UNIVERSITY OF CALIFORNIA

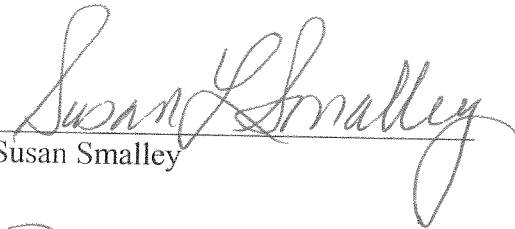Los Angeles

Mixture Modeling with Behavioral Data

A dissertation submitted in partial satisfaction of the

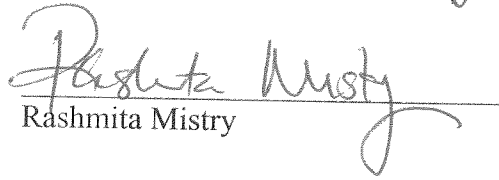requirements for the degree Doctor of Philosophy

in Education

by

Shaunna Lynn Clark
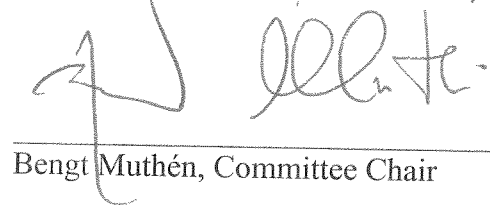
2010

The dissertation of Shaunna Lynn Clark is approved.

_____
Susan Smalley

_____
Rashmita Mistry

_____
Mike Seltzer

_____
Bengt Muthén, Committee Chair

University of California, Los Angeles

2010

ii

To my mother,

who always picked me up, dusted me off, and sent me on my way when I fell.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

My first and primary acknowledgement must go to my dissertation committee, especially my advisor and mentor, Bengt Muthén. It has been an honor and a privilege to learn from and work with him. Mike Seltzer has been a valuable committee member, always providing support when I asked for it. Rashmita Mistry and her social inequalities research group gave me an outlet for developing non-technical explanations of my work. Susan Smalley reminded me that I know more than I think I do and that I'm the "expert".

I would also like to thank two groups for the use of their datasets without out which this dissertation would not have been possible. For the use of the FinnTwin12, I would like to acknowledge Brian d'Onofrio, Dick Rose, and Rick Viken at the Univeristy of Indiana, Bloomington, and Jaakko Kaprio at the University of Helsinki, Finland. I would also like to thank Susan Smalley for the use of the LA ADHD dataset.

There were others, outside those most closely and academically connected to my dissertation, that also provided important support during this process. A big thank you goes out to Shonali Choudhury for her editing help, cupcake runs, and providing a shoulder to cry on. Thanks to Ethan Creech for copy editing and inserting snarky comments in my work to help keep me motivated to edit. Thank you to John French for always giving me emotional support and shoulder to lean on when things got tough. To my TSP family, especially Brisby, I want to thank you for giving me an outlet to vent my frustrations. Thank you to my SRM classmates and colleagues, especially Karen Nylund-Gibson, with whom I have shared my graduate experience. Thank you to the members of

VITA

July 19, 1981	Born, Westlake, California
2003	B.A., Statistics University of California, Berkeley
2003-2005	Research Assistant Center for Research on Evaluation, Standards, & Student Testing (CRESST) University of California, Los Angeles
2004	M.A., Education University of California, Los Angeles
2004-2005	Research Assistant Center for the Assessment and Evaluation of Student Learning (CAESL), UCLA
2005	Teaching Assistant/Special Reader School of Education and Information Studies, UCLA
2005-2009	Graduate Student Researcher School of Education and Information Studies, UCLA
2006	Teaching Assistant/Special Reader School of Education and Information Studies, UCLA
2007	Teaching Assistant/Special Reader School of Education and Information Studies, UCLA
2008	Applied Psychological Measurement Graduate Student Grant School of Education and Information Studies, UCLA

PUBLICATIONS AND PRESENTATIONS

Gearhart, M. & Clark, S. (2006). Developing expertise with classrooms assessment in K-12 science. Paper presented at the annual meeting for the Center for Research on Evaluation, Standards, & Student Testing (CRESST).

Gearhart, M., Nagashima, S., Pfotenhauer, J., Clark, S. Schwab, C., Vendlinski, T., Osmundson, E., Herman, J., & Bernbaum, D. (2006). Developing expertise with

classroom assessment in K-12 science: learning to interpret student work. Educational Assessment, 11(3&4), 237-263.

Clark, S. & Muthén, B. (2007). Latent class analysis of non-independent samples. Paper presented at the annual meeting of the American Educational Research Association.

Clark, S. & Muthén, B. (2007). The structure of ADHD: An application of growth mixture modeling. Paper presented at the annual meeting of the Society for Prevention Research.

Clark, S. & Muthén, B. (2007). How to handle clustered data when deciding on the number of classes in a latent class analysis: A Monte Carlo simulation study. Paper presented at the annual meeting of the Society for Prevention Research.

Clark, S. (July 2007). An Introduction to mixture modeling: Applications in psychology. 2-day short course presented at the Department of Psychology, University of Indiana, Bloomington.

Boomsma, D., Cacioppo, J., Muthén, B., Asparouhov, T. & Clark, S. (2007). Longitudinal genetic analysis for loneliness in Dutch twins. Twin Research and Human Genetics, 10, 267-273.

Muthén, B., Asparouhov, T. & Clark, S. (2008). Multilevel mixture modeling applications. Paper presented at the annual Joint Statistical Meeting.

Muthén, B., Asparouhov, T. & Clark, S. (2008). Multilevel growth mixture modeling: examining trajectories of student aggression. Paper presented at the annual meeting of the Society for Prevention Research.

McGough, J.J., Loo, S.K., McCracken, J.T., Dang, J., Clark, S., Nelson, S.F., & Smalley, S.L. (2008). The CBCL Pediatric Bipolar Disorder Profile and ADHD: A comorbidity study and quantitative trait loci analysis. Journal of the American Academy of Child and Adolescent Psychiatry, 47(10), 1151-7.

Clark, S. & Muthén, B. (2008) When good latent class analyses go bad: Treating latent class membership as an observed variable. Paper presented at the annual meeting of the Society for Prevention Research.

ABSTRACT OF THE DISSERTATION

Mixture Modeling with Behavioral Data

by

Shaunna Lynn Clark

Doctor of Philosophy in Education

University of California, Los Angeles, 2010

Professor Bengt Muthén, Chair

United States schools and students suffer from problems associated with student behavioral disorders. There is a need for innovate statistical methods to analyze data to which will help inform the development of new strategies to deal with the issues associated with behavioral problems. The three papers in this dissertation focus on explicating certain mixture models which have shown promise in analyzing behavioral data. An important interest in mixture modeling is the investigation of what types of individuals belong to each latent class by relating classes to auxiliary variables. The first paper presents results from real data examples and simulations to show how various factors, such as sample size, can impact the estimates and standard errors of auxiliary variable effects and testing mean equality across classes. Based on the results of the examples and simulations, suggestions are made about how to select auxiliary variables for a latent class analysis (LCA). The second paper discusses the factor mixture model (FMM) which uses a hybrid of both categorical and continuous latent variables.

The FMM is a good model for the underlying structure of behavioral disorders because the use of both categorical and continuous latent variables allows the structure to be simultaneously categorical and dimensional. The use of the FMM in behavioral research is not prevalent because there is little research about how the FMM should be applied in practice. This paper explores the FMM by studying two real data examples: conduct disorder and attention-deficit hyperactivity disorder. Through these examples, this paper aims to explain the different formulations of the FMM, the various steps in building a FMM, as well as how to decide between a FMM and alternative models. The third paper explores of the use of two mixture model as potential phenotypes in ACE analysis: LCA and FMM. The use of these models as phenotypes is demonstrated through an example concerning conduct disorder in a sample of Finnish twins. A discussion about extending the models in this dissertation to be applicable to longitudinal data or include gene by environment (GxE) interactions is also presented.

Chapter One: Introduction

Unobserved heterogeneity presents a problem when analyzing behavioral data from randomized control trials (RCTs). Often, data are frequently analyzed as if they are obtained from a single, homogenous population, although it is unlikely that all individuals in a sample have the same set of parameter values (Muthén, 1989). Treating the sample as homogenous, when it is not, may bias the model parameters and impact the inferences made about the effectiveness of a treatment. Mixture modeling is one technique that can be used to analyze data with unobserved heterogeneity resulting from RCTs (Muthén et al., 2002). Mixtures help to model unobserved heterogeneity in a population by identifying different groups of individuals within a population. In RCTs, mixtures are especially useful because they help to identify groups of individuals for whom a treatment is effective. This dissertation will explore specific mixture models and their application to behavioral data.

