	100	.50	.75	All	All	All	All	All	All	All	All	All
	Uneven Transition Probabilities	Even Class Sizes	Thresholds ±1	1250	None	.08	.25	.25	.25	.25	.33	.42	.42	.42	.42
			Thresholds ±2	1000	None	.25	.33	.42	.50	.58	.75	.75	.75	.75	.83
			Thresholds ±3	500	.17	.33	.50	.58	.58	.58	All	All	All	All	All
			Thresholds ±5	500	.25	.42	.50	.58	.75	.75	All	All	All	All	All
		Uneven Class Sizes	Thresholds ±1	5000	None	.08	.17	.25	.33	.33	.42	.42	.42	.50	.50
			Thresholds ±2	1500	.08	.33	.42	.50	.50	.50	.58	.58	.58	.58	.67
			Thresholds ±3	1500	.25	.42	.50	.58	.58	.58	.67	.75	.75	.75	.83
			Thresholds ±5	1500	.67	.42	.50	.58	.58	.58	.75	.92	.92	.92	.92
Poorly Defined	Even Transition Probabilities	Even Class Sizes	Moderate	5000	None	None	None	None	None	None	None	None	None	None	None
			Moderate (Revised)	1500	None	None	None	None	None	None	.67	.67	.67	.67	.92
			High	100	None	.08	25.00	.33	.42	.59	All	All	All	All	All
		Uneven Class Sizes	Moderate	-	None	None	None	None	None	None	.08	.08	.08	.17	.17
			Moderate (Revised)	6000	None	None	None	.08	.25	.25	.50	.67	.67	.67	.75
			High	250	None	.08	.17	.50	.67	.67	All	All	All	All	All
	Uneven Transition Probabilities	Even Class Sizes	Moderate	-	None	None	.08	.08	.08	.17	.25	.17	.17	.17	.17
			Moderate (Revised)	5000	None	.08	.25	.25	.25	.33	.42	.42	.42	.42	.42
			High	1250	None	.17	.33	.33	.33	.33	.67	.75	.75	.83	.92
		Uneven Class Sizes	Moderate	-	None	None	.08	.08	.08	.08	.17	.25	.17	.08	.25
			Moderate (Revised)	-	None	.17	.17	.25	.33	.33	.42	.42	.42	.42	.42
			High	5000	.08	.25	.25	.33	.42	.50	.50	.58	.58	.58	.58

Note. A hyphen (-) indicates that power did not stabilize for any *N* in that model.

Table 14: *Number of Completed Replications per Model*

Model Definition	Transition Probabilities	Class Sizes	Thresholds	N = 100	250	500	1000	1250	1500	5000	6000	7000	8000	10000
Well-Defined	Even Transition Probabilities	Even Class Sizes	Thresholds ±1	933	774	575	669	640	700	967	984	990	993	999
			Thresholds ±2	999	1000	1000	1000	1000	998	1000	1000	1000	1000	999
			Thresholds ±3	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
			Thresholds ±5	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		Uneven Class Sizes	Thresholds ±1	939	756	646	508	499	517	823	872	894	920	948
			Thresholds ±2	995	997	999	999	999	1000	1000	1000	1000	1000	1000
			Thresholds ±3	998	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
			Thresholds ±5	998	1000	1000	999	999	998	999	999	998	999	1000
	Uneven Transition Probabilities	Even Class Sizes	Thresholds ±1	933	774	575	669	640	700	967	984	990	993	999
			Thresholds ±2	999	1000	1000	1000	1000	998	1000	1000	1000	1000	999
			Thresholds ±3	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
			Thresholds ±5	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		Uneven Class Sizes	Thresholds ±1	939	756	646	508	499	517	823	872	894	920	948
			Thresholds ±2	995	997	999	999	999	1000	1000	1000	1000	1000	1000
			Thresholds ±3	873	995	1000	1000	1000	1000	1000	1000	1000	1000	1000
			Thresholds ±5	991	995	997	1000	999	998	1000	999	998	999	1000
Poorly Defined	Even Transition Probabilities	Even Class Sizes	Moderate	927	568	222	148	185	204	421	445	487	508	555
			Moderate (Revised)	918	648	472	531	602	684	965	986	988	996	995
			High	974	960	956	984	991	998	1000	1000	1000	1000	1000
		Uneven Class Sizes	Moderate	919	587	271	133	93	98	209	249	282	296	349
			Moderate (Revised)	906	754	505	424	454	456	626	687	708	736	808
			High	978	988	979	991	996	995	1000	1000	1000	1000	1000
	Uneven Transition Probabilities	Even Class Sizes	Moderate	921	696	414	188	132	115	16	16	13	14	7
			Moderate (Revised)	932	835	784	765	774	762	667	626	611	608	592
			High	892	970	966	985	983	988	1000	998	1000	998	1000
		Uneven Class Sizes	Moderate	915	703	439	222	167	120	25	22	20	16	11
			Moderate (Revised)	925	826	685	553	540	505	524	514	470	470	424
			High	761	928	928	982	986	987	997	1000	999	1000	999

Table 13 includes the sample size at which power stabilizes for all models. This table also includes the amount of latent transition probabilities that met or exceeded the recommended .80 cutoff. These results should be interpreted in conjunction with the power stability results. For example, a model with small *N* may have some latent transition probabilities with adequate power values; however, this model may not stabilize until a higher *N* is reached. Thus, it is advised that results are interpreted with caution and consideration.

4.2 Unusual Patterns in Power Curves

While compiling results, two unusual patterns emerged. Prior to going into deeper analysis of the power curves for each model type, these issues are addressed first to help aid the interpretation of subsequent results. First, one might expect that power would increase as *N* increases for all models. Contrary to this expectation, some models in this study returned unusual power curve patterns. Second, some power curves exhibited “spikes” where power would alternate in higher or lower power with each successive *N*. Reasons for these unusual patterns are explained below.

4.2.1 Artificially high power and stability of power. Some models exhibited an unusual pattern where power started higher at $N = 100$, decreased, then rose again when *N* increased (e.g., see Figure 25). This pattern is counterintuitive as it would imply that for some models, one can anticipate higher power for small *N*s (e.g., $N = 100$) and then again for large *N*s (e.g., $N = 10{,}000$), yet low power for moderate values of *N* (e.g., $N = 5{,}000$). To investigate this unexpected pattern, parameter estimates for each replication were examined to uncover possible issues that may be a result of model specification. This analysis focused on the Well-Defined Model with Uneven Transition Probability and Uneven Class Sizes (see

Table 19 and Figures 26–29). Each replication that endured a fixed standard error for a latent transition probability parameter had the following error message in Mplus output:

```
ONE OR MORE PROBABILITY PARAMETERS WERE FIXED TO AVOID
SINGULARITY OF THE INFORMATION MATRIX. THE SINGULARITY IS
MOST LIKELY BECAUSE THE MODEL IS NOT IDENTIFIED, OR
BECAUSE OF EMPTY CELLS IN THE JOINT DISTRIBUTION OF THE
CATEGORICAL LATENT VARIABLES AND ANY INDEPENDENT
VARIABLES.
```

Investigation of each replications parameters revealed that for smaller *N*s a large proportion of standard errors were being fixed to a very small number near zero. Statistical power is the proportion of replications where the ratio of the parameter estimate to its standard error is significant. Thus, when standard errors are fixed to numbers near zero, these *z*-scores tend to be greater than the 1.96 cutoff and are significant, causing artificial contributions to statistical power for that latent transition probability. They are considered artificial because the only reason they are significant is because they were fixed by the program to avoid estimation problems—which in fact are a signal that the model is not a good one in the first place. Henceforth, this outcome will be referred to as "artificially high power." "Stability" will refer to models where power values are not artificial (e.g., standard errors are not fixed by Mplus). Table 13 includes the sample size at which the model stabilizes. This is also visualized in Figures 14–41 by the dashed line. Power curves to the left of the dashed lines should not be interpreted.

Table 15 below includes the percentage of standard errors out of the first 100 replications that were fixed to zero for *N* = 100, 250, 500, 1,000, and 5,000 for the Well-Defined Model with Uneven Transition Probability and Uneven Class Sizes. When *N* = 100, 60% of standard errors are fixed to near zero. As *N* increases, less standard errors were fixed and by the time *N* reaches 5,000, 0% of standard errors were fixed to zero. Thus, when artificially high power estimates occur, power stabilizes as *N* increases because fewer standard errors are fixed. The Mplus error message indicates that Mplus fixes standard errors for two reasons: if a model is not well-identified or if cells are sparse. It appears that in this model standard errors are fixed due to sparseness because standard errors are only fixed when *N* is small.

Table 15: *Percentage of Standard Errors Fixed to Zero in Well-Defined Model with Uneven Transition Probability and Uneven Class Sizes*

N	% *SE*s set to 0
100	60%
250	50%
500	30%
1000	5%
5000	0%

4.2.2 Model non-convergence and the revised moderate model. Some power curves exhibited "spikes" where power would alternate in higher or lower power with each successive *N*. This pattern can be seen in the Poorly Defined Model with Uneven Transition Probabilities, Even Class Sizes, and Moderate Thresholds in Figure 36. Notice how the power curve for latent transition probability 1:1 starts at .32 for *N* = 100, decreases, then appears to stabilize at *N* = 1,500. Unlike the artificially high power values presented in the previous section, this power curve does not stabilize and does not gradually increase as *N* increases. Instead, it increases at *N* = 5,000, decreases at *N* = 6,000, increases for the next two values of *N*, then decreases again for *N* = 10,000. These "spikes" can also be seen in the Poorly Defined Model with Uneven Transition Probabilities, Uneven Class Sizes, and Moderate and Moderate (Revised) Thresholds in Figures 39 and 40.

A closer look at the output for these models revealed difficulty in model convergence. Table 14 includes the number of replications that completed for each of the model types. Well-defined models with thresholds greater than ±2 had little to no difficulty converging regardless of transition probability matrix or class size. When thresholds for well-defined models were set to ±1, lower and higher *N*s had a greater probability of converging. Moderate *N*s had lower proportions of completed replications. This can be attributed to the same reason why low values of *N* had greater power than higher or moderate values of *N*. For low *N* (e.g., *N* = 100), Mplus is more likely to fix standard errors to zero, which in turn "helps" the model to converge, resulting in high convergence rates. For moderate values of *N*, Mplus is less likely to "help" the model, resulting in moderate convergence rates. By the time *N* is large (e.g., *N* = 5,000), the model can sustain itself and successfully converge without the help of fixed standard errors.

Mplus provides the following error[3] for each replication that does not complete:

```
THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO A
CHANGE IN THE LOGLIKELIHOOD DURING THE LAST E STEP. AN
INSUFFICENT NUMBER OF E STEP ITERATIONS MAY HAVE BEEN
USED.  INCREASE THE NUMBER OF MITERATIONS OR INCREASE THE
MCONVERGENCE VALUE. ESTIMATES CANNOT BE TRUSTED. SLOW
CONVERGENCE DUE TO PARAMETER...
```

Following the suggestion in the error message, the number of iterations was increased from Mplus' default 10 to 100. To investigate difference in convergence rates, iterations were

[3] Mplus simulation output for models in this dissertation included other error messages. These error messages were investigated for potential issues. No major issues were found.

increased for the Poorly Defined model with Uneven Transition Probabilities and Uneven Class Sizes (see Figures 39–41) for this measurement model. For the moderate thresholds model, there was no change in the number of converged replications when the number of iterations was increased from 10 to 100. This shows the importance of having a well-defined measurement model. When measurement models have non-distinct classes, there is too much "noise" to distinguish between classes and as a result, the model will not converge. With non-simulated data, it is unlikely that Mplus would even resolve to this sort of measurement model because it is so poorly defined.

In summary, a model with poor measurement is unlikely to converge especially with the moderate thresholds specified in this simulation study. It may also indicate that a four-class LCA solution would not even emerge under these conditions. Thus, to investigate a similar yet revised model, thresholds were adjusted and henceforth will be referred to as the Moderate (Revised) Thresholds model (refer back to Figure 13). To increase class separation, each class had one distinguishable item. For example, for item 1, class 1 had a 95% probability of endorsing that item while the other three classes had moderate probabilities for that item. Similarly, class 2 had a very low probability (5%) of endorsing item 2 while the other three classes had moderate probabilities. The same pattern continues for class 3 with item 3 and class 4 with item 4. Lastly, item 5 would be considered a "bad" item with all classes maintaining a moderate probability. This revised moderate thresholds model was created with the hopes that a model with greater class and item distinction would have less difficulty converging over 1,000 replications. As seen in Table 14, this moderate (revised) model did in fact converge better than the moderate thresholds model at N = 10,000. With even transition probabilities and even class sizes, almost 100% of replications converged.

With even transition probabilities and uneven class sizes, the moderate (revised) thresholds model had 808 out of 1,000 replications completed. However, with uneven transition probabilities, only about half of replications completed for both even and uneven class sizes. This shows that there is still some difficulty for the moderate (revised) thresholds model to converge when there are uneven transition probabilities. This outcome is explained in further detail in the following subsections. This poorly defined model with moderate (revised) thresholds also serves as a comparison model to examine to what extent the measurement model has an effect on the power of latent transition probabilities. These results are provided in the model results subsections below.

4.3 Well-Defined Model with Even Transition Probabilities

4.3.1 Even class sizes. This model serves as a basis for understanding the power of latent transition probabilities. Power values can be seen in Table 16 and Figures 14–17. Creating a model that has equal transition probabilities and equal class sizes helped show that latent transition probabilities have nearly equal power regardless of if they are on the diagonal or off-diagonals in the latent transition probability matrix. For example, with thresholds of ±3 and N = 1,500, power values range from .92 to .95. This model also helps to show the effect of model measurement and logit thresholds while holding transition probabilities and class sizes constant and equal.

As seen in Figures 14–17, this well-defined model with even transition probabilities and even class sizes has stable power, meaning Mplus did not fix standard errors to zero and power values represent true and trustworthy results. Power is unstable only for $N < 1,500$ for thresholds of ±1. There are no other instances of “artificially high power.” This indicates that Mplus is unlikely to fix standard errors to zero in a well-defined model with even transition probabilities and even class sizes. This model also converged well, again with an exception at low thresholds of ±1 for $N < 1{,}500$.

A closer look at power across all four sets of thresholds helps show the impact of these values on power. When thresholds are set to ±1, adequate power is not met for any of the latent transition probabilities for any value of *N*. Even with a sample size of $N = 10{,}000$, power only reaches as high as .69. However, statistical power improves when thresholds are increased. For thresholds of ± 2 or ±3, there is adequate power for all transition probabilities when $N \geq 250$. For high thresholds of ±5, there is adequate power at all values of *N* included in this study.

Recall that for this simulation study, a well-defined model has four very distinct classes (refer to Figures 7–10 in Chapter 3). Thresholds of ±1 only reflect item-response probabilities of 27% and 73%, respectively, while thresholds of ±5 reflect 1% and 99%, respectively. We see here in these results that even with a well-defined model, if classes are not homogeneous (i.e., item-response probabilities near 0 or 1), it is unlikely that transition probabilities will have adequate power. The relationship between the measurement model and logit thresholds is important, as we see here that well-defined models are likely to have adequate power but only when thresholds are also high.

Thus far, this well-defined model with even transition probabilities and even class sizes shows that:

- Statistical power does not take into consideration whether a transition probability is on the diagonal or off-diagonal. All power values for transition probabilities are approximately equal within a sample size. In other words, holding all things constant, the power to detect probabilities is equal for movers and stayers.