United States schools and students suffer from problems associated with student behavioral disorders. Examples of behavioral disorders include anti-social behavior, mood disorders, and drug and alcohol dependence. These behavioral disorders can have negative effects on the student as an individual, and on the student's classroom and school. Students affected with behavioral disorders are more likely to suffer from failing grades, attendance problems, school removal, having to repeat grades and feelings of not fitting in (Loeber & Farrington, 2000; Lewis et al., 2004) than their unaffected peers. Classrooms with affected students can suffer from classroom management issues and can have negative effects on the unaffected children in the classroom. Schools have to stretch

already limited resources to provide students with behavioral problems with specialized treatments and curricula in order to help them learn. There is a need for innovate statistical methods to analyze data to better analyze data on students with behavioral disorders, which may help overcome the problems stated.

One way to study the effectiveness of a treatment or intervention for students that are affected by behavioral disorders is a randomized control trial (RCT). In a RCT, students or classrooms are randomly assigned into either the intervention condition, where they receive treatment, or the control condition, where they do not receive treatment. This ensures that both known and unknown confounding factors are evenly distributed between the groups. An example of an educational RCT is the Good Behavior Game (GBG) study (Kellam et al., 1994). The GBG study applied a universal intervention aimed at reducing aggressive-disruptive behavior during first and second grade to improve reading and reduce aggression, with outcomes assessed through middle school and beyond.

One limitation of educational RCTs that look at behavioral outcomes is that they typically do not collect any genetic information. There is emerging evidence from both quantitative and molecular genetics studies that there are genetic influences on behavioral disorders. For example, evidence from twin studies (Fu et al. 2002; Hicks et al., 2004; Krueger et al., 2002; Silberg et al., 2003) point to common genetic influences on both antisocial behavior and drug and alcohol dependence. Importantly, a number of genes in the dopaminergic system (e.g., DRD4 and DAT1) have been shown to be associated with impulsiveness (e.g., Keltikangas-Järvinen et al., 2004) and response inhibition (Cheon et

al., 2005; Cornish et al., 2005; Galili-Weisstub et al., 2005), which are both risk factors for antisocial behavior and drug abuse/dependence (Gjone & Stevenson, 1997; Goldsmith & Lemery, 2000; Tarter et al., 1985).

The growing evidence of genetic influences on the development of behavioral disorders suggests that genetic factors may contribute to variation in developmental and intervention outcomes. One way that they can contribute to variation in outcomes is through gene by environment interactions. For example, Tuvblad et al. (2006) found that socioeconomic factors moderated genetic and environmental influences on adolescent antisocial behavior. Similarly, Caspi et al. (2003) reported that a functional polymorphism in the promoter region of the serotonin transporter (5-HT T) gene moderated the influence of stressful life events on depression. Collecting genetic information can help researchers that conduct educational RCTs better understand the effectiveness of their interventions.

One educational RCT that is collecting genetic information in their next wave of data collection is the GBG study (Ialongo, N., personal communication, August 26, 2008). Randomized controlled trials, such as the GBG, provide an opportunity to study genotype by environment interactions with a minimal amount of confounding between genetic and environmental influences given that the participants are randomly assigned to intervention conditions. For example, assume that the GBG study finds an association between antisocial behaviors and a specific candidate gene(s) or genomic marker(s) among the control group participants, but fails to find such an association, or finds a significantly weaker one, among intervention participants. The fact that participants were

randomly assigned to intervention conditions provides a relatively strong basis for inferring that the intervention served as a moderator of gene expression. Put more simply, the intervention may have interrupted the effect of that specific gene(s) or genomic marker(s) on antisocial behavior.

This dissertation focuses on explicating certain mixture models and extending them to include genetic information in a series of three papers. While the methods used in these papers are not currently applied to education data because of the lack of available genetic information, they can be applied to this type of data once the data becomes available. It is important that the methods used to analyze educational data with genetic information be determined ahead of time so that once genetic information becomes available researchers are able to analyze the data efficiently.

Each paper in this dissertation focuses on addressing different issues in the application of mixture models to behavioral data. The first paper, titled "Relating latent class analysis results to variables not included in the analysis," investigates how the method chosen to incorporate covariates into a mixture model can impact the estimates and standard errors in a regression or the results of a mean comparison test. By including covariates into an analysis evaluating the effectiveness of an RCT, rival hypotheses, which might also explain the results, can be ruled out. Two possible ways of testing the effect of covariates are including the covariates as independent variables in a regression analysis or conducting a mean difference test. This paper explores the quality of estimates and standard errors incurred when researchers use the five regression approaches discussed in the paper. Additionally, this study investigates how using most likely class

membership in mean comparison testing can potentially distort the test statistic and its interpretation. Using Monte Carlo simulations, this study examines how various factors, such as the degree to which people are classified correctly into latent classes, can impact the estimates and standard errors of auxiliary variable effects and testing mean equality across classes. Based on the results of the real data examples and simulations, suggestions are made about how to select covariates for an analysis.

The second paper, titled "Models and strategies for factor mixture analysis: Two examples concerning the structure underlying psychological disorders", uses a factor mixture model (FMM) to explore the underlying structure of psychological outcomes. In the psychological literature there has been debate about whether the underlying structure of psychopathology is categorical or dimensional. This paper discusses the FMM and how, through the use of two kinds of latent variables, the FMM models the underlying structure to be simultaneously categorical and dimensional. While the conceptualization of the FMM has been explained in the literature, the use of the FMM is still not prevalent. One reason is that there is little research about how such models should be applied in practice and, once a well fitting model is obtained, how it should be interpreted. The second paper explicates the FMM by discussing its most general form and how the FMM and its interpretation can change based on the amount of measurement invariance in the model. How to apply the FMM in practice, including how to decide on the number and types of latent variables, detecting the presence of measurement invariance, and comparing the FMM to alternative models, is demonstrated through the use of two examples on conduct disorder and attention-deficit hyperactivity disorder (ADHD).

The third paper, titled "ACE Analysis for Latent Variable Phenotypes", explores how latent variable phenotypes can be used in genetic analyses of behavioral outcomes. Two common phenotypes used in behavioral genetics are affection status and a sum score of symptom items. While these phenotypes are easy to obtain, they are not without limitations. The third paper proposes two different mixture models to use as phenotypes in a behavioral genetic analysis to overcome the limitations of traditional phenotypes. The first latent variable phenotype is the latent class analysis (LCA) model, which uses a categorical latent variable to account for the unobserved heterogeneity in a sample. The second latent variable phenotype is the FMM with measurement invariance. This paper elucidates the LCA and FMM as phenotypes and shows how they can be combined with an ACE model, which decomposes the variance of a phenotype in order to explore how much variation is attributable to genes. The use of these models as phenotypes is demonstrated through an example concerning conduct disorder in a sample of Finnish twins.

References

Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, 301, 386-389.

Cheon, K.A., Ryu, Y.H., Kim, J.W., & Cho., D.Y. The homozygosity of 10-repeat allele at dopamine transporter gene and dopamine transporter density in Korean children with attention deficit hyperactivity disorder: Relating to treatment response to methylphenidate. *European Neuropsychopharmacology*, 15:1, 95-101.

Cornish, K.M., Manly, T., Savage, R., Swanson, J., Morisano, D., Butler, N., Grant, C., Cross, D., Bentley, L., & Hollis, C.P. (2005). Association of the dopamine transporter (*DAT1*) 10/10-repeat genotype with ADHD symptoms and response inhibition in a general population sample. *Molecular Psychiatry,* 0, 686-698.

Fu, Q., Heath, A.C., Bucholz, K.K., Nelson, E.C., Glowinski, A.L., Goldberg, J., Lyons, M., et al. (2002) A twin study of genetic and environmental influences on suicidality in men. *Psychological Medicine*, 31:1.

Galili-Weisstub, E., Levy, S., Frisch, A., Gross-Tsur, V., Michaelovsky, E., Kosov, A., Meltzer, et al. (2005). Dopamine transporter haplotype and attention-deficit hyperactivity disorder. *Molecular Psychiatr,* 10, 617-618.

Gjone, H., & Stevenson, J. (1997). The association between internalizing and externalizing behavior in childhood and early adolescence: genetic and environmental common influence. *Journal of Abnormal Child Psychology*, 25:4, 277-286.

Goldsmith, H.H., & Lemry, K.S. (2000). Linking temperamental fearfulness and anxiety symptoms: A behavior-genetic perspective. *Biological Psychiatry*, 48:12, 119-1209.

Hicks, B.M., Krueger, R.F., Iacono, W.G., McGue, M., & Patrick, C.J. (2004). Family transmission and heritability of externalizing disorders: A twin-family study. *Archives of General Psychiatry*, 61, 922-928.

Kellam, S.G., Rebok, G.W., Ialongo, N., and Mayer, L.S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry*, 25, 359-382.

Keltikangas-Järvinen, L., Räikkönen, K., Ekelund, J., & Peltonen, L. (2004). Nature and nuture in novelty seeking. *Molecular Psychiatry*, 9, 308-311.

Krueger, R.F., Hicks, B.M., Patrick, C.J., Carlson, S.R., Iacono, W.G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior, and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology*, 111, 411-424.

Lewis, L.M., Davis, B., & Hops, H. (2004). Childhood social predictor of adolescent antisocial behavior: Gender differences in predictive accuracy and efficacy. *Journal of Abnormal Child Psychology*, 27:4, 277-292.

Loeber, R., & Farrington, D.P. (2000). Young children who commit crime: Epidemiology, developmental origins, risk factors, early interventions, and policy implications. *Development and Psychopathology*, 12, 737-762.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Presidential

   address to the Psychometric Society, July, 1989. *Psychometrika*, 54, 557-585.

Silberg, J., Rutter, M., D'Onofrio, D., Eaves, J. (2003). Genetic and environmental risk

   factors in adolescent substance use. *Journal of Child Psychology and Psychiatry*,

   44:5, 664-676.

Tarter,R., Alterman, A., Edwards, K. (1985). Vulnerability to alcoholism in men: a

   behavior-genetic perspective. *Journal for the Study of Alcohol*, 46: 329-351.

Tuvblad, C., Grann, M., & Lichtenstein, P. (2006). Heritability for adolescent antisocial

   behavior differences with socioeconomic status: gene-environment interaction.

   Journal of Child Psychology and Psychiatry, 47:7, 734-743.

Chapter Two:

Relating Latent Class Analysis Results to Variables not Included in the Analysis

Mixture modeling in the form of latent class analysis and growth mixture modeling has become an important tool for researchers (for an overview see Muthén, 2008). Mixtures help model unobserved heterogeneity in a population by identifying different latent classes of individuals based on their observed response pattern. An important interest in mixture modeling is the investigation of what types of individuals belong to each class by relating classes to covariates, concurrent outcomes, and distal outcomes, also known as auxiliary variables. This paper compares techniques for relating latent classes to auxiliary variables.

As a first step in investigating the relationship between latent classes and auxiliary variables, many researchers utilize mean comparisons tests, such as *t*-tests, ANOVAs, or chi-square tests, to establish whether a relationship is present. In order to conduct these tests, the first step is to estimate the mixture model based only on latent class indicators, obtaining each individual's most likely class membership, with assignment into classes being based on the highest probability of being in a given class. Using these assigned class memberships, the mean comparison tests can then be performed.

Regression models are used to further explore the relationship between latent classes and auxiliary variables. There are four commonly used regression approaches:

- Most likely class regression: Regression of most likely class membership on the covariates,

- Probability regression: Regression of an individual's logit-transformed posterior probability to be in a given class on the covariates,

- Probability-weighted regression: Regression that is weighted by an individual's posterior probability to be in a given class,

- Single-step regression: Including the covariates in the analysis while forming the latent classes.

In both the mean- and regression-oriented approaches, a problem with using most likely class membership is that class is treated as an exact, observed variable. To illustrate this problem, suppose a two-class model and take two individuals, one with a probability of 1.0 for belonging to Class 1 and 0.0 for Class 2 and the other with a probability of 0.51 for belonging to Class 1 and 0.49 for Class 2. Both individuals would be assigned and treated as members of Class 1 in the subsequent analyses. But the analyses does not take into account that the two individuals have different probabilities of being in the same class and instead are treated as if they both have a probability of 1.0 of being in Class 1. This will distort estimates because individuals are forced into their most likely latent classes. The standard errors will also be incorrect because the analysis does not take into account the uncertainty of the classification but treats it as an observed variable. This poses a problem because incorrect standard errors can lead to erroneous conclusions about the significance of an effect.

As in the most likely class regression, the first step of the probability and probability-weighted regressions is the estimation of the latent class model based only on the latent class indicators. In the second step, instead of having assigned class

membership as the outcome, the probability regression uses an individual's logit-transformed posterior class probability as the outcome. For the probability-weighted regression, the regression of class membership on the covariates is weighted by each individual's posterior probability. Using the probabilities of being in a given class may give less bias to regression coefficients but is still problematic because the probabilities are estimates and an analysis will not take into account the error associated with those estimates. So, the standard errors of a regression between the posterior probabilities and an auxiliary variable will be incorrect.

In the single-step approach, the problem of incorrect estimates and standard errors is circumvented because the analysis allows individuals to be fractional members of all classes and the latent class variable is not treated as observed. Such an approach, however, may be cumbersome when many auxiliary variables are involved because of the increased computation time associated with the inclusion of more auxiliary variables. Furthermore, a researcher may not always want auxiliary variables to influence the determination of class membership because the inclusion of auxiliary variables can potentially change the substantive interpretation of the latent classes.

A fifth approach, which has recently been put forward, is pseudo-class draws (Asparouhouv & Muthén, 2007; Wang et al., 2005). Here, several random draws are made from each individual's posterior probability distribution to determine an individual's class membership. Based on these draws, mean tests and regression estimates can be computed.

This paper explores the quality of estimates and standard errors incurred when researchers use the five regression approaches introduced above. Additionally, this study investigates how using most likely class membership in mean comparison testing can potentially distort the test statistic and its interpretation. Using Monte Carlo simulations, this study also examines how various factors, such as the degree to which people are classified correctly into latent classes, can impact the estimates and standard errors of auxiliary variables and testing mean equality across classes. Based on the results of the real data examples and simulations, suggestions are made about how to select covariates for an analysis.

The first section of this paper introduces the latent class analysis model and describes the approaches for examining the relationship between the latent classes and auxiliary variables. The next section provides two real data examples to demonstrate the problem of treating class membership as an observed variable and also to show how incorrect the estimates and standard errors can be when including many auxiliary variables. The third section describes the simulation study and its results to confirm the results of the real data examples as well as to show the extent of the problem. The final section, presents highlighted results, suggests under what conditions it is appropriate to use the methods examined, and recommends a process by which to select auxiliary variables for an analysis.

Background

Latent Class Analysis Model

The latent class analysis (LCA) model, introduced by Lazarsfeld and Henry (1968), is used to identify subgroups, or classes, of a study population. A diagram of an example of a latent class analysis model is shown in Figure 2.1a. There are two major concepts depicted in Figure 2.1a, the latent class itself and the observed outcomes or items that define the class. These can be seen in Figure 2.1 as the C, and u_1-u_r, respectively. The boxes, u_1 to u_r, represent the observed response items or outcomes. The outcomes in a LCA model can be categorical or continuous, though this paper will specifically focus on dichotomous, categorical items. The circle with the letter C in the middle is the unordered, categorical latent class variable with K classes. The arrows pointing from the latent class variable to the boxes above indicate that those items are measuring the latent class variable. This means that class membership is based on the observed response pattern of items. An important assumption, called the conditional or local independence assumption, implies that the correlation among the observed outcomes is explained by the latent class variable. Because of this, there is no residual correlation between the items.

For a LCA model with categorical outcomes, there are two types of model parameters: conditional item probabilities and class probabilities. The conditional item probabilities are specific to a given class and provide information about the probability that an individual in that class will endorse that item. The class probabilities specify the relative size of each class, or the proportion of the population that is in a particular class.

14

Consider a LCA model with r observed binary items, u, has a categorical latent variable C with K classes ($C = k$; $k = 1, 2, \ldots$, K). The marginal item probability for item $u_j = 1$ is

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

Assuming conditional independence, the joint probability of all the r observed items is

$$P(u_1, u_2, \ldots, u_r) = \sum_{k=1}^{K} P(C = k)P(u_1 | C = k)P(u_2 | C = k)\ldots P(u_r | C = k). \tag{2}$$

A product of LCA is the estimated class probabilities for each individual, called posterior probabilities, analogous to factor scores in a factor analysis (Muthén 2001). These are estimates of

$$P(C = k | u_1, u_2, \ldots, u_r) = \frac{P(C = k)P(u_1 | C = k)P(u_2 | C = k)\ldots P(u_r | C = k)}{P(u_1, u_2, \ldots, u_r)}. \tag{3}$$

Note that each individual is allowed fractional class membership and may have non-zero values for many classes. It is from these probabilities that class membership is assigned. An individual is assigned to be a member of a class based on their highest probability of being in a given class, even though an individual may have several classes to which they are a partial member. Once assigned to a class, an individual is assumed to be a part of that class 100%. Therefore, the fractional membership in multiple classes disappears when using most likely class membership as an observed variable in subsequent analyses.

One way to examine how well individuals have been classified is to look at the entropy of the latent class model using the following formula:

$$E_k = 1 - \frac{\sum_i \sum_k (-\hat{p}_{ik} \ln \hat{p}_{ik})}{n \ln K}, \tag{4}$$

where $\hat{p}_{ik}$ denotes the estimated posterior probability for individual i in class k. Entropy is measured on a zero to one scale with a value of one indicating the individuals are perfectly classified into latent classes. Higher values of entropy indicate better classification of individuals.