- This model is fairly stable except when N is low with thresholds of ±1.
- Increasing threshold values helps the model to stabilize and results in greater power.

4.3.2 Uneven class sizes. This model only differs from the above model by its class sizes. Transition probabilities are still even across all cells and the measurement model is congruent with the previous model. However, the power curves of these two model types are very different (see Table 17, Figures 18–21). In the previous even class size model, all conditions are set equal to each other and thus, all transition probabilities have nearly equal power within a sample size. For this uneven class sizes model, we see differences between each latent transition probability. For a single N, some power values are high while others are low. In general, there is lower power for latent transition probabilities with small class sizes. For example, power is lower when transitioning out of, or staying in class 4 (which had only 5% of the sample), versus transitioning out of or staying in class 1 (which had 50% of the sample). When comparing this model to the previous even class sizes model, the effect of sparseness of cells is evident in the low power of the transition probabilities. Cells with fewer individuals had less power. For adequate power for *all* transition probabilities in the transition matrix, results indicate that we need $N \geq 5{,}000$ with thresholds of ±2, $N \geq 1{,}000$ with thresholds of ±3, or $N \geq 500$ with thresholds ±5.

With low thresholds, this model has some difficulty stabilizing, meaning standard errors were fixed for many *N*s and thus, power values are "artificially high." For thresholds of ±1, power stabilizes after $N = 5{,}000$. When thresholds are increased to ±2, power stabilizes at $N = 250$. When thresholds are ±3 or ±5, power is stable for all N. Caution should be taken when interpreting power results for this model.

In summary, this well-defined model with even transition probabilities and uneven class sizes contributes the following findings in addition to what we have seen in results thus far:

- Sparseness negatively affects statistical power. Larger class sizes have greater power compared to smaller latent class sizes.
- Transition probabilities from smaller class sizes had more difficulty stabilizing, as expected.

Table 16: *Power Values for Well-Defined Model with Even Transition Probabilities and Even Class Sizes*

	N	C1 to C1	C2 to C1	C3 to C1	C4 to C1	C1 to C2	C2 to C2	C3 to C2	C4 to C2	C1 to C3	C2 to C3	C3 to C3	C4 to C3
Thresholds ±1	100	0.42	0.36	0.37	0.40	0.36	0.44	0.39	0.43	0.35	0.38	0.44	0.44
	250	0.39	0.30	0.33	0.34	0.32	0.42	0.32	0.38	0.29	0.33	0.39	0.38
	500	0.47	0.34	0.37	0.35	0.38	0.44	0.36	0.40	0.33	0.37	0.45	0.45
	1000	0.52	0.45	0.49	0.46	0.46	0.50	0.44	0.48	0.44	0.47	0.56	0.51
	1250	0.57	0.52	0.51	0.55	0.53	0.57	0.52	0.54	0.52	0.51	0.62	0.54
	1500	0.56	0.49	0.53	0.54	0.48	0.53	0.52	0.54	0.51	0.51	0.58	0.53
	5000	0.59	0.59	0.60	0.65	0.62	0.62	0.64	0.63	0.63	0.66	0.65	0.67
	6000	0.62	0.62	0.63	0.67	0.63	0.62	0.66	0.64	0.62	0.66	0.63	0.64
	7000	0.63	0.62	0.64	0.66	0.64	0.62	0.67	0.66	0.66	0.68	0.66	0.68
	8000	0.62	0.65	0.65	0.67	0.66	0.66	0.69	0.67	0.66	0.67	0.65	0.67
	10000	0.63	0.64	0.64	0.66	0.66	0.65	0.67	0.67	0.67	0.69	0.67	0.69
Thresholds ±2	100	0.55	0.51	0.54	0.54	0.50	0.56	0.50	0.53	0.53	0.50	0.57	0.51
	250	0.82	0.80	0.81	0.86	0.81	0.80	0.85	0.81	0.81	0.83	0.82	0.81
	500	0.87	0.85	0.87	0.90	0.87	0.87	0.92	0.88	0.86	0.89	0.89	0.91
	1000	0.86	0.87	0.88	0.90	0.87	0.88	0.89	0.90	0.87	0.89	0.91	0.91
	1250	0.88	0.87	0.90	0.92	0.85	0.87	0.89	0.90	0.87	0.89	0.89	0.90
	1500	0.88	0.88	0.90	0.92	0.87	0.89	0.91	0.91	0.88	0.90	0.90	0.94
	5000	0.89	0.89	0.90	0.92	0.88	0.91	0.90	0.91	0.89	0.91	0.92	0.93
	6000	0.88	0.88	0.89	0.91	0.90	0.90	0.90	0.90	0.89	0.89	0.90	0.90
	7000	0.89	0.88	0.89	0.90	0.88	0.88	0.89	0.89	0.88	0.89	0.89	0.91
	8000	0.90	0.91	0.91	0.92	0.90	0.91	0.91	0.91	0.90	0.90	0.91	0.91
	10000	0.87	0.88	0.88	0.89	0.90	0.90	0.89	0.90	0.89	0.89	0.90	0.91
Thresholds ±3	100	0.80	0.75	0.75	0.79	0.77	0.82	0.79	0.79	0.78	0.79	0.78	0.76
	250	0.96	0.93	0.96	0.95	0.94	0.95	0.95	0.94	0.94	0.96	0.94	0.94
	500	0.95	0.94	0.95	0.96	0.94	0.96	0.96	0.96	0.95	0.97	0.97	0.96
	1000	0.93	0.92	0.93	0.96	0.92	0.94	0.95	0.94	0.92	0.94	0.95	0.96
	1250	0.94	0.93	0.94	0.96	0.90	0.93	0.93	0.94	0.91	0.92	0.94	0.95
	1500	0.93	0.92	0.93	0.95	0.92	0.94	0.94	0.95	0.92	0.95	0.95	0.94
	5000	0.92	0.93	0.94	0.96	0.92	0.93	0.94	0.95	0.92	0.94	0.94	0.95
	6000	0.94	0.94	0.95	0.96	0.93	0.94	0.94	0.95	0.93	0.95	0.95	0.95
	7000	0.94	0.94	0.95	0.95	0.94	0.95	0.95	0.96	0.94	0.95	0.96	0.96
	8000	0.95	0.95	0.95	0.96	0.94	0.95	0.95	0.96	0.95	0.95	0.96	0.96
	10000	0.94	0.94	0.95	0.95	0.95	0.96	0.96	0.96	0.95	0.96	0.96	0.96
Thresholds ±5	100	0.87	0.84	0.87	0.87	0.87	0.88	0.88	0.87	0.88	0.86	0.88	0.85
	250	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99	1.00
	500	1.00	0.99	1.00	1.00	0.99	0.99	1.00	0.99	0.99	1.00	1.00	1.00
	1000	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	0.99	1.00
	1250	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.99
	1500	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99
	5000	0.99	0.98	0.98	0.99	0.97	0.98	0.98	0.98	0.97	0.98	0.98	1.00
	6000	0.98	0.98	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.99	0.99
	7000	0.98	0.97	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.99
	8000	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.99
	10000	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99

Figure 14. *Well-Defined Model/Even Transition Probabilities/Even Class Sizes/Thresholds ±1.*

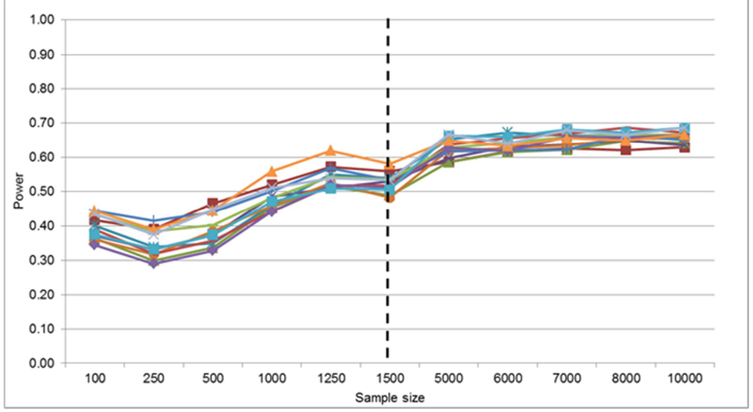
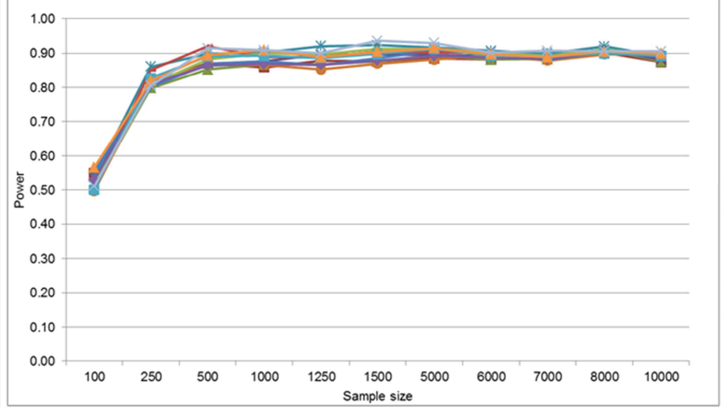


Figure 15. *Well-Defined Model/Even Transition Probabilities/Even Class Sizes/Thresholds ±2.*



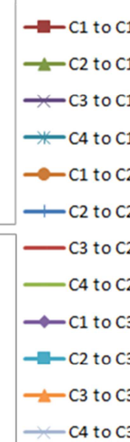


Figure 16. *Well-Defined Model/Even Transition Probabilities/Even Class Sizes/Thresholds ±3.*

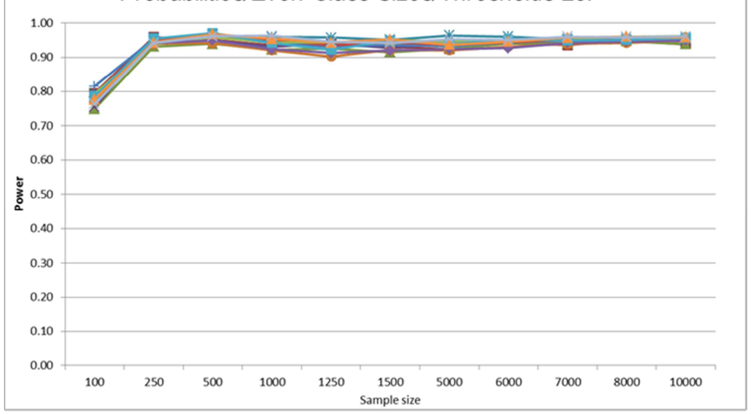


Figure 17. *Well-Defined Model/Even Transition Probabilities/Even Class Sizes/Thresholds ±5.*

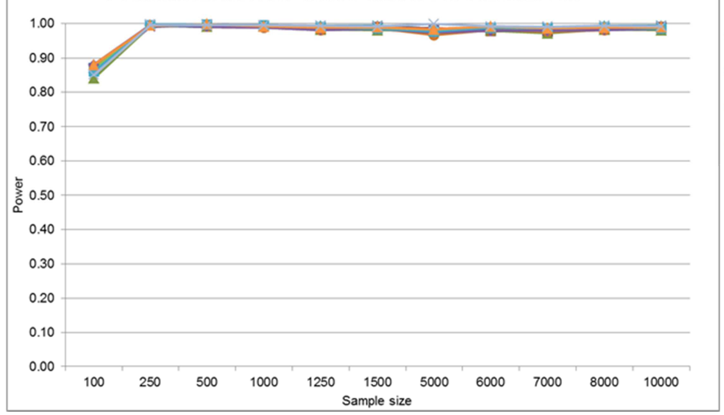


Table 17: *Power Values for Well-Defined Model with Even Transition Probabilities and Uneven Class Sizes*

		1 1	2 1	3 1	4 1	1 2	2 2	3 2	4 2	1 3	2 3	3 3	4 3
Thresholds ±1	100	0.43	0.36	0.41	0.50	0.38	0.41	0.44	0.51	0.41	0.40	0.43	0.55
	250	0.41	0.31	0.31	0.35	0.36	0.41	0.32	0.37	0.40	0.37	0.36	0.38
	500	0.49	0.34	0.30	0.29	0.48	0.44	0.27	0.28	0.44	0.38	0.40	0.31
	1000	0.69	0.54	0.34	0.21	0.57	0.50	0.30	0.25	0.53	0.43	0.39	0.22
	1250	0.68	0.57	0.34	0.12	0.57	0.53	0.30	0.18	0.54	0.48	0.42	0.15
	1500	0.77	0.63	0.40	0.12	0.59	0.52	0.34	0.16	0.55	0.47	0.42	0.13
	5000	0.99	0.99	0.83	0.26	0.83	0.76	0.71	0.30	0.80	0.78	0.62	0.20
	6000	1.00	1.00	0.90	0.34	0.88	0.83	0.79	0.34	0.85	0.84	0.66	0.22
	7000	1.00	1.00	0.94	0.40	0.92	0.88	0.87	0.39	0.89	0.88	0.76	0.31
	8000	1.00	1.00	0.97	0.48	0.96	0.91	0.91	0.43	0.93	0.92	0.81	0.36
	10000	1.00	1.00	0.99	0.59	0.98	0.96	0.97	0.52	0.98	0.97	0.89	0.47
Thresholds ±2	100	0.68	0.52	0.35	0.39	0.73	0.62	0.39	0.42	0.74	0.59	0.45	0.42
	250	0.93	0.83	0.61	0.20	0.88	0.81	0.62	0.25	0.89	0.85	0.65	0.26
	500	0.99	0.98	0.93	0.39	0.94	0.91	0.86	0.38	0.94	0.94	0.85	0.38
	1000	1.00	0.99	0.99	0.73	0.96	0.95	0.96	0.65	0.96	0.96	0.93	0.64
	1250	1.00	1.00	0.99	0.84	0.98	0.98	0.97	0.76	0.98	0.98	0.96	0.78
	1500	1.00	0.99	1.00	0.91	0.98	0.98	0.98	0.83	0.98	0.98	0.97	0.82
	5000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	6000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	7000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	8000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Thresholds ±3	100	0.93	0.80	0.48	0.28	0.95	0.85	0.54	0.30	0.95	0.86	0.53	0.30
	250	0.99	0.98	0.88	0.37	0.98	0.96	0.90	0.34	0.97	0.97	0.89	0.37
	500	1.00	1.00	0.99	0.77	0.98	0.98	0.97	0.73	0.98	0.99	0.97	0.73
	1000	1.00	1.00	1.00	0.98	0.99	0.99	0.98	0.95	0.99	0.99	0.98	0.95
	1250	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.98
	1500	1.00	1.00	0.99	1.00	0.99	1.00	0.99	0.99	0.99	1.00	0.99	0.98
	5000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	6000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	7000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	8000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Thresholds ±5	100	0.99	0.94	0.57	0.25	1.00	0.93	0.60	0.25	1.00	0.94	0.59	0.24
	250	1.00	1.00	0.97	0.47	1.00	1.00	0.98	0.48	1.00	1.00	0.98	0.49
	500	0.96	0.94	0.93	0.92	0.95	0.95	0.92	0.92	0.95	0.94	0.93	0.92
	1000	0.96	0.96	0.95	0.93	0.95	0.94	0.95	0.94	0.95	0.95	0.96	0.93
	1250	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	6000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	7000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	8000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Figure 18. *Well-Defined Model/Even Transition Probabilities/Uneven Class Sizes/Thresholds ±1.*