Mean Comparison Testing. One method to investigate the relationship between the latent classes and auxiliary variables is to conduct a mean difference test across the classes. This is done by first estimating the latent class model and then classifying people into latent classes based on their highest posterior probability of being in a given class. These memberships are then used as the groups in the mean comparisons tests, although, which test to use depends on the type of covariate.

If the covariate is continuous and approximately normally distributed, a one-way analysis of variance (ANOVA) is appropriate because the test requires a continuous outcome and can handle having two or more latent classes. When there are only two class means to compare, one could also choose to use a *t*-test but, with only two classes, the *t*-test and the ANOVA are equivalent. One assumption of ANOVA is that being a member of one class has no effect on whether an individual is a member of a different class. This assumption is incorrect when using latent classes as the group in an ANOVA because if an individual is not a member of class one, then they have to be a member of class two. Even though the analyses do not meet the assumptions of this test, it is still commonly

used by researchers when comparing the mean of an outcome across latent classes. So, it is important to investigate the impact of using this test.

If the covariate is categorical, a chi-square test of equal proportions will test whether the frequency of the covariate is the same across all classes. The chi-square test can be computed as:

$$Q_P = \sum_{j=1}^{K} \frac{(f_j - e_j)^2}{e_j},$$ (5)

where f_j denotes the frequency of the covariate in class j (or the number of observations in class j) for $j = 1, 2, \ldots, K$. Since the null hypothesis specifies equal proportions of the total sample size for each class, the expected frequency for each class equals the total sample size divided by the number of classes, or:

$$e_j = n / K \text{ for } j = 1, 2, \ldots, K.$$ (6)

Most Likely Class Regression. In this approach, most likely class membership is related to the auxiliary variables. First, the LCA is conducted based on the latent class indicators and individuals are assigned to their most likely class. In the second step, the membership assignments from the previous step are used as an observed variable in the regression. In the covariate case, the assigned membership, m, which can take on values from 1 to K, is regressed onto the covariate using a multinomial logistic regression where:

$$P(m_i = k \mid x_i) = \frac{e^{\alpha_k + \beta_k x_i}}{\sum_{s=1}^{K} e^{\alpha_s + \beta_s x_i}},$$ (7)

17

where $\alpha_K = 0$, $\beta_K = 0$ so that $e^{\alpha_K + \beta_K x_i} = 1$, implying that the log odds of comparing Class k to the last Class K is

$$\log [P(m_i = k \mid x_i) / P(m_i = K \mid x_i)] = \alpha_k + \beta_k x_i . \qquad (8)$$

In the case of a distal outcome, the distal outcome is regressed on the assigned membership and the type of regression used will depend on the distal outcome variable type.

Probability regression and probability-weighted regression. One way around using the assigned class membership, is to use class probabilities. Two ways of incorporating class probabilities are discussed in this study. The first method is the probability regression in which an individual's posterior class probability is converted to the logistic scale and then regressed on the covariates using a linear regression. The probabilities are converted because, on a probability scale, the values can only range from 0 to 1, which is not suitable for linear regression. On a logistic scale, the converted probabilities can have any value. In the second approach, the regression is weighted by each individual's posterior probability of being in a given class.

Using the probabilities of being in a given class may give less bias to regression coefficients but is still problematic because the probabilities are also estimates and an analysis will not take into account the error associated with those estimates. So, the standard errors of a regression between the posterior probabilities and an auxiliary variable will be incorrect. This poses a problem because incorrect standard errors can lead to erroneous conclusions about the significance of an effect.

Single-Step Approach. In order to see who is a member of each class, covariates can be added to the LCA model. Figure 2.1b shows the diagram of this model. As with the plain LCA model, there is a latent class variable, C, and the items that measure the latent class variable. What has been added is the box to the left of the latent class variable, with the x in the middle, which represents the covariate. There is an arrow starting from the covariate box and ending at the latent class variable. This indicates that the latent class variable is being regressed on the covariate, x. More specifically, this regression is a multinomial logistic regression because the outcome, the latent class variable, is categorical with potentially more than two categories.

LCA with covariates has been considered by Bandeen-Roche et al. (1997), Dayton and Macready (1988), Formann (1992), and Heijden et al. (1996). This modeling considers a covariate, x, where the probability that individual i, falls in class k of the latent class variable C is expressed through multinomial logistic regression as

$$P(C_i = k \mid x_i) = \frac{e^{\alpha_k + \gamma_k x_i}}{\sum_{s=1}^{K} e^{\alpha_s + \gamma_s x_i}},$$

(9)

where $\alpha_K = 0$, $\gamma_K = 0$ so that $e^{\alpha_K + \gamma_K x_i} = 1$, implying that the log odds of comparing class k to the last class K is

$$\log [P(C_i = k \mid x_i) / P(C_i = K \mid x_i)] = \alpha_k + \gamma_k x_i .$$

(10)

Muthén and Muthén (2000) gave an example of LCA with covariates applied to antisocial behavior classes related to age, gender, and ethnicity.

Similar to covariates, distal outcomes can be added to the LCA model to see how class membership predicts the distal outcome. Here, the latent class variable is an

exogenous instead of endogenous variable. This is one method that substantive

researchers use to investigate the predictive validity of the latent classes. Figure 2.1c

shows an LCA model with a distal outcome. Again, there is the traditional latent class

model, but the addition is the box to the right of the latent class variable with a y in the

middle. This box represents the distal outcome. There is also an arrow pointing from the

latent class variable to the distal outcome box indicating that the distal outcome is being

regressed on to the latent class variable. This regression can be linear, logistic, or another

type of regression depending on the form of the distal outcome.

Pseudo-class draws. Instead of using assigned class membership, another option

is to use pseudo-class draws. When doing latent class analysis, every individual has a

posterior class distribution or, stated differently, every individual has a posterior

probability of being in each class. The distribution of these probabilities is multinomial

because there are potentially more than two classes. The draws are made by taking

random samples from this multinomial distribution. By having multiple random samples,

individuals are given a chance to change membership to neighboring classes, which gives

a sense of the variation associated with the distribution. The pseudo-class draws are

similar to multiple imputation in missing data analysis, except in this case, the latent

classes are missing. Given the pseudo-class draws, class specific means, variances, mean

equality tests and regressions can be computed. For a more technical treatment of pseudo-

class draws, see Asparohouv & Muthén (2007) and Wang et al. (2005).

Previous Work. While many researchers have identified the problem of using

class membership as an observed variable in an analysis, none have shown how

problematic it can be. Hagenaars (1993) points out that because there can be high misclassification of individuals into classes, using most likely class assignment can be problematic because those individuals that are assigned to the wrong class will skew the true relationship between the classes and the external variables. Clogg (1995) also points out the dangers of using this strategy and named it the *classify-analyze* strategy. Nagin and Tremblay (2001) show that after individuals are assigned to classes, comparisons can be made across the classes on an outcome of interest by inspecting the means of the outcome. They continue on to point out that this technique is problematic because there is no valid basis for computing the standard errors and therefore confidence intervals or hypothesis tests cannot be computed (Roeder et. al. 1999). Heijden et al. (1996) offer that one way to avoid the problems of the *classify-analyze* strategy is to estimate class membership and the relationship to external variables all in one step. Heijden also points out that there are three advantages to conducting the analyses in one step. First, by doing the analyses in one step the classification issue is avoided. Second, it is better to work with one model and its model fit than to worry about model fit for two separate models, the latent class analysis and the subsequent regression. Finally, it is possible to investigate models that have zero degrees of freedom or are unidentified in ordinary latent class analysis by using covariates in the analysis.

This study shows how much of a problem it is to use class membership as an observed variable by demonstrating how distorted the estimates and standard errors of a covariate or distal outcome effect can be. It also shows results for a new approach for relating latent classes to auxiliary variables, namely using pseudo-class draws.

Real Data Examples

In this section two real data examples will be discussed to explore how the problem of treating assigned class membership as an observed variable is compounded when there are multiple auxiliary variables in an analysis. The first example examines antisocial behavior in 16-23 year olds. The second example investigates aggressive behavior of first grade children in the classroom.

Antisocial Behavior Example

The data for this example are the Antisocial Behavior (ASB) data which were taken from the National Longitudinal Survey of Youth (NLSY) that is sponsored by the Bureau of Labor Statistics. These data are made available to the public by Ohio State University. The data were obtained as a multistage probability sample with an over-sampling of African Americans, Hispanics, and economically disadvantaged non-black and non-Hispanics.

The ASB data include 17 antisocial behavior items that were collected in 1980 when respondents were between the ages of 16 and 23. The ASB items assessed the frequency of various behaviors during the past year. A sample of 7,326 respondents has complete data on the antisocial behavior items and the covariates used in this example. For covariates, gender, age and ethnicity will be considered. Gender and ethnicity are dichotomous variables with ethnicity being split into two separate variables, Black and Hispanic, which are referenced to the predominant ethic group in the sample, White.