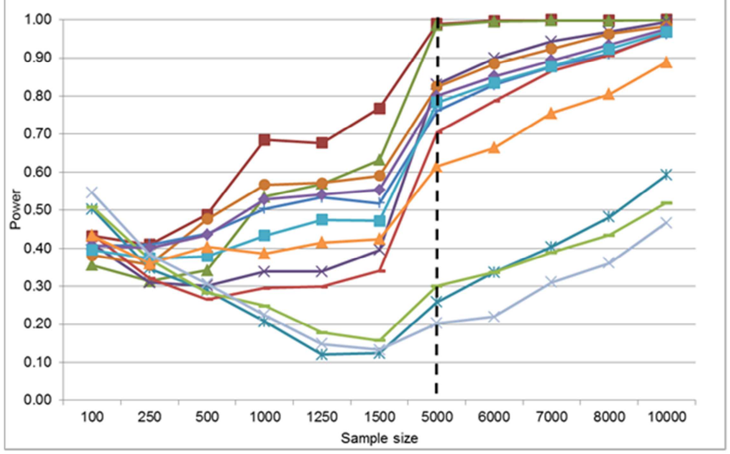
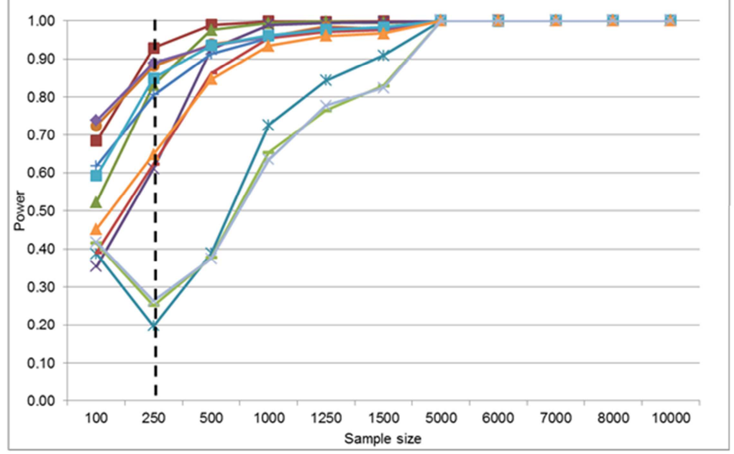


Figure 19. *Well-Defined Model/Even Transition Probabilities/Uneven Class Sizes/Thresholds ±2.*



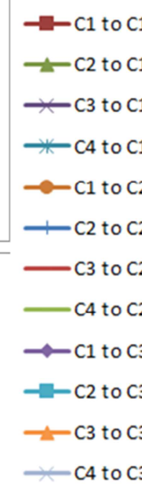


Figure 20. *Well-Defined Model/Even Transition Probabilities/Uneven Class Sizes/Thresholds ±3.*

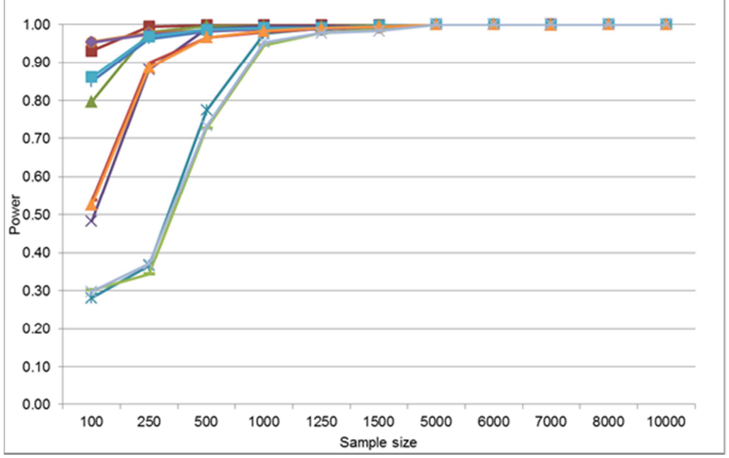
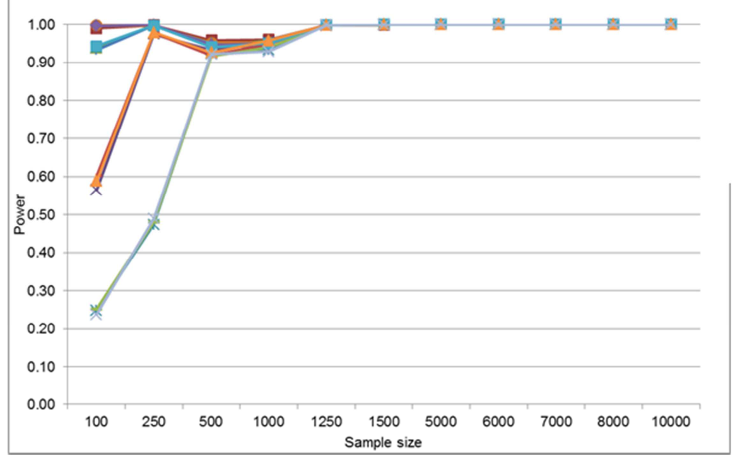


Figure 21. *Well-Defined Model/Even Transition Probabilities/Uneven Class Sizes/Thresholds ±5.*



4.4 Well-Defined Model with Uneven Transition Probabilities

4.4.1 Even class sizes. This model differs from the initial model in section 4.3.1 only by its transition probabilities. When compared to that model, one can clearly see the impact that uneven transition probabilities have on statistical power and stability (see Table 18 and Figures 22–25). For all four sets of threshold values, there are stability issues. Standard errors were fixed for many of these models, resulting in "artificially high" power. Power stabilizes at $N \geq 1{,}250$ for threshold ±1, $N \geq 1500$ for thresholds ±2, and $N \geq 500$ for high thresholds of ±3 or ±5. It should be noted that only some of the transition probabilities have difficulty stabilizing. The transition probabilities that start with high power and stay high are diagonal transition probabilities, which did not have difficulty stabilizing. These transition probabilities are much larger than the off-diagonal transition probabilities (e.g., a probability of .85 versus .04). The power curves on the lower half of the curve are power values for off-diagonal transition probabilities. Larger off-diagonal transition probabilities (e.g., a probability of .15) are more likely to stabilize sooner than smaller off-diagonal transition probabilities (e.g., a probability of .01). In sum, larger transition probabilities have higher power and are easier to stabilize, as we would expect.

In the similar model with even transition probabilities in section 4.3.1, all power values were nearly equal for each value of *N*. In this model we could expect that if every condition was held equal, power would be roughly equal. However, in this model with uneven transition probabilities, some power values are low while others are high. We see here that power is related to the value of the transition probability. As seen before, larger transition probabilities leads to greater power.

This condition also displays the influence that the size of the threshold values has on power, specifically that larger thresholds lead to better power. When thresholds are ±1 or ±2, there is adequate power for some, but not all, of the transition probabilities. Low transition probabilities (i.e., .01) never reach power of .80 when thresholds are ±1. When thresholds are higher (i.e., ±3, ±5), there is adequate power for all transition probabilities at $N \geq 5{,}000$. For $N < 5{,}000$, some transition probabilities have adequate power whereas others do not (see Figures 24 and 25). Low transition probabilities (i.e., .01) tend to have power values less than the .80 cut-off for small *N*s.

In summary, this well-defined model with uneven transition probabilities and even class sizes contributes the following findings in addition to what we have seen in results thus far:

- Larger transition probabilities have less difficulty stabilizing than smaller transition probability. Diagonal values in this model had no difficulty stabilizing.
- Larger transition probabilities have greater power than smaller transition probabilities.

4.4.2 Uneven class sizes. Power values for this model were less stable than in the similar model above with even class sizes (see Table 19 and Figures 26–29). The combination of small transition probabilities and small class sizes resulted in artificially high power, or unstable power values. In general, power stabilizes at sample sizes between N = 1,500 and N = 5,000 for all thresholds. Similar to the previous subsection, not all transition probabilities in a single model have difficulty stabilizing. Larger transition probabilities stabilize sooner than smaller transition probabilities. Larger transition probabilities (typically diagonal probabilities) are stable at $N = 100$. Additionally, off-diagonal transition probabilities are more likely to stabilize sooner if coming from a larger class size such as class 1 or 2.

This is the first instance that we see a combination of uneven transition probabilities and uneven class sizes. This model shows the interplay of these two conditions on statistical power. Power is lower for smaller transition probabilities or smaller class sizes (e.g., classes 3 and 4). In fact, the combination of these two attributes creates a small cell size which in turn results in even smaller power. Some transition probabilities reach adequate power, but even with thresholds of ±5 and high *N*s, not all transition probabilities reached power values of .80. Transition probabilities that do have high power are for classes that are large at $t = 1$ and/or had diagonal transition probabilities.

In summary, this well-defined model with uneven transition probabilities and uneven class sizes contributes the following findings in addition to what we have seen in results thus far:

- The combination of small transition probability and small class size results in even lower power.

Table 18: *Power Values for Well-Defined Model with Uneven Transition Probabilities and Even Class Sizes*

	N	1 1	2 1	3 1	4 1	1 2	2 2	3 2	4 2	1 3	2 3	3 3	4 3
Thresholds ±1	100	0.54	0.37	0.45	0.45	0.28	0.59	0.40	0.48	0.38	0.35	0.58	0.46
	250	0.80	0.27	0.30	0.28	0.19	0.82	0.23	0.33	0.25	0.21	0.76	0.29
	500	0.95	0.16	0.21	0.13	0.22	0.95	0.15	0.19	0.13	0.13	0.90	0.16
	1000	0.98	0.08	0.08	0.06	0.37	0.98	0.10	0.09	0.07	0.18	0.97	0.06
	1250	0.99	0.08	0.06	0.06	0.48	0.99	0.09	0.06	0.06	0.23	0.99	0.05
	1500	0.98	0.07	0.06	0.05	0.52	0.98	0.12	0.06	0.10	0.25	0.98	0.08
	5000	0.99	0.06	0.04	0.06	0.95	1.00	0.33	0.07	0.18	0.73	1.00	0.06
	6000	0.99	0.08	0.04	0.05	0.97	0.99	0.35	0.10	0.22	0.80	1.00	0.08
	7000	0.99	0.06	0.04	0.05	0.97	0.99	0.41	0.09	0.29	0.84	0.99	0.07
	8000	1.00	0.07	0.04	0.06	0.98	1.00	0.47	0.12	0.31	0.90	1.00	0.08
	10000	1.00	0.07	0.05	0.07	0.98	1.00	0.58	0.13	0.39	0.95	1.00	0.11
Thresholds ±2	100	0.64	0.34	0.41	0.32	0.17	0.69	0.22	0.38	0.23	0.14	0.77	0.34
	250	0.99	0.18	0.22	0.12	0.64	0.98	0.15	0.17	0.14	0.35	0.99	0.16
	500	0.97	0.09	0.08	0.06	0.95	0.99	0.40	0.07	0.28	0.77	0.99	0.07
	1000	1.00	0.06	0.06	0.05	1.00	1.00	0.80	0.19	0.66	0.98	1.00	0.15
	1250	1.00	0.06	0.07	0.06	1.00	1.00	0.89	0.26	0.78	1.00	1.00	0.20
	1500	1.00	0.08	0.09	0.08	1.00	1.00	0.95	0.36	0.86	1.00	1.00	0.27
	5000	1.00	0.37	0.50	0.19	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.85
	6000	1.00	0.49	0.61	0.23	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.92
	7000	1.00	0.54	0.70	0.29	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.95
	8000	1.00	0.64	0.75	0.33	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
	10000	1.00	0.73	0.87	0.30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Thresholds ±3	100	0.78	0.50	0.57	0.51	0.29	0.80	0.23	0.50	0.27	0.19	0.83	0.45
	250	0.94	0.23	0.27	0.22	0.81	0.94	0.36	0.18	0.18	0.71	0.97	0.16
	500	1.00	0.12	0.11	0.08	1.00	0.99	0.83	0.17	0.58	0.98	0.99	0.13
	1000	1.00	0.11	0.13	0.10	1.00	1.00	0.99	0.49	0.95	1.00	1.00	0.44
	1250	1.00	0.18	0.19	0.16	1.00	1.00	1.00	0.68	0.98	1.00	1.00	0.59
	1500	1.00	0.25	0.28	0.22	1.00	1.00	1.00	0.78	0.99	1.00	1.00	0.73
	5000	1.00	0.94	0.96	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	6000	1.00	0.97	0.98	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	7000	1.00	0.98	0.99	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	8000	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10000	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Thresholds ±5	100	1.00	0.65	0.64	0.68	0.53	1.00	0.28	0.56	0.35	0.37	1.00	0.55
	250	1.00	0.22	0.20	0.26	0.98	1.00	0.61	0.19	0.29	0.93	1.00	0.19
	500	1.00	0.11	0.10	0.10	1.00	1.00	0.97	0.24	0.73	1.00	1.00	0.25
	1000	1.00	0.23	0.25	0.26	1.00	1.00	1.00	0.71	0.99	1.00	1.00	0.71
	1250	1.00	0.35	0.37	0.40	1.00	1.00	1.00	0.85	1.00	1.00	1.00	0.86
	1500	1.00	0.47	0.51	0.51	1.00	1.00	1.00	0.93	1.00	1.00	1.00	0.93
	5000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	6000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	7000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	8000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Figure 22. *Well-Defined Model/Uneven Transition Probabilities/Even Class Sizes/Thresholds ±1.*

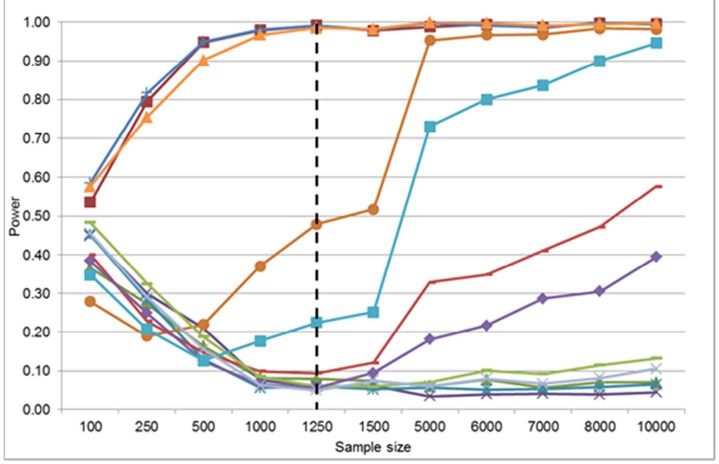
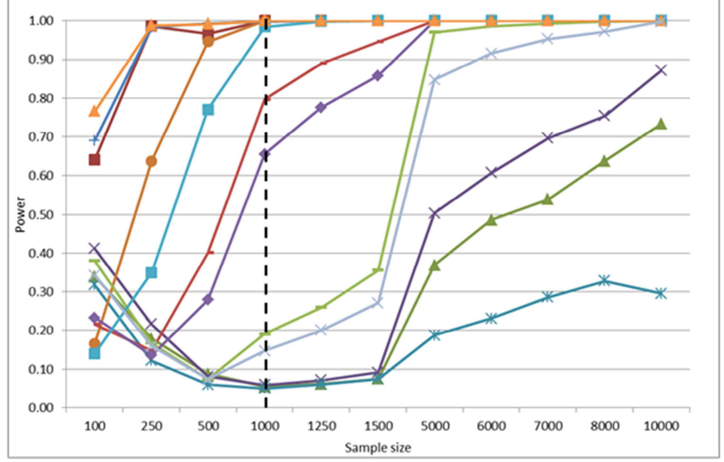


Figure 23. *Well-Defined Model/Uneven Transition Probabilities/Even Class Sizes/Thresholds ±2.*



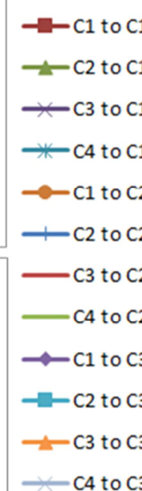


Figure 24. *Well-Defined Model/Uneven Transition Probabilities/Even Class Sizes/Thresholds ±3.*

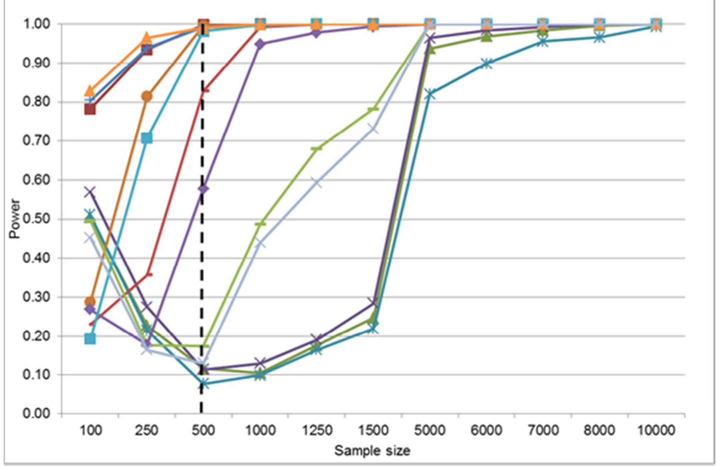


Figure 25. *Well-Defined Model/Uneven Transition Probabilities/Even Class Sizes/Thresholds ±5.*

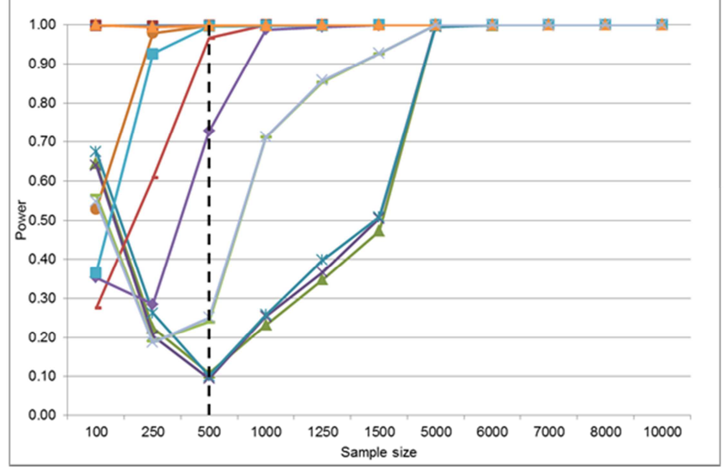


Table 19: *Power Values for Well-Defined Model with Uneven Transition Probabilities and Uneven Classes*

	N	1 1	2 1	3 1	4 1	1 2	2 2	3 2	4 2	1 3	2 3	3 3	4 3
Thresholds ±1	100	0.73	0.37	0.46	0.48	0.26	0.57	0.40	0.49	0.31	0.36	0.46	0.51
	250	0.85	0.34	0.44	0.43	0.27	0.79	0.35	0.43	0.23	0.27	0.61	0.43
	500	0.93	0.24	0.35	0.36	0.45	0.89	0.30	0.38	0.19	0.22	0.75	0.34
	1000	0.99	0.21	0.22	0.22	0.74	0.98	0.18	0.28	0.15	0.23	0.89	0.24
	1250	0.99	0.15	0.21	0.20	0.81	0.99	0.16	0.24	0.14	0.27	0.92	0.20
	1500	0.99	0.15	0.19	0.17	0.89	0.99	0.15	0.19	0.16	0.32	0.93	0.19
	5000	1.00	0.06	0.10	0.08	0.99	1.00	0.15	0.10	0.57	0.87	0.99	0.10
	6000	1.00	0.07	0.09	0.09	0.99	1.00	0.16	0.10	0.66	0.92	0.99	0.11
	7000	1.00	0.07	0.08	0.07	0.99	0.99	0.18	0.08	0.74	0.94	0.99	0.09
	8000	1.00	0.07	0.08	0.05	1.00	1.00	0.23	0.08	0.83	0.97	1.00	0.09
	10000	1.00	0.07	0.05	0.06	1.00	1.00	0.26	0.07	0.89	0.98	1.00	0.08
Thresholds ±2	100	0.86	0.39	0.56	0.47	0.39	0.76	0.31	0.55	0.23	0.23	0.63	0.54
	250	0.98	0.30	0.39	0.30	0.92	0.96	0.17	0.38	0.24	0.46	0.94	0.42
	500	1.00	0.17	0.21	0.16	0.99	1.00	0.20	0.23	0.63	0.86	0.99	0.24
	1000	1.00	0.09	0.08	0.06	1.00	1.00	0.46	0.12	0.97	0.99	1.00	0.14
	1250	1.00	0.11	0.06	0.04	1.00	1.00	0.57	0.08	0.99	1.00	1.00	0.12
	1500	1.00	0.12	0.05	0.04	1.00	1.00	0.65	0.08	1.00	1.00	1.00	0.10
	5000	1.00	0.48	0.22	0.04	1.00	1.00	1.00	0.10	1.00	1.00	1.00	0.09
	6000	1.00	0.55	0.28	0.05	1.00	1.00	1.00	0.14	1.00	1.00	1.00	0.10
	7000	1.00	0.65	0.36	0.05	1.00	1.00	1.00	0.16	1.00	1.00	1.00	0.15
	8000	1.00	0.71	0.44	0.05	1.00	1.00	1.00	0.18	1.00	1.00	1.00	0.18
	10000	1.00	0.81	0.55	0.05	1.00	1.00	1.00	0.25	1.00	1.00	1.00	0.23
Thresholds ±3	100	0.91	0.53	0.75	0.76	0.72	0.92	0.38	0.77	0.20	0.32	0.92	0.77
	250	0.98	0.33	0.49	0.49	0.96	0.96	0.25	0.54	0.50	0.83	0.94	0.53
	500	0.99	0.20	0.28	0.29	0.99	1.00	0.48	0.35	0.92	0.99	0.99	0.36
	1000	1.00	0.18	0.12	0.09	1.00	1.00	0.90	0.15	1.00	1.00	1.00	0.13
	1250	1.00	0.26	0.11	0.06	1.00	1.00	0.98	0.11	1.00	1.00	1.00	0.11
	1500	1.00	0.36	0.11	0.04	1.00	1.00	0.99	0.11	1.00	1.00	1.00	0.11
	5000	1.00	0.97	0.72	0.07	1.00	1.00	1.00	0.42	1.00	1.00	1.00	0.35
	6000	1.00	0.99	0.83	0.10	1.00	1.00	1.00	0.53	1.00	1.00	1.00	0.46
	7000	1.00	1.00	0.92	0.10	1.00	1.00	1.00	0.62	1.00	1.00	1.00	0.58
	8000	1.00	1.00	0.94	0.11	1.00	1.00	1.00	0.72	1.00	1.00	1.00	0.62
	10000	1.00	1.00	1.00	0.15	1.00	1.00	1.00	0.83	1.00	1.00	1.00	0.80
Thresholds ±5	100	1.00	0.68	0.83	0.92	0.93	1.00	0.39	0.89	0.25	0.50	1.00	0.88
	250	1.00	0.37	0.57	0.75	1.00	1.00	0.33	0.69	0.73	0.98	1.00	0.70
	500	1.00	0.21	0.32	0.55	1.00	1.00	0.77	0.48	0.98	1.00	1.00	0.49
	1000	1.00	0.35	0.18	0.26	1.00	1.00	0.98	0.22	1.00	1.00	1.00	0.21
	1250	1.00	0.50	0.16	0.17	1.00	1.00	0.99	0.15	1.00	1.00	0.99	0.16
	1500	1.00	0.64	0.21	0.11	1.00	1.00	1.00	0.13	1.00	1.00	1.00	0.13
	5000	1.00	1.00	0.91	0.21	1.00	1.00	1.00	0.72	1.00	1.00	1.00	0.70
	6000	1.00	1.00	1.00	0.25	1.00	1.00	1.00	0.82	1.00	1.00	1.00	0.83
	7000	1.00	1.00	0.98	0.31	1.00	1.00	1.00	0.90	1.00	1.00	1.00	0.89
	8000	1.00	1.00	0.99	0.38	1.00	1.00	1.00	0.95	1.00	1.00	1.00	0.95
	10000	1.00	1.00	1.00	0.57	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.98

Figure 26. *Well-Defined Model/Uneven Transition Probabilities/Uneven Class Sizes/Thresholds ±1.*

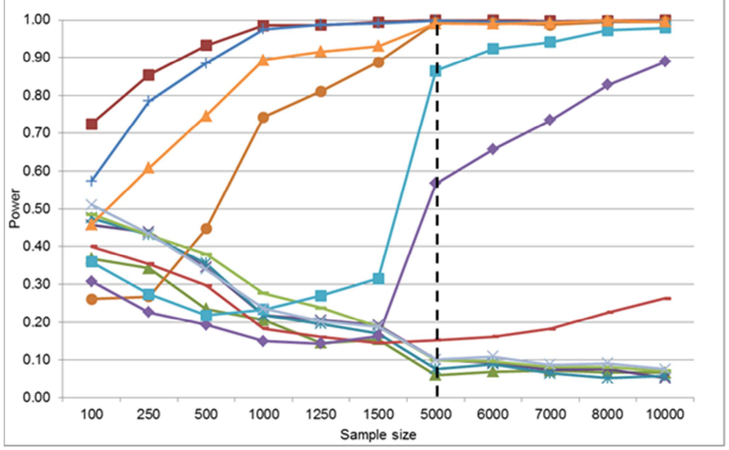
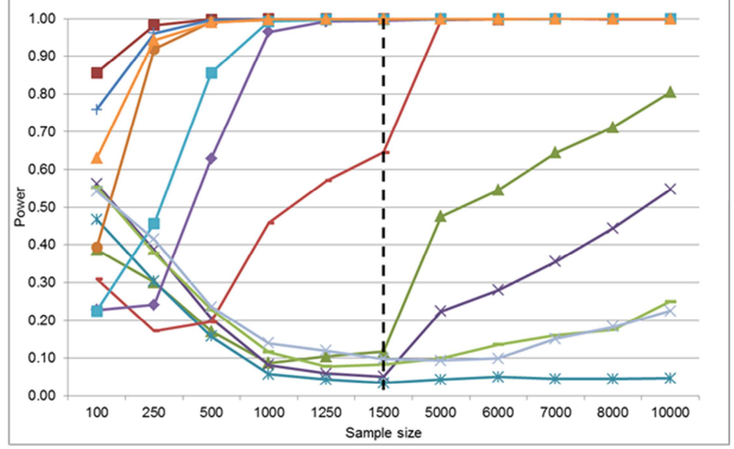


Figure 27. Well-Defined Model/*Uneven Transition Probabilities/Uneven Class Sizes/Thresholds ±2.*



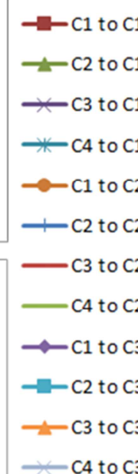


Figure 28. *Well-Defined Model/Uneven Transition Probabilitiesl/Uneven Class Sizes/Thresholds ±3.*

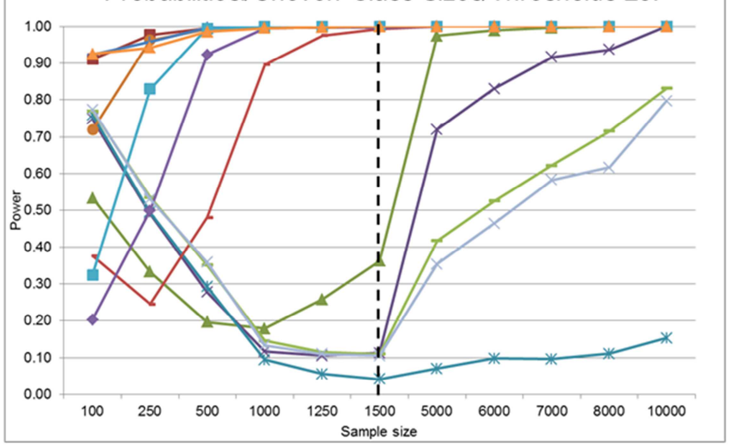
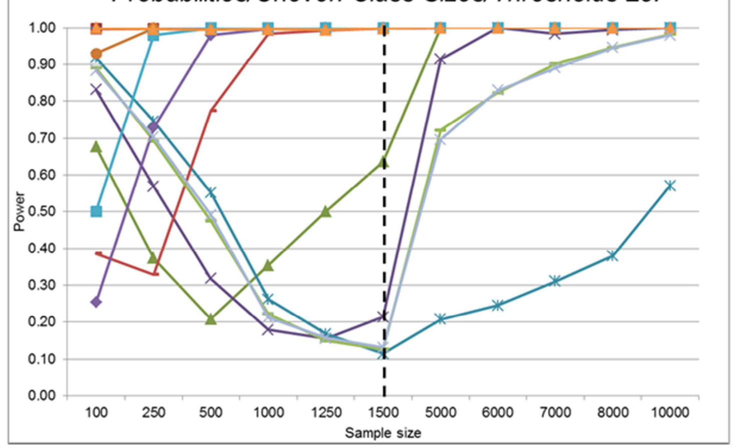


Figure 29. *Well-Defined Model/Uneven Transition Probabilities/Uneven Class Sizes/Thresholds ±5.*



4.5 Poorly Defined Model with Even Transition Probabilities

4.5.1 Even class sizes. This poorly defined model with even transition probabilities and even class sizes model serves as a basis to compare all other variations of this model. See Table 21 and Figures 30–32 for power values. In the well-defined model with even transition probabilities and even class sizes, power values were nearly equal for each transition probability at each *N* because all conditions were equal. The well-defined model had high class separation. Because each class was clearly defined, there were marginal differences between the power values of transition probabilities among the classes.

Unlike the well-defined model, the poorly defined model has some variance in power values depending on which classes the transition probability were attributed. For example, in the high thresholds model, there were four transition probabilities that had higher power than the rest at *N* = 1,000. These four transition probabilities were all classes transitioning *into* class 1: class 1 into class 1, class 2 into class 1, class 3 into class 1, and class 4 into class 1. Because all other conditions are held constant (i.e., transition probabilities and class size), the measurement model itself must be attributing to the differences in power values for each transition probability. This result shows that there is higher power when transitioning into a class that is distinct from the rest. In this case, that distinct class is class 1.

For this poorly defined model with even transition probabilities and even class sizes, when thresholds were moderate, power stabilized near *N* = 5,000. None of the transition probabilities ever reached .80. The revised moderate thresholds model performed better, stabilizing after *N* = 1,500. Most of the transition probabilities met the .80 recommendation at *N*s ≥ 5,000. Interestingly, the four transition probabilities that did not meet the cutoff were all transitions *into* class 3, a class that is the least distinct from the rest. In other words, many of the item-response probabilities in class 3 look like the other classes. This result is similar to the one above, that sample size and transition probabilities are not the only influence on statistical power; the separation of the latent class influences it as well.