The data were first analyzed by conducting the LCA and inclusion of the covariates in one single step. The next phase was to conduct the other strategies:

regressing most likely class membership on the covariates, regressing the class

probabilities on the covariates, using the class probabilities as weights in a regression, or

pseudo-class regression. Also, once class membership was established, the mean

comparison tests were conducted. The results of these analyses can be seen in Table 2.1

for the mean comparisons and Table 2.2 for the regression. In Table 2.2, there are four

classes in the analysis, which means that when the latent classes are regressed on the

covariates a multinomial logistic regression is used. Because of this, the results are in

terms of logits and the regressions are in reference to the last class which has been set to

be the largest class.

Before discussing the results, it is important to understand the class structure of

this dataset so it can be understood how using different methods of incorporating

covariates can affect the overall interpretation. Figure 2.2 shows a class profile plot for

these data with four classes. The first class, and the smallest class in size, was considered

to be a high antisocial behavior class because the probability of endorsing almost all of

the items was high. The second class was considered the person offense class because the

items that had a high probability of endorsement are items related to aggression against

another person such as fighting. The third class was considered the drug class because

only items relating to drugs were highly endorsed. The fourth class, and largest in size,

was considered the normative class because almost all items had a low probability of

endorsement, with marijuana use being the only exception. The entropy was 0.74.

The results in Table 2.1 suggested that the pseudo-class Wald chi-square and

equal proportion chi-square yielded similar results. The absolute values of the chi-square

statistics were similar across both types of tests, with the equal proportions chi-square test yielding higher values. Additionally, the significance of the tests was equivalent across all the covariates, though this may be due to chi-square being sensitive to the large sample size of this data set. Even though the values and significance of the tests were similar across the two methods, the potential for problems becomes apparent. Differences between the tests are likely to arise when the tests are close to the borderline of being significant. When this occurs it is likely that the equal proportions chi-square will lead one to conclude that there are significant differences across the classes because it has higher test values, and the pseudo-class Wald chi-square will lead one to conclude that there are no differences across the classes.

As is shown in Table 2.2, how covariates are incorporated into an analysis can affect the interpretation of the relationship between the covariate and the latent class variable. Using the assigned class membership, the class probability-weighted regression or pseudo-class draws yielded smaller regression estimates and standard errors than conducting the analyses in one step. Pseudo-class regression estimates were the furthest away from the estimates obtained when including the covariates in the analysis, followed by probability-weighted regression. Most likely class membership estimates and standard errors came closest to the results from including the covariates in the analysis.

In Table 2.2, the results for the probability regression are presented on a continuous dependent variable scale while the rest of the table is presented on a logistic scale. Because the results in the table were not directly comparable, the comparison between the methods will focus on the sign and significance of the regression estimates.

In most of the regressions, the sign and significance of the regression estimate were the same across all methods. For the regressions of the drug class on gender and the person offense class on Hispanic, the sign of the probability regression estimate differed from the other methods. Additionally, for the person offense class on Hispanic regression, the estimate was significant for the probability regression, but it was insignificant for the other methods.

One interesting point of difference between the five approaches occurred in the case of the regressions of the high class on Black and Hispanic. In these two cases, using most likely class membership, probability regression, probability-weighted regression, or pseudo-class regression would lead to the conclusion that the estimate was significant since the ratio of the estimate to the standard error is greater than 1.96. But, including covariates while forming the latent classes leads one to conclude that the estimate was insignificant because the ratio of the estimate to the standard error was less than 1.96. These two cases highlight the problem associated with underestimated standard errors. This also suggests that the standard error difference among the methods is larger than the difference in the estimates.

This example has shown that using assigned class membership regression, the class probability-weighted regression, or pseudo-class draw regression underestimates the regression effects of the classes on the covariates. The probability regression sometimes showed different signs and significance of the regression estimates when compared with the other method. The mean comparisons had different values for each test, but the

overall significance was the same which may be in part do to chi-square's sensitivity to large sample sizes.

Aggression Example

The data for this example come from a randomized preventive field trial conducted in Baltimore public schools (Dolan et al., 1993; Ialongo et al., 1999). The study applied a universal intervention aimed at reducing aggressive-disruptive behavior during first and second grade to improve reading and reduce aggression with outcomes assessed through middle school and beyond (Kellam et al., 1994). The outcome variables of interest are teacher ratings of each child's aggressive behavior in classroom. The ratings were made using the Teacher's Observation of Classroom Adaptation Revised (TOCA-R) scaling instrument, with 10 items. For simplicity, the items were dichotomized with a value of one representing presence of the aggressive behavior and zero representing no symptom presence. Information was also collected on concurrent and distal outcomes, including school removal and juvenile court records.

This example will focus on analyzing the pre-intervention data of Cohort 1 when the children entered into the intervention study at the beginning of first grade. In total, there were 1,174 children participating in this study. For covariates, gender, ethnicity, and treatment condition will be considered. All three covariates are dichotomous with gender indicating whether an individual is male, ethnicity indicating whether a child is African American, as opposed to Caucasian, and the treatment variable indicating whether a child's classroom was assigned to the treatment or control condition. Since this example focuses on a pre-intervention time point and classrooms were assigned to

26

treatment or condition randomly, the relationship between the latent class variable and the treatment condition indicator should be non-significant.

This analysis was executed similarly to the ASB example. The data was first analyzed by conducting the LCA and inclusion of the covariates in a single step. The next phase was to conduct the other strategies: regressing most likely class membership on the covariates, regressing the logistic class probabilities on the covariates, conducting a class probability-weighted regression, or pseudo-class regression. Also, once class membership was established, the mean comparison tests were conducted. As in the previous example, the probability regression results are on a different scale and comparison of this method to the others will focus on the sign and significance of the regression estimate. The results of these analyses can be seen in Table 2.3 for the mean comparisons and Table 2.4 for the regression results.

 A three-class solution was found to fit these data. The profiles for the TOCA data can be found in Figure 2.3. The first class had a high average probability of endorsing all of the items, so this class is named the high aggression class. The second class had moderate endorsement for those items dealing with verbal aggressions and low endorsement for the other items so this class is called the verbal aggression class. The final class had low endorsement of all items, was the largest percentage of the sample, and was named the normative class. The entropy was 0.79.

Comparing Table 2.3 to the results seen in the ASB example, a similar pattern emerged. When comparing the pseudo-class chi-square results to the equal proportions chi-square, the values of the statistics were similar in size with the equal proportions chi-

square having a higher value. For the gender and race covariates, the significance of the test was equivalent across the two statistics and both tests would lead one to conclude that the effects of gender and race were significantly different across the classes. For the treatment indicator, both tests indicate that there were no significant differences across the classes because the p-value associated with both statistics was greater than 0.05.

The regression results presented in Table 2.4 showed a similar pattern to what was seen in the ASB data example. A comparison of the most likely class membership regression, probability-weighted regression, and pseudo-class regression showed that all of the methods yielded similar estimates and the standard errors for the regression between the classes and the covariate. But, when compared with including covariates during the formation of the latent classes, the other three methods tended to produces slightly smaller estimates and standard errors. In all cases, except for the regression of the verbal Aggression class on Black, the significance of the effect was the same across the methods. For the regression of the verbal aggression class on Black, the pseudo-class regression was not significant, while the other three methods were significant.

When comparing the sign and significance of the probability regression estimate to the other methods, Table 2.7 showed that there were differences. For the regressions of the verbal class on gender and Black, the estimate from the probability regression was not significant, but the estimate was significant for the other methods. The regression of high class on treatment condition shows the opposite pattern, where the estimate was significant for the probability regression but not for the other methods. Any regression with treatment condition as a covariate was not expected to be significant because the

data being analyzed are a pre-intervention time point. A significant effect of any class on treatment condition would have suggested that randomization did not work. For the regression verbal class on gender, the probability estimate had a negative sign while the other methods had a positive one.

Comparing the two different data examples, there were consistencies in the performances of the methods investigated. For the mean comparisons, in both data sets, the statistics yielded similar values and led to similar conclusions about the significance of the covariate of interest. For the regression results, in the TOCA and ASB data, most likely class membership regression, probability-weighted regression, and pseudo-class regression performed similarly, but in comparison to including covariates while forming the latent classes the other methods tended to underestimate the effect and the standard errors. The probability regressions showed a tendency to have different signs and significance of the regression coefficients when compared to the other methods. A question to ask is under what conditions will these differences that are seen in the regression results arise? Potential explanations for the differences between the methods are how well individuals are being classified into their latent classes, sample size, and the size of the covariate effect. One way to investigate this question is to do a Monte Carlo simulation study where the settings of the analysis, such as sample size, entropy, and the size of the covariate effect, can be controlled to fully explore potential reasons for differences in the methods.