For models with high thresholds, power was stable for all *N* and *N*s ≥ 5,000 had adequate power for all transition probabilities. Transition probabilities for classes with more distinct thresholds were more likely to have higher power. Similar to the moderate and revised moderate thresholds model, there was greater power when transitioning into class 1 and lower power when transitioning into class 3.

In summary, this poorly defined model with even transition probabilities and even class sizes contributes the following findings in addition to what we have seen in results thus far:

- The poorly defined model has more difficulty stabilizing than the well-defined model.
- Measurement models with better class separation have higher power.
- Transitions into distinct classes have higher power than transitions into non-distinct classes. Non-distinct classes in models with larger thresholds had greater power than lower thresholds, as larger thresholds increase homogeneity.

4.5.2 Uneven class sizes. This model experienced instability and difficulty converging for the moderate thresholds and moderate revised thresholds models (see Table 22 and Figures 33–35). For moderate thresholds, power does not stabilize under the sample sizes included in this study. This model also had trouble converging, with only about a third of replications completing even when *N* = 10,000. Difficulties converging are exhibited in "spikes" in its power curve. With these thresholds, only two of the transition probabilities ever meet the .80 recommendation. As seen in the previous model, these higher power transition probabilities were going *into* class 1a latent class with logit thresholds more distinct from the other three classes. The next two highest power values were also for transition probabilities going *into* class 1.

The revised moderate thresholds model had less difficulty converging, with 99.5% of replications completing when N = 10,000. However, less than half of the replications converged when N < 1,500. This model eventually stabilizes after N = 6,000 and did not exhibit “spikes” in its power curve. Furthermore, not all transition probabilities met the .80 recommendation. Even when N = 10,000, three transition probabilities do not meet the .80 cutoff. These three transition probabilities are all from a very small class (class 4), which only has 5% of the overall sample at t = 1. Again, this shows that the effect of sparseness due to small class size. Recall that in the previous subsection, there was strong power for a transition probability of .01 going from class 4 to class 1 when class sizes were equal. Now, in this case with a very small class size, the power is much lower and is even one of the lowest power values in the entire model.

In the previous subsection, we saw that transition probabilities were lower when going *into* class 3 which was the least distinct of all the classes. Here, we can see how this finding fares by comparing class size and power for the revised moderate thresholds with even versus uneven class sizes (see Table 20 below). These models are identical in measurement with the same transition probabilities. The only difference is in class sizes at t = 1. For the transition probability from class 1 to class 3, power increased from .70 to .84 when class size doubled from 25% to 50%. For the transition probability from class 4 to class 3, power decreased from .69 to .15 when class size decreased from 25% to 5%. In summary, although the power for latent transition probabilities going into class 3 are generally lower than other classes, power values increase when there is less sparseness in that particular cell.

Table 20: *Comparison of Power for Even versus Uneven Class Size Model in Poorly Defined Model with Even Transition Probabilities and Revised Moderate Thresholds for N = 6,000*

	Even Class Sizes		Uneven Class Sizes	
	Class Size	Power	Class Size	Power
C1 to C3	25%	0.70	50%	0.84
C2 to C3	25%	0.71	30%	0.81
C3 to C3	25%	0.66	15%	0.63
C4 to C3	25%	0.69	5%	0.15

The high thresholds model had less difficulty stabilizing and converging. This model stabilized around $N = 250$. At sample size $N = 5{,}000$, there was adequate power for all transition probabilities. Again, we see the consistent finding that transition probabilities for larger classes have higher power. We also see that the combination of transitioning from a large class and transitioning into class 1 results in higher power.

In summary, this poorly defined model with even transition probabilities and uneven class sizes contributes the following findings in addition to what we have seen in results thus far:

- Although the poorly defined model has difficulty converging, poorly defined models with higher thresholds are more likely to converge and stabilize on lower values of *N*.

Table 21: *Power Values for Poorly Defined Model with Even Transition Probabilities and Even Class Sizes*

	N	1 1	2 1	3 1	4 1	1 2	2 2	3 2	4 2	1 3	2 3	3 3	4 3
Moderate Thresholds	100	0.40	0.40	0.40	0.43	0.39	0.44	0.40	0.46	0.41	0.41	0.47	0.46
	250	0.39	0.33	0.37	0.37	0.33	0.39	0.33	0.42	0.37	0.37	0.45	0.44
	500	0.33	0.28	0.32	0.37	0.23	0.38	0.28	0.37	0.29	0.32	0.40	0.38
	1000	0.41	0.43	0.34	0.39	0.41	0.44	0.42	0.46	0.35	0.28	0.39	0.41
	1250	0.42	0.36	0.38	0.38	0.34	0.40	0.35	0.43	0.34	0.36	0.42	0.39
	1500	0.46	0.37	0.40	0.37	0.39	0.44	0.39	0.41	0.38	0.39	0.42	0.45
	5000	0.51	0.45	0.44	0.52	0.43	0.42	0.39	0.43	0.41	0.39	0.45	0.47
	6000	0.55	0.49	0.52	0.56	0.47	0.44	0.43	0.47	0.51	0.47	0.51	0.53
	7000	0.60	0.51	0.53	0.62	0.43	0.44	0.40	0.45	0.50	0.46	0.48	0.51
	8000	0.63	0.55	0.54	0.62	0.46	0.43	0.40	0.47	0.47	0.43	0.46	0.51
	10000	0.66	0.57	0.58	0.68	0.46	0.44	0.42	0.48	0.51	0.44	0.46	0.51
Moderate (Revised) Thresholds	100	0.40	0.37	0.40	0.41	0.38	0.42	0.42	0.44	0.40	0.38	0.46	0.46
	250	0.42	0.34	0.36	0.39	0.33	0.39	0.39	0.41	0.36	0.36	0.45	0.42
	500	0.38	0.34	0.30	0.40	0.31	0.40	0.36	0.42	0.35	0.30	0.39	0.38
	1000	0.49	0.48	0.49	0.48	0.49	0.53	0.49	0.52	0.43	0.41	0.45	0.45
	1250	0.52	0.50	0.47	0.47	0.52	0.56	0.48	0.52	0.42	0.42	0.44	0.42
	1500	0.56	0.59	0.52	0.56	0.64	0.66	0.55	0.61	0.43	0.41	0.46	0.40
	5000	0.85	0.86	0.83	0.85	0.87	0.88	0.85	0.88	0.63	0.66	0.62	0.65
	6000	0.86	0.87	0.83	0.86	0.88	0.88	0.86	0.87	0.70	0.71	0.66	0.69
	7000	0.90	0.90	0.87	0.90	0.90	0.90	0.88	0.89	0.72	0.72	0.71	0.72
	8000	0.90	0.91	0.88	0.91	0.91	0.91	0.90	0.90	0.78	0.77	0.73	0.77
	10000	0.93	0.93	0.91	0.93	0.93	0.94	0.92	0.93	0.84	0.83	0.79	0.84
High Thresholds	100	0.60	0.49	0.43	0.59	0.46	0.50	0.42	0.51	0.39	0.37	0.41	0.46
	250	0.75	0.57	0.64	0.85	0.51	0.53	0.50	0.56	0.53	0.48	0.52	0.65
	500	0.84	0.73	0.82	0.91	0.59	0.57	0.58	0.64	0.64	0.55	0.60	0.71
	1000	0.94	0.89	0.93	0.97	0.72	0.66	0.69	0.77	0.73	0.69	0.69	0.73
	1250	0.97	0.95	0.96	0.99	0.79	0.72	0.77	0.84	0.74	0.71	0.71	0.77
	1500	0.98	0.97	0.98	0.99	0.86	0.78	0.83	0.88	0.79	0.77	0.77	0.79
	5000	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.97	0.98
	6000	0.98	0.99	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	7000	0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.98	0.98	0.98	0.99
	8000	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.97	0.98	0.98	0.99
	10000	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

Figure 30. *Poorly Defined Model/Even Transition Probabilities/Even Classes/Moderate Thresholds.*

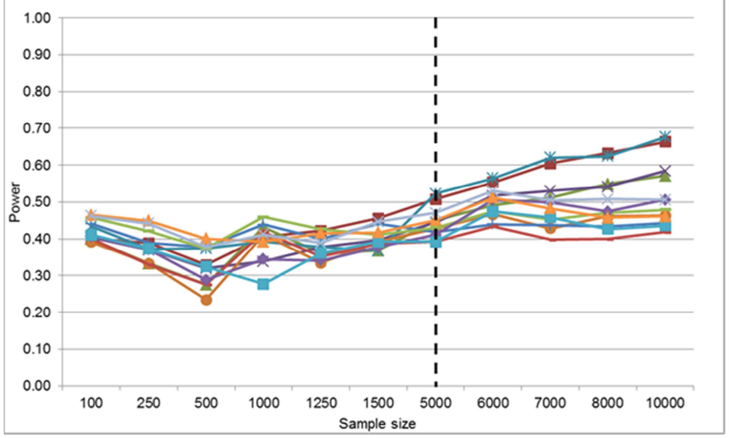


Figure 31. *Poorly Defined Model/Even Transition Probabilities/Even Classes/Moderate (Revised) Thresholds.*

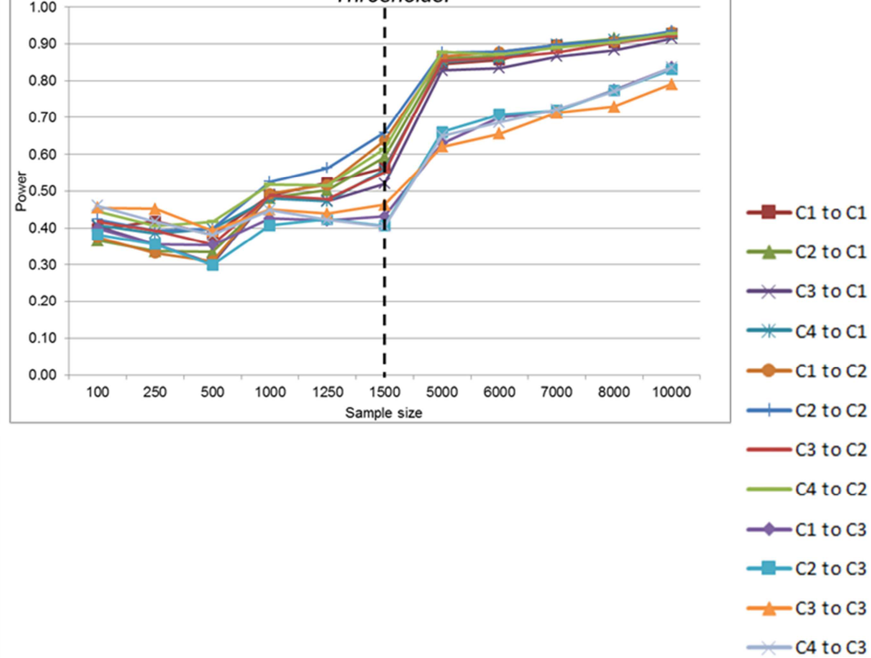


Figure 32. *Poorly Defined Model/Even Transition Probabilities/Even Classes/High Thresholds.*

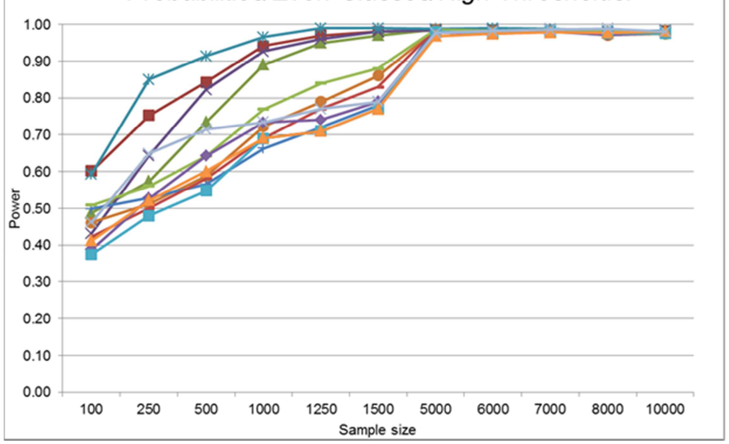


Table 22: *Power Values for Poorly Defined Model with Even Tprob and Uneven Class Sizes*

	N	1 1	2 1	3 1	4 1	1 2	2 2	3 2	4 2	1 3	2 3	3 3	4 3
Moderate Thresholds	100	0.39	0.36	0.44	0.48	0.39	0.44	0.46	0.52	0.39	0.42	0.51	0.51
	250	0.44	0.32	0.36	0.47	0.34	0.41	0.39	0.47	0.38	0.36	0.48	0.51
	500	0.40	0.28	0.37	0.41	0.34	0.34	0.35	0.36	0.29	0.34	0.41	0.40
	1000	0.50	0.31	0.22	0.15	0.28	0.23	0.20	0.23	0.23	0.26	0.32	0.20
	1250	0.51	0.33	0.18	0.16	0.18	0.19	0.24	0.19	0.28	0.26	0.31	0.25
	1500	0.62	0.42	0.33	0.09	0.35	0.36	0.29	0.22	0.43	0.33	0.41	0.25
	5000	0.90	0.70	0.46	0.26	0.38	0.36	0.28	0.22	0.35	0.28	0.32	0.22
	6000	0.93	0.78	0.57	0.38	0.40	0.39	0.28	0.25	0.41	0.36	0.38	0.20
	7000	0.94	0.78	0.54	0.37	0.37	0.35	0.25	0.18	0.37	0.30	0.32	0.16
	8000	0.95	0.82	0.59	0.41	0.36	0.35	0.21	0.20	0.39	0.31	0.35	0.19
	10000	0.95	0.86	0.58	0.51	0.38	0.33	0.23	0.19	0.37	0.31	0.28	0.19
Moderate (Revised) Thresholds	100	0.42	0.35	0.39	0.49	0.38	0.42	0.42	0.50	0.43	0.40	0.49	0.50
	250	0.44	0.33	0.36	0.42	0.38	0.42	0.35	0.41	0.39	0.33	0.41	0.44
	500	0.57	0.45	0.35	0.39	0.54	0.46	0.31	0.40	0.43	0.35	0.39	0.42
	1000	0.82	0.72	0.34	0.36	0.77	0.66	0.30	0.33	0.50	0.39	0.36	0.34
	1250	0.87	0.81	0.42	0.33	0.82	0.72	0.37	0.28	0.55	0.43	0.32	0.28
	1500	0.87	0.84	0.49	0.25	0.82	0.75	0.37	0.25	0.55	0.48	0.34	0.22
	5000	0.98	0.98	0.90	0.32	0.96	0.95	0.87	0.31	0.76	0.76	0.56	0.16
	6000	0.99	0.99	0.96	0.41	0.98	0.97	0.94	0.36	0.84	0.81	0.63	0.15
	7000	0.99	1.00	0.98	0.44	0.99	0.99	0.98	0.38	0.89	0.89	0.72	0.19
	8000	1.00	1.00	0.99	0.46	0.99	1.00	0.99	0.43	0.93	0.91	0.76	0.21
	10000	1.00	1.00	0.99	0.52	1.00	1.00	1.00	0.47	0.96	0.96	0.85	0.28
High Thresholds	100	0.72	0.50	0.39	0.33	0.55	0.52	0.39	0.35	0.59	0.38	0.41	0.39
	250	0.89	0.70	0.50	0.35	0.61	0.54	0.48	0.42	0.72	0.53	0.39	0.29
	500	0.98	0.89	0.75	0.64	0.71	0.60	0.50	0.48	0.77	0.67	0.47	0.35
	1000	1.00	0.99	0.96	0.93	0.86	0.74	0.62	0.55	0.81	0.76	0.61	0.47
	1250	1.00	1.00	0.99	0.97	0.92	0.83	0.73	0.63	0.88	0.84	0.72	0.57
	1500	1.00	1.00	0.99	0.99	0.95	0.89	0.79	0.70	0.90	0.88	0.76	0.64
	5000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	6000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	7000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	8000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Figure 33. *Poorly Defined Model/Even Transition Probabilities/Uneven Class Sizes/Moderate Thresholds.*

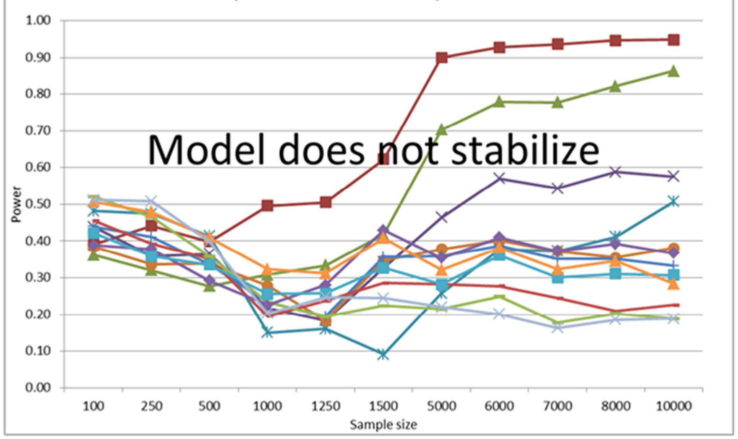
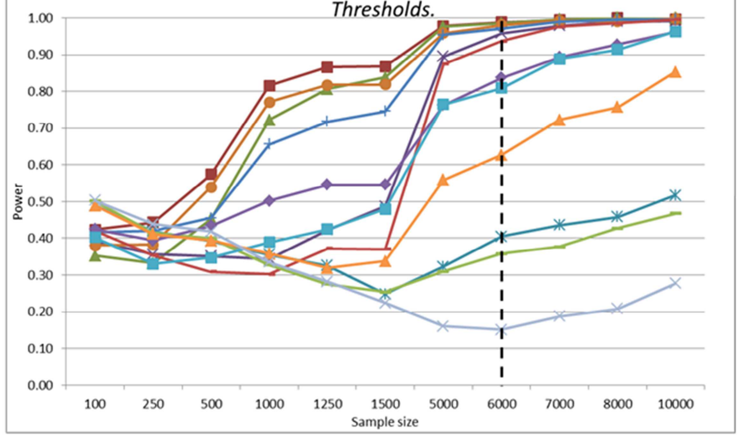


Figure 34. *Poorly Defined Model/Even Transition Probabilities/Uneven Class Sizes/Moderate (Revised) Thresholds.*



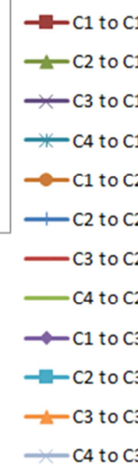
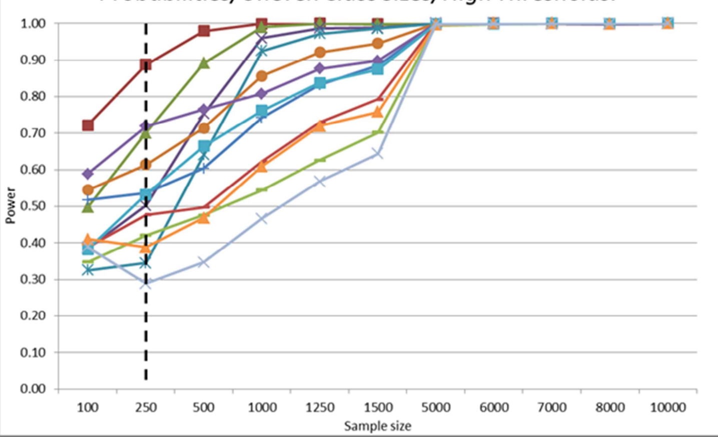


Figure 35. *Poorly Defined Model/Even Transition Probabilities/Uneven Class Sizes/High Thresholds.*



4.6 Poorly Defined Model with Uneven Transition Probabilities

4.6.1 Even class sizes. Similar to other poorly defined models, this model experienced difficulty converging, as seen in the "spikes" in the moderate thresholds model (see Table 23 and Figures 36–38). This poorly defined model also exhibited instability at all threshold levels. In fact, the moderate thresholds model never stabilized. Only 7 out of 1,000 replications completed when N = 10,000. Because so few replications converged, this model is not interpretable. It can be concluded that Mplus would never arrive at this sort of solution with such a poor measurement model and these thresholds.

The revised moderate thresholds model performed better, yet still with only 592 out of 1000 replications completed at N = 10,000. Because this model had difficulty converging and did not stabilize, results should be interpreted with great caution. For the same revised moderate thresholds model with even transition probabilities, 808 replications completed (rather than the 592 we saw before). Thus, there was increased difficulty for the model to converge on a plausible solution when transition probabilities were small. Similar to previous findings, some transition probabilities performed better than others. In this model, larger transition probabilities had the greatest power. The next best power values were for transition probabilities equal to .15 (class 1 to class 2), .12 (class 2 to class 3) and .07 (class 3 to class 2). The lowest power values were attributed to very small transition probabilities (i.e., .01 and .02). Again, this result shows the effect that the value of a transition probability has on its statistical power.

With high thresholds, power stabilizes at $N \geq 1{,}250$. When N = 10,000, power values for all transition probabilities are near or exceed .80. We see the same patterns seen earlier in similar models. For instance, some transition probabilities are high for all N while others are much lower. The high power transition probabilities are for those going into class 1, which was deemed distinct from the other classes. The low power transition probabilities are for those going into class 3, which was deemed indistinct from the other classes.

In summary, this poorly defined model with uneven transition probabilities and even class sizes contributes the following findings in addition to what we have seen in results thus far:

- A combination of "poor" conditions makes it difficult for statistical programs to converge on a solution. It can be concluded that in a non-simulated study, Mplus would not reach a solution on this model if there were moderate thresholds.

4.6.2 Uneven class sizes. From the findings thus far, one would expect this model to have the poorest power of the ones considered. Up to this point, we have seen that small transition probabilities, small class sizes and poorly defined models have lower power. Prior to looking at results, one would expect:

- the moderate thresholds model to have extreme difficulty converging,
- the revised moderate model to have improved yet still some difficulty converging,
- the high thresholds model to have little to no difficulty converging,
- small class sizes to have lower power,
- large class sizes to have higher power,
- small transition probabilities to have lower power,
- diagonal transition probabilities to have higher power,

- transitioning into a distinct class (i.e., class 1) will have higher power than transitioning into an indinstict class (i.e., class 3), and
- larger transition probabilities and higher thresholds can help recover loss in power.

These hypotheses are consistent with findings related to this poorly defined model with uneven transition probabilities and uneven class sizes (see Table 24 and Figures 39–41). The moderate thresholds model never stabilized. Only 11 out of 1,000 replications converged when *N* = 10,000. Again, this moderate thresholds model is deemed uninterpretable. The revised moderate thresholds model also had difficulty converging and never reached stability, although 424 out of 1000 replications completed at *N* = 10,000. Lastly, the high thresholds model had little difficulty converging, with 999 completed replications at *N* = 10,000. Additionally, this model's power curve reached stability at *N* = 5,000.

Findings regarding the interplay of class sizes and transition probability values also returned true with this model. For the high thresholds model, 6 out of 12 transition probabilities reached adequate power at *N* = 5,000. As expected, these higher power transition probabilities had one or more of the following attributes: larger transition probability, larger class size, and/or transitioning into a distinct class (i.e., class 1). For *N* = 10,000, the transition probabilities that met the .80 recommendation had a large class size (class 1 or class 2 at *t* = 1) and/or large transition probability (diagonal value or transition probability greater than or equal to .12).

In summary, this model corroborates all other findings and expectations. Trends regarding stability, convergence, sample size, transition probabilities, class sizes, homogeneity and class separation persist throughout all models. The following chapter provides a discussion of all results, practical recommendations, limitations, and future directions.

Table 23: *Power Values for Poorly Defined Model with Uneven Transition Probabilities and Even Classes*

	N	1 1	2 1	3 1	4 1	1 2	2 2	3 2	4 2	1 3	2 3	3 3	4 3
Moderate Thresholds	100	0.45	0.45	0.47	0.47	0.32	0.48	0.43	0.50	0.38	0.39	0.50	0.51
	250	0.65	0.33	0.40	0.40	0.24	0.54	0.33	0.40	0.31	0.31	0.50	0.43
	500	0.80	0.31	0.36	0.34	0.18	0.60	0.32	0.33	0.20	0.27	0.57	0.31
	1000	0.91	0.15	0.20	0.18	0.15	0.68	0.20	0.19	0.13	0.14	0.60	0.19
	1250	0.96	0.10	0.17	0.15	0.14	0.69	0.13	0.14	0.03	0.09	0.61	0.13
	1500	0.97	0.06	0.08	0.10	0.15	0.81	0.09	0.11	0.09	0.10	0.73	0.08
	5000	1.00	0.06	0.06	0.00	0.31	0.94	0.06	0.00	0.06	0.13	0.94	0.06
	6000	1.00	0.00	0.00	0.06	0.19	0.69	0.00	0.06	0.00	0.06	0.88	0.00
	7000	1.00	0.08	0.08	0.08	0.31	0.92	0.00	0.00	0.00	0.00	0.69	0.00
	8000	1.00	0.00	0.00	0.07	0.57	0.86	0.14	0.00	0.00	0.00	0.64	0.07
	10000	1.00	0.14	0.14	0.00	0.29	0.43	0.14	0.00	0.00	0.14	1.00	0.00
Moderate (Revised) Thresholds	100	0.56	0.45	0.49	0.45	0.29	0.54	0.41	0.43	0.39	0.33	0.48	0.44
	250	0.83	0.40	0.44	0.41	0.20	0.80	0.35	0.37	0.31	0.29	0.64	0.40
	500	0.95	0.33	0.40	0.33	0.28	0.93	0.26	0.31	0.26	0.17	0.84	0.32
	1000	0.97	0.26	0.26	0.26	0.59	0.97	0.18	0.25	0.16	0.21	0.93	0.24
	1250	0.99	0.23	0.23	0.22	0.74	0.99	0.18	0.23	0.14	0.26	0.97	0.25
	1500	0.98	0.17	0.19	0.17	0.81	0.99	0.15	0.14	0.13	0.29	0.97	0.17
	5000	1.00	0.06	0.04	0.05	0.99	1.00	0.39	0.10	0.17	0.83	0.99	0.07
	6000	1.00	0.06	0.03	0.04	0.99	1.00	0.48	0.10	0.18	0.89	0.99	0.07
	7000	1.00	0.06	0.04	0.03	0.98	1.00	0.51	0.12	0.22	0.92	0.99	0.05
	8000	1.00	0.06	0.03	0.02	0.99	1.00	0.57	0.12	0.25	0.96	1.00	0.06
	10000	1.00	0.06	0.04	0.05	1.00	1.00	0.70	0.13	0.31	0.98	1.00	0.07
High Thresholds	100	0.70	0.44	0.50	0.42	0.18	0.55	0.35	0.41	0.24	0.25	0.48	0.45
	250	0.95	0.31	0.44	0.33	0.55	0.85	0.25	0.27	0.19	0.21	0.78	0.32
	500	0.99	0.20	0.25	0.25	0.88	0.96	0.17	0.18	0.16	0.29	0.93	0.20
	1000	1.00	0.13	0.09	0.16	0.99	0.99	0.17	0.22	0.35	0.49	0.99	0.19
	1250	1.00	0.10	0.07	0.18	0.99	0.99	0.21	0.27	0.50	0.57	0.99	0.23
	1500	1.00	0.14	0.09	0.17	1.00	1.00	0.26	0.35	0.63	0.66	1.00	0.28
	5000	1.00	0.60	0.53	0.45	1.00	1.00	0.76	0.89	1.00	0.99	1.00	0.91
	6000	1.00	0.67	0.62	0.54	1.00	1.00	0.84	0.94	1.00	1.00	1.00	0.95
	7000	1.00	0.76	0.69	0.59	1.00	1.00	0.89	0.98	1.00	1.00	1.00	0.98
	8000	1.00	0.81	0.77	0.66	1.00	1.00	0.93	0.99	1.00	1.00	1.00	0.99
	10000	1.00	0.89	0.87	0.78	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00

Figure 36. *Poorly Defined Model/Uneven Transition Probabilities/Even Class Sizes/Moderate Thresholds.*

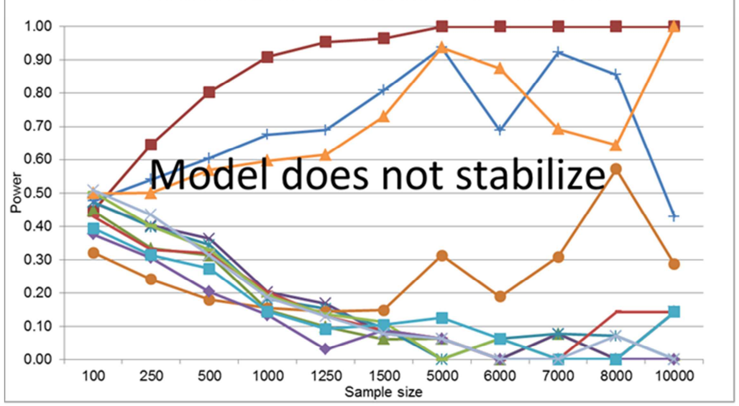
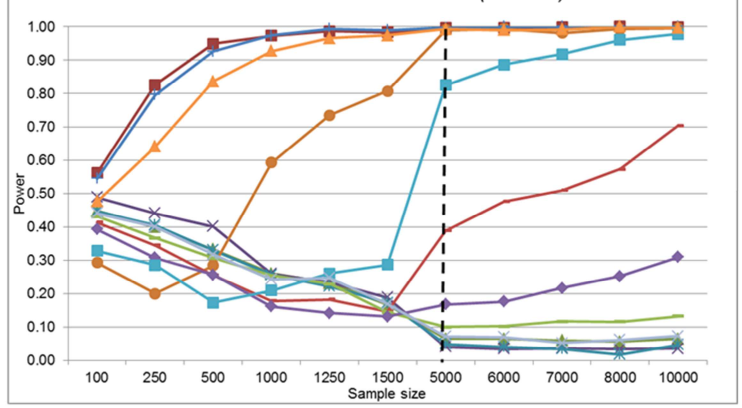


Figure 37. *Poorly Defined Model/Uneven Transition Probabilities/Even Class Sizes/Moderate (Revised) Thresholds.*



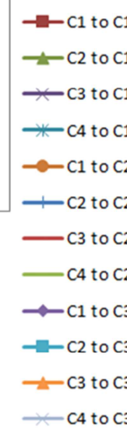


Figure 38. *Poorly Defined/Uneven Transition Probabilities/Even Class Sizes/High Thresholds.*