Monte Carlo Simulation Study

One of the main questions of this study is how to select important covariates for an analysis. This is investigated by comparing the effectiveness of the mean comparison and regression methods described above. One way to conduct these investigations is to do a Monte Carlo simulation study where the covariate effect, classification of people into latent classes, and model population parameters are known and how to treat latent class membership is explored under a variety of different conditions.

In this study, the sample size was chosen to be 250 and 1,000. The number of replications of data generation was chosen to be 500 for all the models since this assures sufficient reliability of the summary information.

For the model population, or the model from which the data is generated, there are two possible sources of variation that are examined in this paper: 1) degree to which individuals are correctly classified into latent classes, and 2) the strength of the relationship between the latent classes and the covariate. Both sources of variation are discussed in detail below.

Data Generation Models

The overall plan of the simulation was to generate the data using a latent class model with a covariate effect. Specifically, data was generated using the model shown in Figure 2.1b. Ten items were used in the analyses. All of the items in the analysis were chosen to be binary so that the conditional item probabilities, the probability of endorsing an item given membership in a specific class, could be used to distinguish among the

classes. Also, binary items are common when studying behavioral disorders because they can indicate either presence or absence of a symptom.

When generating latent class models, it is important to specify the number of classes in each model and what the class profiles look like. The number of classes for each of the latent class variables was chosen to be two for simplicity. Another key feature of latent class models is the class profile or item endorsement profile. One way to visually inspect the class profile is to plot the items on the x-axis and the conditional item probabilities are on the y-axis. In this study, a complex profile was used in which the two class profile cross. In Figure 2.1d, which shows an example of a complex profile, class one has high endorsement of the first five items and low endorsement of the last five. Class two has the opposite pattern of class one, with low endorsement of the first five items and high endorsement of the last five items.

One hypothesis is that as classification of people into latent classes becomes worse (as entropy approaches zero) treating the latent class variable as observed further distorts the estimate of the covariate or distal effect and the standard error of that parameter. In order to explore this hypothesis, different entropy settings were used in this study: perfect, high, medium and low entropy. For the perfect entropy case, the average entropy value was 1.0 meaning that people were perfectly classified into latent classes. The high entropy had an average value of 0.80 across the replications, which says that 80% of the time individuals were correctly classified in latent classes. The medium entropy case had an average value of 0.60 and the low entropy case had an average value of 0.40.

Also, the strength of the covariate or distal outcome effect was studied using two different settings. In Figure 2.1b, the covariate effect can be seen as the path emanating from the covariate x and ending at the latent class variable c. The first setting was to set this path to 0.5 indicating that there is a strong, positive relationship. The final setting assumes that there is no relationship at all and so the path was set to zero.

Overall, there were eight data generation models specified. There were four different entropy settings (perfect, almost perfect, high and low) and two different covariate effects giving eight data generation models.

Data Analysis Models

The data was generated using the model specifications that were described above. The data were then analyzed using each of the five different regression approaches described and once most likely class membership information was obtained, mean comparison tests were conducted. The simulation and analysis of the data was carried out using the Monte Carlo facilities in Mplus version 5.0 (Muthén & Muthén, 2008). Mplus input scripts for each method are presented in the Appendix A for the case where the sample size is 1,000 and the entropy is 0.8.

Two techniques were used to evaluate the mean differences across classes on the outcome of interest. The first uses pseudo-class draws to conduct a pseudo-class chi-square test. The other technique is a t-test which was conducted by classifying individuals into their most likely latent class and then comparing mean differences in each class using an independent samples t-test. In order to be able to compare the results of the pseudo-class chi-square test and the most likely class membership t-test, the value of the

t-test was squared in Table 2.5 in order to make the *t* value asymptotically equivalent to χ^2. In the real data examples, a chi-square test of equal proportions was compared to the pseudo-class chi-square test instead of a *t*-test because in both real data examples there were more than two classes and the covariates were categorical. The presence of only two classes and a continuous covariate in the simulation required that a *t*-test be used. As mentioned previously, five different regression approaches were used to analyze the data and obtain estimates and standard errors of the covariate or distal outcome effect: most likely class membership regression, probability regression, probability-weighted regression, pseudo-class regression, and single-step regression.

Simulation Results

This section summarizes results from the simulation study. The results of the mean comparison tests are first examined followed by a comparison of the five approaches used to incorporate covariates into a latent class analysis.

Mean Comparison Simulation Results. In some instances, it is not possible to include the covariates when conducting the latent class analyses. When this is the case, one way to examine differences in covariates is to compare the means across classes. As mentioned previously, two techniques were used to investigate which is the best method for testing mean equality: pseudo-class chi-square test and a most likely class membership *t*-test. The results for these comparisons are presented in Table 2.5.

In Table 2.5, the results of the means comparisons are broken down by the size of the relationship between the latent classes and the covariate. The top half of the table presents results for when there was no relationship between the latent classes and the

covariate. Because there was no relationship, it was expected that across all settings the tests will not be significantly different from zero. The *p*-value of the average statistic, across all settings, was greater than 0.05 which indicated that the test is not significant, which is what was expected. The bottom two rows of this case give an idea of the type I error, or when a difference is observed when in truth there is no difference, associated with these tests because these rows give the proportion of replications that have a significant p-value. Given that the significance level had been set to 5%, it was expected that if the test is working properly, 5% of the time the tests will find a significant difference when there is none. In the bottom row of the table, the pseudo-class chi-square had the proportion of replications giving significant *p*-values close to zero for nearly all the settings examined. This indicated that the test was not performing as it should because a 5% error rate was not recovered. For the most likely class membership *t*-statistic, the proportion had values from 0.036 to 0.052, which is close to the expected value of 0.05.

The bottom half of the table is the largest relationship with the class on covariate effect being 0.5. This setting showed similar results to what was seen in the real data examples with the average value of the statistics being close in size, but with the most likely class membership *t*-value average being larger than the pseudo-class chi-square. The difference between the averages of the two statistics increases in size as both the entropy and the sample size increases. More importantly, however, was that the decision of whether or not the covariate was significantly different across the classes did not change across the two methods. Both *p*-values lead to the conclusion that the covariate

34

was significantly different across the classes. Even though the average p-values for both statistics indicated that the covariate was significantly different across the classes, comparing the number of replications that had significant p-values ($p < 0.5$) did show a difference. For the smaller sample sizes, the most likely class membership t-statistic produced a larger number of significant replications. When the sample size was equal to 1,000 the number of replications with significant p-values increased when compared with the smaller sample size of 250. Additionally, as the entropy increased, the number of significant replication approached 500 or 100% of the replications showing significant p-values. One explanation for these results in the higher sample size was that the pseudo-class chi-square and the most likely class membership t-statistic, which, when squared, is asymptotically equivalent to a chi-square test, were picking up on a chi-square test's sensitivity to large sample sizes and will always conclude that covariate is significantly different across the classes.

The simulation study showed that the pseudo-class chi-square and the most likely class membership t-value yielded similar results for the magnitude of their values and the significance of the covariate effect. This was similar to what was seen in the real data examples.

Regression Simulation Results. The comparison of the approaches examined in this paper was executed by simulating data under a variety of conditions and then analyzing the simulated data using all five of the regression approaches: most likely class membership regression, probability regression, probability-weighted regression, pseudo-class regression, and single-step regression. When examining the results it is important to

examine the power and mean square error (MSE), in addition to the estimates of the coefficient. The power, or the probability of rejecting the null hypothesis when it is false, provides an estimate of whether there is enough information in the data to detect a covariate effect. The MSE quantifies the amount by which an estimate differs from the true value of the regression coefficient. The results of these analyses are displayed in Tables 2.6 and 2.7.

Table 2.6 displays the results for when the relationship between the latent class variable and the covariates is specified to be 0, or no relationship. The table rows are broken down into five sections, one for each of the methods used to incorporate covariates. The top of the table shows results for when most likely class membership is regressed on the covariates. Following that, there are sections for when the class probabilities are regressed on the covariates, when posterior class probabilities are used as weights in the regression of class membership on the covariates, pseudo-class regression, and when covariates are included while the latent classes are formed. The single-step regression was placed at the bottom of the table for easier comparisons since this was the model from which the data was generated and the other models will be compared to. Moving from left to right in the table, the entropy, or the degree to which individuals are correctly classified into their latent classes, increases.

Starting with the two rightmost columns of the table, where individuals were perfectly classified, a comparison of the five methods showed that for this entropy setting the methods were equivalent. When there was perfect entropy, every individual had a probability of 1.0 of being in their latent class, which means that there was no error

36

involved when classifications were made into most likely latent class. With the pseudo-class draws, all of the draws were equivalent to one another since there would be no class switching because individuals are perfectly classified.

Comparing the estimates obtained by the five methods, all of them were able to recover the true effect of zero, but when comparing the standard errors of the methods a similar pattern to what was seen in the real data examples emerged. When comparing the standard errors to including the covariates in the model, the four other methods underestimated the standard errors. Standard errors associated with using class probabilities as weights were the smallest overall.