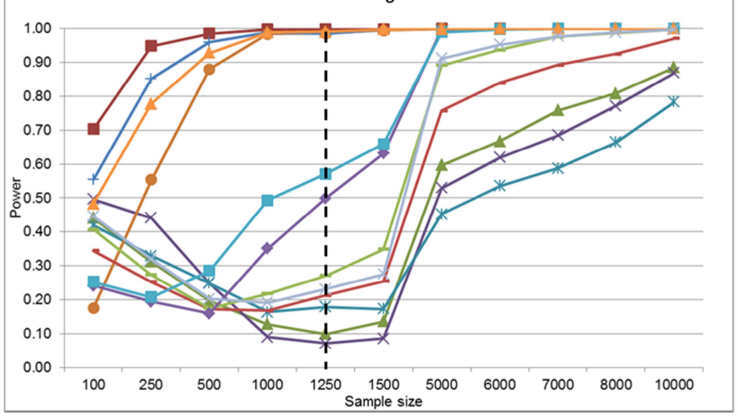


Table 24: *Power Values for Poorly Defined Model with Uneven Transition Probabilities and Uneven Classes*

		1 1	2 1	3 1	4 1	1 2	2 2	3 2	4 2	1 3	2 3	3 3	4 3
Moderate Thresholds	100	0.64	0.43	0.48	0.49	0.26	0.52	0.46	0.52	0.32	0.38	0.49	0.52
	250	0.74	0.40	0.46	0.49	0.23	0.61	0.43	0.49	0.25	0.31	0.47	0.51
	500	0.82	0.36	0.43	0.46	0.24	0.60	0.34	0.47	0.23	0.26	0.47	0.53
	1000	0.90	0.26	0.36	0.47	0.30	0.72	0.37	0.46	0.19	0.25	0.52	0.48
	1250	0.90	0.21	0.40	0.43	0.26	0.71	0.38	0.47	0.16	0.24	0.49	0.51
	1500	0.92	0.17	0.23	0.34	0.43	0.74	0.25	0.33	0.13	0.17	0.51	0.33
	5000	1.00	0.08	0.08	0.08	0.48	0.88	0.00	0.08	0.00	0.04	0.60	0.12
	6000	0.96	0.00	0.05	0.09	0.64	0.96	0.09	0.05	0.00	0.05	0.82	0.05
	7000	1.00	0.00	0.10	0.00	0.75	0.90	0.10	0.05	0.00	0.00	0.70	0.05
	8000	1.00	0.00	0.00	0.00	0.50	0.75	0.00	0.00	0.00	0.00	0.69	0.00
	10000	1.00	0.00	0.00	0.18	0.64	0.91	0.00	0.00	0.00	0.09	0.82	0.00
Moderate (Revised) Thresholds	100	0.78	0.45	0.51	0.52	0.27	0.62	0.44	0.51	0.37	0.35	0.43	0.51
	250	0.87	0.40	0.48	0.50	0.38	0.84	0.41	0.47	0.28	0.29	0.55	0.48
	500	0.92	0.39	0.46	0.42	0.58	0.91	0.33	0.46	0.23	0.25	0.66	0.46
	1000	0.96	0.36	0.42	0.41	0.85	0.95	0.33	0.44	0.22	0.21	0.76	0.43
	1250	0.97	0.32	0.37	0.37	0.89	0.96	0.27	0.39	0.22	0.31	0.83	0.41
	1500	0.98	0.30	0.40	0.42	0.95	0.98	0.26	0.39	0.22	0.36	0.88	0.38
	5000	1.00	0.21	0.25	0.32	1.00	1.00	0.23	0.26	0.49	0.88	0.99	0.27
	6000	0.99	0.19	0.25	0.29	0.99	1.00	0.25	0.24	0.55	0.92	0.98	0.26
	7000	1.00	0.21	0.20	0.27	0.99	1.00	0.28	0.22	0.65	0.94	0.99	0.24
	8000	1.00	0.16	0.20	0.24	0.99	1.00	0.31	0.21	0.68	0.97	0.99	0.22
	10000	1.00	0.14	0.16	0.22	0.99	1.00	0.24	0.16	0.79	0.98	0.99	0.18
High Thresholds	100	0.88	0.53	0.69	0.65	0.52	0.71	0.48	0.70	0.31	0.37	0.54	0.71
	250	0.97	0.33	0.53	0.46	0.81	0.90	0.36	0.50	0.16	0.26	0.64	0.54
	500	0.97	0.33	0.53	0.46	0.81	0.90	0.36	0.50	0.16	0.26	0.64	0.54
	1000	1.00	0.16	0.21	0.22	0.98	0.99	0.20	0.15	0.67	0.67	0.96	0.24
	1250	1.00	0.20	0.19	0.22	0.99	0.99	0.19	0.11	0.81	0.75	0.98	0.22
	1500	1.00	0.20	0.12	0.15	0.99	1.00	0.19	0.10	0.88	0.82	0.98	0.14
	5000	1.00	0.77	0.21	0.10	1.00	1.00	0.47	0.21	1.00	1.00	1.00	0.13
	6000	1.00	0.84	0.26	0.12	1.00	1.00	0.54	0.24	1.00	1.00	1.00	0.18
	7000	1.00	0.89	0.32	0.10	1.00	1.00	0.62	0.29	1.00	1.00	1.00	0.22
	8000	1.00	0.93	0.38	0.11	1.00	1.00	0.66	0.32	1.00	1.00	1.00	0.26
	10000	1.00	0.97	0.48	0.10	1.00	1.00	0.77	0.42	1.00	1.00	1.00	0.36

Figure 39. *Poorly Defined Model/Uneven Transition Probabilities/Uneven Class Sizes/Moderate Thresholds.*

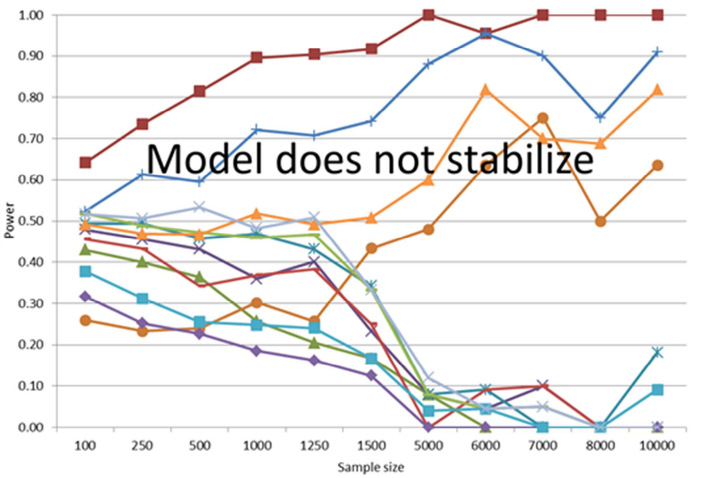
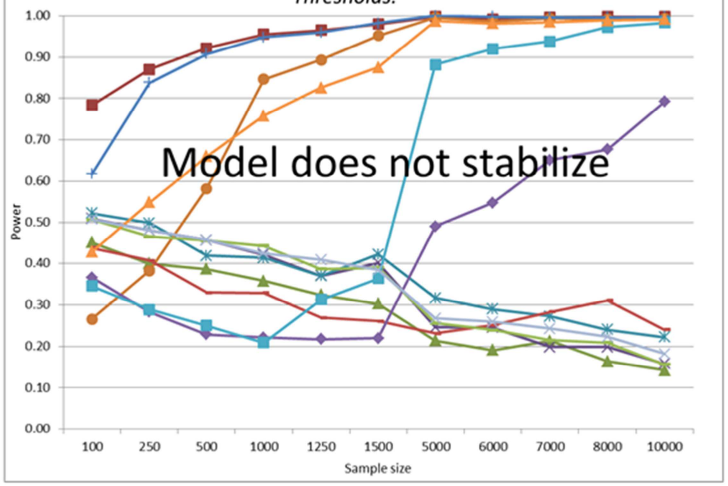


Figure 40. *Poorly Defined Model/Uneven Transition Probabilities/Uneven Class Sizes/Moderate (Revised) Thresholds.*



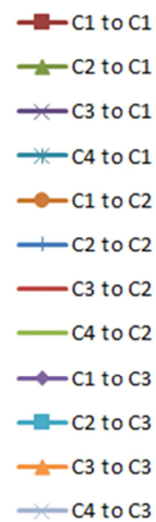
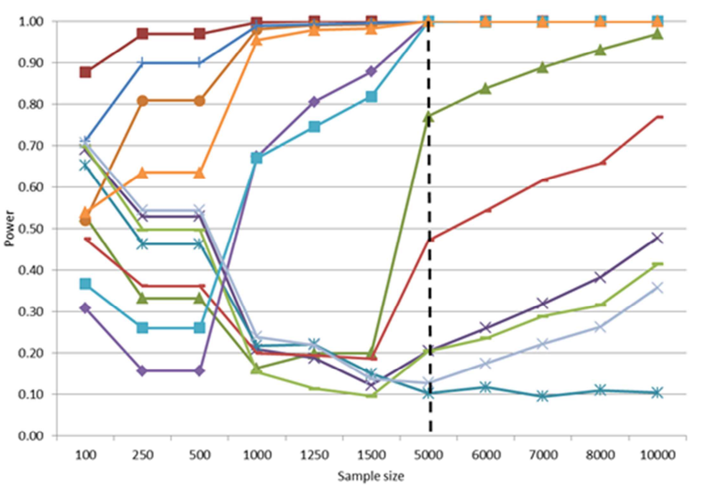


Figure 41. *Poorly Defined Model/Uneven Transition Probabilities/Uneven Class Sizes/High Thresholds.*



Chapter 5 Discussion

5.1 General Overview

LTA is a statistical model that can be used to study how individuals transition from or stay in latent classes over time. LTA is commonly used and gaining increasing popularity in many fields including educational, health, and behavioral research. Because only two simulation studies have been conducted to examine the how the model specification and sample size requirements for this model , the purpose of this study was to investigate the sample size needed to establish statistical power to detect latent transition probabilities under various model conditions. These conditions included sample size, well-defined versus poorly defined measurement models, equal versus unequal transition probability matrices, equal versus unequal class sizes, and variations of item-response logit thresholds which relate to the measurement quality of the latent class models. A meta-analysis was conducted to explore common characteristics of recently published LTA studies. Using these attributes, Monte Carlo simulations were conducted to help examine what level of power one can expect under those conditions. This discussion section will review the general trends found across results. This section will also provide general recommendations for applied researchers using LTA. Lastly, this section will discuss limitations to this study and future directions for future LTA power studies.

5.2 Summary of Dissertation

In total, 308 models were generated across the five conditions included in this Monte Carlo simulation study:

- Two time points
- 11 sample sizes

- Two measurement models: well-defined, poorly defined
- Two sets of transition probability matrices: equal, unequal
- Two sets of class sizes: equal, unequal
- Three to four variations in logit thresholds:
    - ±1, ±2, ±3, and ±5 for the well-defined model
    - moderate, moderate (revised), and high for the poorly defined model

Power values for 12 transition probabilities were provided and studied in Mplus output, summing to a total of 3,696 power values across all 308 models. Taken together, results from this study revealed the effect of each condition on the power to detect latent transition probabilities, as well as the impact of the combination of two or more conditions on power.

5.3 Key Trends and Findings

The following section will walk through key trends and findings across all models included in this dissertation. Multiple regressions were conducted using Stata 12.1 (Statacorp, 2011) to statistically examine the relationship between power and all of the model conditions. First, two measurement models were included in this dissertation to examine how power performs in a well-defined model in comparison to a poorly defined model. Results show that the measurement model is a key component on whether or not a transition probability will have adequate power. Collins and Lanza (2010) state that class separation and homogeneity are two desirable attributes when considering an LCA model. First, this study helped show how high class separation is related to higher power in the LTA model. When comparing well-defined and poorly defined models that were equivalent on all other characteristics, the well-defined model always had higher power than the poorly defined model. In the multiple regress analysis, holding all conditions constant, well-defined models had significantly higher power than poorly defined models (β = .35, p < .001).

We also saw the effect of class separation in the poorly defined models. Even when all other conditions were equal, the power for latent transition probabilities going into class 1 was greater than going into class 3. A closer look at the measurement model revealed that class 1 was the most distinct (i.e., had the best measurement qualities) from the other three classes, while the item-response probabilities for class 3 were similar to another class. Thus, this study helped show that high class separation in a well-defined model led to greater power than a poorly defined model and also showed how on a smaller scale, within a poorly defined model a class that is more separate than the others will likely have higher power.

By varying logit thresholds, we could also see how homogeneity impacts the power of a latent transition probability. Higher thresholds imply increased homogeneity, which is synonymous with item-response probabilities near 0 or 1, considered a good measurement quality since we know with certainty how individuals in a given class responded. When thresholds were increased, we saw two important results. First, power was always higher in models that had more homogeneous classes. Second, when thresholds were increased, unstable models were now more stable because Mplus was less likely to fix standard errors. Regression results support this finding. Thresholds were, in fact, significant predictors of power in both poorly defined (β = .46, p < .001) and well-defined (β = .30, p < .001) models.

Other model attributes led to greater power. As one might expect, larger sample sizes had better power and also more stability. In fact, sample size was the greatest predictor of both for both poorly defined models (β = .50, p < .001) and well-defined models (β = .95, p < .001). Additionally, larger transition probabilities from larger class sizes had higher power than small transition probabilities from a small class. For example, in a single model, a transition probability of .80 from a class size of 50% at $t = 1$ always had higher power than a transition probability of .15 from a class size of 5%. This result helped prove hypotheses that sparseness would affect statistical power. Larger class sizes have greater power than smaller class. In other words, cells with more individuals would have greater power. Small class sizes also had more difficulty stabilizing. Although regression results did not reveal significant interactions between transition probabilities and class size, there were significant three-way interaction effects between transition probabilities, class size, and thresholds. In other words, larger thresholds have a greater effect of transition probabilities and class size on the power to detect latent transition probabilities.

In summary, the following are the key findings of this study:

- The measurement model matters. Models with highly separated and homogeneous classes are likely to have higher power.
- Large sample sizes—larger than we usually see in applied work—is needed to establish power for all parameters of the model, especially small latent transition probabilities.
- Although the measurement and sample size are key predictors of power, applied researchers must consider all model conditions (e.g., transition probabilities, class size) when determining what sample size is necessary for adequate power.
- Non-convergence and fixed standard errors are indications of poor model measurement. This is also referred to as solution propriety (Wolf, Harrington, Clark, & Miller, 2013). In applied work, poor power and under-identification can

be indicated by a large number of errors and the need to re-specify models numerous times.

5.4 Practical Recommendations

Researchers often ask what sample size is necessary for their statistical model. Regarding LTA, this extensive Monte Carlo simulation study suggests the following response: it depends. The sample size necessary to attain adequate statistical power for latent transition probabilities depends on a number of characteristics including the measurement model, item-response probabilities, latent transition probabilities, class sizes and sparseness of cells. Looking at results from this study, it can be argued that the measurement model and sample size are the leading factors in whether a transition probability has adequate statistical power. However, even with a well-defined model, the sample size needed for adequate power depends on other characteristics.

In a simulation study, the researcher has the ability to control and manipulate the model and all conditions. In this study, five different conditions were varied and specified in Mplus. However, realistically and practically, most conditions are not controllable when working with real-life data. A researcher might have the ability to increase the sample size of a study. However, other characteristics such as thresholds, class separation, transition probabilities and class sizes emerge from parameter estimation. Although these conditions are not controllable, one can expect a latent transition probability to have better power if it has the following characteristics:

- Within a model with homogeneous classes
- Within a model with high class separation
- In a class that is distinct from other classes

- Larger transition probability value
- Larger class size at $t = 1$
- Larger N

Researchers can use Tables 16–19 and Tables 21–24 as a guide for what one might expect under those conditions. Table 13 shows how many latent transition probabilities met the .80 in each model, while Table 14 shows how many replications converged in each model. This study can help researchers understand how poor or strong model attributes impacts the statistical power to detect a latent transition probability. For instance, a researcher could say, “If I have a sample size of 500, my LCA model has homogeneous classes but poor class separation, the power to detect a very small latent transition probability from a small class will likely be low.”

Results from this dissertation indicate that we need sample sizes larger than we are used to seeing in applied social science studies. These larger sample sizes are needed to say, with confidence, that all parameters in the model have sufficient power. Additionally, it is important to have good measurement models. This is a difficult requirement to have a priori because it is unknown what classes will emerge due to the exploratory nature of LCA. Researchers can look at similar previously conducted LCA studies to speculate what the $t = 2$ classes may look like. Applied researchers should conduct simulation studies in Mplus using LCA results from previously conducted studies to speculate possible results and to ensure that their parameters will have sufficient power.

Applied researchers often run power simulations studies to justify their sample size and results. However, the “artificially high” patterns that emerged in this dissertation reveal some issues with this approach. For example, a researcher may run a single simulation and find that there is adequate power with a sample size of $N = 100$ when in fact this power value is artificially high due to the fixing of parameters that Mplus does. If this researcher is not aware that many standard errors are being fixed to zero by reading the error messages provided in the output, the researcher will have incorrect justification to support the small sample size. Applied researchers should carefully look into parameter estimates and standard errors to ensure that power values are correctly estimated and not artificially adequate.

Importantly, statistical power is the probability to detect an effect when there is in fact an effect. When power is low, there is greater chance for Type II error. These Monte Carlo simulations showed how varying one or more conditions could increase or decreases power. Inversely, these simulations showed how varying one or more conditions could decrease or increase the probability of making a Type II error. Researchers must consider the effects of model characteristics on the power to detect latent transition probabilities and the chance of committing a Type II error.

5.5 Limitations and Future Directions

Because only two other LTA power studies have been conducted to date, the scope of this dissertation was to examine how a set of conditions influence the statistical power of latent transition probabilities. The conditions included in the simulations were based on commonalities found in recently published LTA studies. This study did not exhaust all possible variations of an LTA model. First, only two time points were simulated with 4-class solutions at both time points. Additionally, measurement invariance was assumed, though it is not a necessary condition of the LTA model. Extensions of this study should include more time points, more or less classes, and even examine power when measurement invariance is not assumed. This study also only examined 5 categorical indicators. Additional categorical and the inclusion of continuous variables can add to the results found in this study. One of the two other LTA simulation studies to date (Collins & Wugalter, 1992) aimed to determine if adding additional indicators in a LTA model would provide better measurement or more sparse contingency tables. The study found that including more indicators improved standard errors even when the contingency tables were sparse. Adding more indicators to this study could help eliminate issues such as artificially high power and inability to converge as a result of sparseness and poor model measurement.

The variations of a latent transition probabilities matrix are seemingly endless. This study looked at two matrices to see how power fared among larger or smaller probabilities. Other transition probabilities such as .50 could be included in a future simulation study. Future studies can also look at the importance of power for small transition probabilities such as .01. Further research can help answer questions that were not covered in this dissertation, such as whether each transition probability needs adequate power or if adequate power for the majority of transition probabilities would suffice. Additionally, different combinations of class size can be examined. This study included 11 sample sizes, though further simulations should include sample sizes between N = 1,500 and N = 5,000. This could reveal earlier instances of stability in the power curves. Lastly, extensions of this study can include latent transition analysis models that include covariates and distal outcomes and examine the power to detect latent transition probabilities, given these explanatory variables.

Future extensions of this study could examine the 3-step approach in mixture modeling (Asparouhov & Muthén, 2014). This approach is gaining popularity for its advantage in correcting for classification error. A future simulation study could incorporate the 3-step method to examine power in LTA. For example, in a measurement model with high entropy yet a small sample size, the researcher can fix individuals to classes and conduct a cross-tabulation of class proportions at $t = 1$ and $t = 2$. This study can help reveal the intersection between entropy, the measurement model, and sample size on power.

LTA is a valuable method that has been used to conduct research a number of different fields. This dissertation helped answer questions about how various model conditions can influence the statistical power to detect an effect in latent transition probabilities. It also has the potential to help researchers understand level of power they can expect under certain circumstances. This simulation study is the very beginning of a body of work that has yet to be conducted on LTA methodology. Future studies can help uncover other mysteries that still remain regarding LTA.

References

Abar, C. C. (2012). Examining the relationship between parenting types and patterns of student alcohol-related behavior during the transition to college. *Psychology of Addictive Behaviors*, *26*(1), 20-29.

Asparouhov, T. & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 329-341.

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*(1), 41-75.

Cain, A. S., Epler, A. J., Steinley, D., & Sher, K. J. (2012). Concerns related to eating, weight, and shape: Typologies and transitions in men during the college years. *International Journal of Eating Disorders, 45*(6), 768-775.

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*(1), 166-181.

Castellini, G., Fioravanti, G., Lo Sauro, C., Rotella, F., Lelli, L., Ventura, L., ... & Ricca, V. (2013). Latent profile and latent transition analyses of eating disorder phenotypes in a clinical sample: A 6-year follow-up study. *Psychiatry Research*, *207*(1), 92-99.

Cochran, W. G. (1954). Some methods for strengthening the common X2 tests. *Biometrics, 10*.

Collins, L. M., & Flaherty, B. P. (2002). Latent class models for longitudinal data. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis*, 287-303. New York: Cambridge University Press.

Collins, L., & Lanza, S. (2010). *Latent Class and Latent Transition Analysis with Applications in the Social, Behavioral and Health Science*. Hoboken, NJ: John Wiley & Sons, Inc.

Collins, L. M., & Tracy, A. J. (1997). Estimation in complex latent transition models with extreme data sparseness. *Kwantitatieve Methoden*, *18*, 57-71.

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, *27*(1).

Ding, C., Richardson, L., & Schnell, T. (2013). A developmental perspective on word literacy from kindergarten through the second grade. *The Journal of Educational Research*, *106*(2), 132-145.

Fan, X. (2003). Power of latent growth modeling for detecting group differences in linear growth trajectory parameters. *Structural Equation Modeling*, *10*(3), 380-400.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*(2), 215-231.

Haberman, S.J. (1979). *Analysis of Qualitative Data*, Vol 2, New Developments. New York: Academic Press.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424-453.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1).

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling*, *12*(3), 368-390.

Langeheine, R., & van de Pol, F. (2002). Latent Markov chains. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied Latent Class Analysis* (304-341). New York: Cambridge Universty Press.

Lanza, S. T., & Collins, L. M. (2006). A mixture model of discontinuous development in heavy drinking from ages 18 to 30: The role of college enrollment. *Journal of Studies on Alcohol, 67*(4).

Lanza, S. T., & Collins, L. M. (2008). A new SAS procedure for latent transition analysis: transitions in dating and sexual risk behavior. *Developmental Psychology*, *44*(2), 446-456.

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

Lee, M. R., Chassin, L., & Villalta, I. K. (2013). Maturing out of alcohol involvement: Transitions in latent drinking statuses from late adolescence to adulthood. *Development and Psychopathology*, *25*(4.1), 1137-1153.

Marcoulides, G. A., Gottfried, A. E., Gottfried, A. W., & Oliver, P. H. (2008). A latent transition analysis of academic intrinsic motivation from childhood through adolescence. *Educational Research and Evaluation: An International Journal on Theory and Practice, 14*(5), 411-427.

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology* (Vol. 2, 375-393). New York: Oxford University Press.

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods*, *3*(1), 45-48.

Morovati, D. (2014). *The intersection of sample size, number of indicators, and class enumeration in LCA: A Monte Carlo study* (Order No. 3637458). Available from ProQuest Dissertations & Theses A&I. (1617402700). Retrieved from http://search.proquest.com/docview/1617402700?accountid=14522

Muthén, B. (2008). *Multilevel growth models* [PDF document]. Retrieved from UCLA website http://pages.gseis.ucla.edu/faculty/muthen/Ed255C/Lecture/Lecture%209/Multilevel%20Growth%20Models.pdf

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599-620.

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Myers, N. D., Ahn, S., & Jin, Y. (2011). Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: a Monte Carlo approach. *Research Quarterly for Exercise and Sport*, *82*(3), 412-423.

Palen, L. A., Smith, E. A., Caldwell, L. L., Mathews, C., & Vergnani, T. (2009). Transitions to substance use and sexual intercourse among South African high school students. *Substance Use & Misuse*, *44*(13), 1872-1887.

Peterson, C. B., Crow, S. J., Swanson, S. A., Crosby, R. D., Wonderlich, S. A., Mitchell, J. E., ... & Halmi, K. A. (2011). Examining the stability of DSM–IV and empirically derived eating disorder classification: Implications for DSM–5. *Journal of Consulting and Clinical Psychology*, *79*(6), 777-783.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535-569.

Quaiser-Pohl, C., Rohe, A. M., & Amberger, T. (2010). The solution strategy as an indicator of the developmental stage of preschool children's mental-rotation ability. *Journal of Individual Differences, 31*(2), 95-100.

Rindskopf, R. & Rindskopf, W. (1986). The value of latent class analysis in medical diagnoses. *Statistics in Medicine*, *5*, 21-27.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, *58*(7), 935-943.

Spinhoven, P., de Rooij, M., Heiser, W., Smit, J. H., & Penninx, B. W. (2012). Personality and changes in comorbidity patterns among anxiety and depressive disorders. *Journal of Abnormal Psychology, 121*(4), 874-884.

StataCorp. (2011). *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.

Tueller, S. J., Drotar, S., & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling*, *18*(1), 110-131.

Velicer, W. F., Martin, R. A., & Collins, L. M. (1996). Latent transition analysis for longitudinal data. *Addiction*, *91*(12), 197-210.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913-934.

Appendix Sample Mplus Output with Annotated Comments

Well-Defined Model with Even Transition Probabilities, Even Class Sizes and Thresholds ±1 with N = 250

```
montecarlo:
!names of indicator variables
      names are u11-u15 u21-u25;

!the (1) indicates binary variables
      generate = u11-u15 u21-u25 (1);

!indicate that these are categorical and not continuous variables
      categorical = u11-u15 u21-u25;

!create 4 classes at each time point
      genclasses = c1(4) c2(4);
      classes = c1(4) c2(4);

!sample size
      nobservation = 250;

!number of replications
      nreps = 1000;

!saves parameter estimates for each replication
     results = 4c_even_well_even_1_250.csv;

!indicate this is a mixture model
  analysis:
      type=mixture;
      parameterization=probability;

!parameter values for overall population
  model population:

!class sizes
      %overall%
      [c1#1*.25];
      [c1#2*.25];
      [c1#3*.25];
      [c2#1*.25];
      [c2#2*.25];
      [c2#3*.25];

!latent transition probabilities
        c2#1 on c1#1*.25;
        c2#1 on c1#2*.25;
        c2#1 on c1#3*.25;
        c2#1 on c1#4*.25;

        c2#2 on c1#1*.25;
        c2#2 on c1#2*.25;
        c2#2 on c1#3*.25;
        c2#2 on c1#4*.25;
```

```
        c2#3 on c1#1*.25;
        c2#3 on c1#2*.25;
        c2#3 on c1#3*.25;
        c2#3 on c1#4*.25;


!item-response logit thresholds for time 1
  model population-c1:

!item-response logit thresholds for time 1 class 1
      %c1#1%
      [u11$1*1] (1);
      [u12$1*1] (2);
      [u13$1*1] (3);
      [u14$1*1] (4);
      [u15$1*1] (5);

!item-response logit thresholds for time 1 class 4

      %c1#2%
      [u11$1*1] (6);
      [u12$1*1] (7);
      [u13$1*-1] (8);
      [u14$1*-1] (9);
      [u15$1*-1] (10);

!item-response logit thresholds for time 1 class 3
      %c1#3%
      [u11$1*-1] (11);
      [u12$1*-1] (12);
      [u13$1*-1] (13);
      [u14$1*-1] (14);
      [u15$1*-1] (15);

!item-response logit thresholds for time 1 class 4
      %c1#4%
      [u11$1*-1] (16);
      [u12$1*-1] (17);
      [u13$1*1] (18);
      [u14$1*1] (19);
      [u15$1*1] (20);


!item-response logit thresholds for time 2 these should be identical to
values above because we are assuming measurement invariance

  model population-c2:
      %c2#1%
      [u21$1*1] (1);
      [u22$1*1] (2);
      [u23$1*1] (3);
      [u24$1*1] (4);
      [u25$1*1] (5);

      %c2#2%
      [u21$1*1] (6);
      [u22$1*1] (7);
```

```
      [u23$1*-1] (8);
      [u24$1*-1] (9);
      [u25$1*-1] (10);

      %c2#3%
      [u21$1*-1] (11);
      [u22$1*-1] (12);
      [u23$1*-1] (13);
      [u24$1*-1] (14);
      [u25$1*-1] (15);

      %c2#4%
      [u21$1*-1] (16);
      [u22$1*-1] (17);
      [u23$1*1] (18);
      [u24$1*1] (19);
      [u25$1*1] (20);

!parameter values for overall model
  Model:
        %overall%

        c2#1 on c1#1*.25;
        c2#1 on c1#2*.25;
        c2#1 on c1#3*.25;
        c2#1 on c1#4*.25;

        c2#2 on c1#1*.25;
        c2#2 on c1#2*.25;
        c2#2 on c1#3*.25;
        c2#2 on c1#4*.25;

        c2#3 on c1#1*.25;
        c2#3 on c1#2*.25;
        c2#3 on c1#3*.25;
        c2#3 on c1#4*.25;

  model c1:

      %c1#1%
      [u11$1*1] (1);
      [u12$1*1] (2);
      [u13$1*1] (3);
      [u14$1*1] (4);
      [u15$1*1] (5);

      %c1#2%
      [u11$1*1] (6);
      [u12$1*1] (7);
      [u13$1*-1] (8);
      [u14$1*-1] (9);
      [u15$1*-1] (10);

      %c1#3%
      [u11$1*-1] (11);
      [u12$1*-1] (12);
      [u13$1*-1] (13);
```

```
      [u14$1*-1] (14);
      [u15$1*-1] (15);

      %c1#4%
      [u11$1*-1] (16);
      [u12$1*-1] (17);
      [u13$1*1] (18);
      [u14$1*1] (19);
      [u15$1*1] (20);

  model c2:
      %c2#1%
      [u21$1*1] (1);
      [u22$1*1] (2);
      [u23$1*1] (3);
      [u24$1*1] (4);
      [u25$1*1] (5);

      %c2#2%
      [u21$1*1] (6);
      [u22$1*1] (7);
      [u23$1*-1] (8);
      [u24$1*-1] (9);
      [u25$1*-1] (10);

      %c2#3%
      [u21$1*-1] (11);
      [u22$1*-1] (12);
      [u23$1*-1] (13);
      [u24$1*-1] (14);
      [u25$1*-1] (15);

      %c2#4%
      [u21$1*-1] (16);
      [u22$1*-1] (17);
      [u23$1*1] (18);
      [u24$1*1] (19);
      [u25$1*1] (20);

!tech 1 provides parameter values
!tech 9 provides information on each replication such as errors
  Output: tech1 tech9;
```