The methods can be further compared by examining the mean square error (MSE), which quantifies the amount by which an estimate differs from the true value of the regression coefficient. An MSE value of zero indicates that there is no difference between the estimated regression coefficient and its true value. Thus, when comparing the MSE of two models, the model with the smallest MSE is interpreted as being the best model for explaining the variability in the observations. In Table 2.6, the MSE of each of the techniques was close to zero for all settings. When comparing among the techniques, for the two smallest entropies, pseudo-class regression and probability-weighted regression had the smallest MSEs. When the entropy increases to 0.80, probability regression had the smallest MSEs. But, the difference between the probability regression MSEs and the next smallest MSEs, from pseudo-class regression, was only 0.005.

One way of examining how well the parameters and their standard errors were being estimated in a simulation study is to look at the coverage. Coverage was defined as

the proportion of replications for which the 95% confidence interval contains the population parameter value (Muthén & Muthén, 2008). The closer the coverage was to one, the better the estimates of the parameters and their standard errors. Coverage was high across the methods and in all settings, with the lowest value in the entire table being 0.938. Probability-weighted regression had slightly smaller coverage values than the other methods, but they were still high.

In Table 2.6, the last row in each section of the table examines whether the correct alpha level is recovered. Because the population value for this table is equal to zero, the values in this row are estimates of Type I error, or the probability of rejecting the null hypothesis when it is false. If a method is performing correctly, then a value of 0.05 is expected based on the 5% significance level that was chosen for this test. For the most likely class membership, probability regression, class probability-weighted regression and including the covariates in the model, the values in these cells were all close to 0.05 which shows that these methods are performing as expected. The pseudo-class regression cells have values close to zero across all settings indicating that the method was not performing as expected.

In Table 2.7, where the relationship between the classes and covariates is 0.5, the equivalency of the methods that was seen when there was no relationship breaks down as the entropy decreases. The most likely class membership, probability, probability-weighted, and pseudo-class draw estimates were all smaller than the true value of 0.5, with the pseudo-class regression and probability-weighted regression having the smallest values across all settings. As the entropy settings decrease, the estimates for all four

methods get further away from the true value. Including the covariates in the models tended to slightly overestimate the true value across the settings with the biggest difference occurring for the small sample size and lowest entropy setting.

While not as big of a problem as having incorrect parameter estimates, there were also differences in the standard errors that each method is obtaining. As the entropy decreases, the standard errors for the most likely class membership, probability, class probability weighted regression, and pseudo-class regression were underestimated when compared with the ones obtained by including the covariates while forming the latent classes. Having underestimated standard errors is problematic because it could lead a researcher to conclude that an effect is significant when it may not be. The standard errors for the pseudo-class regression, while not exactly the same as the ones obtained when including the covariates during the formation of the latent classes, were relatively close but still underestimated while the standard errors for the probability-weighted regression are the most underestimated.

In Table 2.7, the MSEs of the each of the methods was relatively close to zero. When comparing among the methods, for the highest entropy settings, the methods have similar MSEs with probability regression having the highest MSEs and including covariates in the model having the smallest. For the lower entropy settings, including covariates in the models tended to have the smallest MSEs.

For the two highest entropy settings, the coverage was good with most methods and settings having coverage of at least 0.80, except in the probability case where the coverage was 0.50 or lower in the high entropy setting. As the entropy decreases, the

coverage remained around 0.95 for the case where covariates were included while forming the latent classes, which indicated that the parameters and standard errors were being estimated well for this method. For the two lower entropy settings, the methods using class probabilities had the lowest coverage of all the methods. The pseudo-class regression and most likely class membership had similar coverage across the settings, but it was still low in the two lowest entropy settings.

The last row in each section of Table 2.7, displays the proportion of the replications for which the null hypothesis that a parameter is equal to zero is rejected at the 0.05 level (two-tailed test with a critical value of 1.96). For parameters with population values different from zero, this value is an estimate of power with respect to a single parameter, that is, the probability of rejecting the null hypothesis when it is false. A proportion of 0.80 or greater is considered to be good power to reject that the parameter is zero. For the two highest entropy settings, all of the methods had a high proportion of replications for which the null hypothesis is rejected, except for pseudo-class regression with a sample size of 250 and an entropy setting of 0.80. This exception hints at what was seen in the rest of table, which was that pseudo-class regression had lower power across most of the settings when compared with the other four methods. This was especially noticeable in the smaller sample sizes and lower entropies in first few columns. But, as the entropies and sample sizes become smaller, the coverage does decrease across all methods. In the first column, with the small sample size and small entropy, none of the methods had a high power to reject the null hypothesis. Across all settings, probability-weighted regression had the highest power.

The simulation study shed light on the hypotheses. For the mean

comparison simulations, the results showed that the squared *t*-test produced higher test

statistic values than the pseudo-class Wald chi-square test but, the significance of the

tests was equivalent. For the regression simulation approaches, the results showed that for

approaches that do not include covariates during the formation of the latent classes, the

estimates and standard errors were different when compared to including the covariates.

Further comparisons among the regression methods and recommendations for which

method to use in practice will be made in the discussion section below.

Discussion

The purpose of this study was to investigate how the different methods for

treating latent class variables can impact the relationship between the latent classes and

auxiliary variables. This was investigated by exploring two different real data examples

that showed similar results when latent class membership was used as the dependent

variable in a regression analysis. This issue was further investigated by utilizing Monte

Carlo simulation techniques to explore how changing different settings in an analysis and

varying the technique by which latent classes were incorporated can impact the estimates

and standard errors of a regression analysis between latent classes and auxiliary variables.

Given the results of the real data examples and the simulation study, the question

of which method or methods are best to use to incorporate auxiliary variables in an

analysis arises. Table 2.6 shows that when there is no relationship between the

covariates, all five of the regression approaches examined are able to recover the true

effect of zero and that the effect is not significant. When there is no effect, it does not

matter which approach is used, but most researchers will not know ahead of time whether or not a covariate will be significant. When there is a relationship between the covariate and latent class the question posed becomes more difficult to answer. Across all of the settings, including the covariate while forming the latent classes performed the best. The method was able to recover the true effect and had high coverage and power in most settings. But this method could be problematic to use in many real data analyses because the inclusion of many covariates will significantly increase computation time and the covariates may potentially influence the formation and interpretation of the latent classes.

If including the covariates while forming the latent classes is not an option because of the reasons stated above, one alternative is to use the most likely class membership, but only if the entropy is 0.80 or greater. When the entropy was high, most likely class membership was the best performing method in terms of recovering the true value used in the simulation study, and had relatively good coverage and power in the settings examined. But, when using most likely class membership, researchers need to be aware that this method does have the potential for underestimating the standard errors of the parameter. Therefore, when deciding on the significance of auxiliary variables, a more stringent criterion than the 5% level for deciding on significance should be employed. If the entropy is lower than 0.60, it is unclear as to which method is best to use. None of the methods were able to recover the true effect and all had various problems with underestimated standard errors, coverage and power.

One issue that arises with latent class analysis, and other mixture models, is how to decide on which covariates should be included in an analysis when there are a large

42

number of covariates to select from. Based on the results of this study, a strategy for how to select covariates is suggested below. The first step is to conduct the latent class analysis without any covariates in the model in order to understand the substantive interpretation of the latent classes. In the second step, a pseudo-class Wald chi-square test is conducted to examine whether a covariate has some impact on the latent classes in terms of the mean differences. Next, a pseudo-class regression should be conducted to further winnow down the pool of potential covariates. A pseudo-class regression was chosen for this step since this method had the highest power to detect covariate effects of all the methods examined. Since the pseudo-class regression had biased estimates, as a final step, an analysis with the covariates included while forming the latent classes should be conducted in order to obtain unbiased estimates of the regression coefficient.

Limitations and Future Directions

A major limitation of all simulation studies is the failure of any study to cover all possibilities that are seen in real data analysis. In the present study, the simulations only considered the impact of one continuous covariate, but in many real data analyses, many covariates, of possibly different types, would be included. The simulation study, however, was not meant to cover all possible situations that researchers may encounter, but instead to point out that researchers need to be conscientious of the magnitude of the problem associated when using these different methods. Future work in this area should investigate how the addition of multiple covariates and different types and combinations of covariates impact the estimates and standard errors obtained by these methods.

This paper has considered the impact on estimates and standard errors when class membership is used as the outcome in a regression analysis. But this raises the question of what happens when class membership is used as a predictor in a regression of a distal outcome on class membership. Petras and Masyn (2009) considered this issue in their study looking at how to incorporate covariates and distal outcomes in growth mixture modeling. In their study, the authors found that treating class membership as an observed variable lead to different conclusions when compared with incorporating the distal outcome while forming the latent classes. Future work in this area should examine under what conditions it is appropriate to treat class membership as an observed variable when it is used as a predictor of a distal outcome.

One aspect that has not been discussed is the comparability of the regression models. With most likely class membership, probability regression, probability-weighted regression, and pseudo-class draws the covariate is included after the latent class analysis has already been conducted. In the single-step method, the covariate was included at the same time as the latent class analysis is conducted. By including the covariate with the latent class analysis, there is a potential for a direct effect between the covariates and the latent class indicators. This direct effect can potentially impact the relationship between the covariate and the classes because the covariate effect is partially filtered through the relationship between the covariate and the latent class indicators. In the simulation study, this was not problematic because the models were specified so that there was no direct effect between the covariate and the latent class indicators, but, the potential for direct

effects does arise in the real data examples. Nylund and Masyn (2008) are currently investigating the impact of mis-specifying direct effects in latent class analyses.

Despite the limitations discussed, this study is not without its strengths. While other authors have discussed the problems with using most likely class membership, this is one of the first studies to look at how problematic it can be when used in regression and mean comparisons. This study takes other authors' work one step further by also comparing the results using most likely class membership regression to those obtained from other methods: probability-weighted regression, probability, pseudo-class draws, and including the covariate while the latent classes are formed. This is the first study to make suggestions about when it is appropriate to use these techniques in practice.

Table 2.1

Antisocial Behavior Mean Comparison Results

	Sex	Black	Hispanic	Age
Pseudo-Class Chi-Square	287.6 (0)	112.9 (0)	8.49 (0)	101.2 (0)
Equal Proportions Chi-square	627.4 (0)	122.7 (0)	14.83 (0)	255.5 (0)

Table 2.2

Antisocial Behavior Regression Results

	Gender			Age		
	High	Person	Drug	High	Person	Drug
Most Likely Class Regression						
Est.	2.18	0.96	0.25	-0.16	-0.16	0.06
S.E.	0.11	0.06	0.07	0.02	0.02	0.02
Est./S.E.	19.4	15.9	3.78	-8.02	-11.1	3.85
p-value	0	0	0	0	0	0
Logistic Class Probability Regression						
Est.	2.77	0.35	-0.45	-0.19	-0.18	0.21
S.E.	0.117	0.084	0.087	0.027	0.019	0.02
Est./S.E.	23.65	4.22	-5.14	-7.15	-9.48	10.39
p-value	0	0	0	0	0	0
Probability-Weighted Regression						
Est.	2.04	0.92	0.24	-0.15	-0.15	0.06
S.E.	0.1	0.06	0.06	0.012	0.013	0.013
Est./S.E.	20.9	16.9	3.97	-8.28	-11.3	4.35
p-value	0	0	0	0	0	0
Pseudo-class Regression						
Est.	1.99	0.86	0.29	-0.14	-0.13	0.05
S.E.	0.12	0.07	0.07	0.02	0.02	0.02
Est./S.E.	17.3	11.8	3.97	-6.19	-7.63	3.12
p-value	0	0	0	0	0	0.002
Single Step Regression						
Est.	2.65	1.42	0.22	-0.22	-0.26	0.08
S.E.	0.16	0.09	0.1	0.03	0.02	0.21
Est./S.E.	16.1	15.5	2.14	-8.36	-10.8	3.61
p-value	0	0	0.03	0	0	0

	Black			Hispanic		
	High	Person	Drug	High	Person	Drug
Most Likely Class Regression						
Est.	-0.22	0.36	-0.72	-0.32	-0.02	-0.53
S.E.	0.1	0.07	0.08	0.12	0.08	0.09
Est./S.E.	-2.17	5.35	-8.66	-2.65	-0.24	-5.75
p-value	0.03	0	0	0.008	0.81	0
Logistic Class Probability Regression						
Est.	-0.53	1.19	-0.91	-0.37	0.42	-0.67
S.E.	0.134	0.096	0.1	0.158	0.11	0.12
Est./S.E.	-3.96	12.48	-9.11	-2.36	3.66	-5.77
p-value	0	0	0	0.018	0	0
Probability-Weighted Regression						
Est.	-0.23	0.36	-0.61	-0.31	-0.02	-0.47
S.E.	0.096	0.06	0.07	0.11	0.07	0.08
Est./S.E.	-2.41	5.98	-8.69	-2.78	-0.25	-5.81
p-value	0.02	0	0	0.005	0.8	0
Pseudo-class Regression						
Est.	-0.28	0.33	-0.6	-0.32	-0.01	-0.44
S.E.	0.11	0.08	0.09	0.13	0.09	0.09
Est./S.E.	-2.45	4.31	-6.79	-2.51	-0.11	-4.47
p-value	0.014	0	0	0.012	0.92	0
Single Step Regression						
Est.	0.22	0.84	-0.96	-0.19	0.21	-0.69
S.E.	0.15	0.12	0.14	0.134	0.12	0.12
Est./S.E.	0.16	6.94	-7.01	-1.46	1.77	-5.89
p-value	0.88	0	0	0.145	0.08	0

Table 2.3

Aggression Mean Comparison Results

	Race	Gender	Treatment
Pseudo-Class Chi-Square	8.37 (0.02)	42.57 (0)	0.33 (0.85)
Equal Proportions Chi-Square	13.95(0.001)	45.2 (0)	2.31 (0.32)

Table 2.4

Aggression Regression Results

	Gender		Black	
	High	Verbal	High	Verbal
Most Likely Class Membership as Outcome				
Est.	1.343	0.347	0.713	0.47
S.E.	0.199	0.14	0.206	0.154
Est./S.E.	6.743	2.478	3.457	3.057
p-value	0	0.013	0.001	0.002
Logistic Class Probability Regression				
Est.	2.719	-0.265	0.884	0.291
S.E.	0.356	0.23	0.38	0.246
Est./S.E.	7.64	-1.15	2.33	1.184
p-value	0	0.248	0.02	0.236
Probability - Weighted Regression				
Est.	1.228	0.339	0.621	0.366
S.E.	0.17	0.126	0.179	0.135
Est./S.E.	7.208	2.697	3.469	2.706
p-value	0	0.007	0.001	0.007
Pseudo-Class Regression				
Est.	1.199	0.335	0.56	0.313
S.E.	0.2	0.157	0.21	0.167
Est./S.E.	5.995	2.142	2.667	1.874
p-value	0	0.032	0.008	0.061
Covariates in Model				
Est.	1.52	0.38	0.716	0.442
S.E.	0.235	0.175	0.229	0.184
Est./S.E.	6.47	2.174	3.129	2.404
p-value	0	0.03	0.002	0.016

	Treatment	
	High	Verbal
Most Likely Class Membership as Outcome		
Est.	-0.005	0.192
S.E.	0.186	0.139
Est./S.E.	-0.025	1.382
p-value	0.98	0.167
Logistic Class Probability Regression		
Est.	-0.789	0.063
S.E.	0.359	0.232
Est./S.E.	-2.19	0.273
p-value	0.028	0.785
Probability - Weighted Regression		
Est.	-0.15	0.055
S.E.	0.167	0.126
Est./S.E.	-0.897	0.434
p-value	0.37	0.664
Pseudo-Class Regression		
Est.	-0.137	0.035
S.E.	0.191	0.156
Est./S.E.	-0.718	0.225
p-value	0.473	0.822
Covariates in Model		
Est.	-0.154	0.096
S.E.	0.198	0.173
Est./S.E.	-0.779	0.554
p-value	0.44	0.58

Table 2.5

Monte Carlo Covariate Mean Comparison

Entropy	0.4				0.6			
Sample Size	250		1,000		250		1,000	
	Pc	Mlcm	Pc	Mlcm	Pc	Mlcm	Pc	Mlcm
Class on Covariate = 0								
Value	0.353	0.063	0.312	0.001	0.498	0.001	0.52	0.002
S.D.	0.532	0.994	0.411	1.02	0.676	0.967	0.698	0.981
P-value of average stat	0.552	0.083	0.576	0.972	0.48	0.976	0.471	0.964
Average p-value	0.669	0.285	0.674	0.282	0.596	0.287	0.603	0.281
Number rep p < 0.05	1	21	0	26	3	21	1	18
Proportion rep p < 0.05	0.002	0.042	0	0.052	0.006	0.042	0.002	0.036
Class on Covariate = 0.5								
Value	1.72	2.96	6.39	17.8	5.25	8.01	18.9	32.9
S.D.	1.55	1.05	2.94	1.04	3.31	1.03	6.62	1.02
P-value of average stat	0.19	0.085	0.011	0	0.022	0.005	0	0
Average p-value	0.31	0.136	0.03	0.004	0.081	0.038	0	0
Number rep p < 0.05	98	195	410	493	302	380	499	500
Proportion rep p < 0.05	0.196	0.39	0.82	0.986	0.604	0.76	0.998	1

Note. Pc-Pseudo-class chi-square; Mlcm – Most likely class membership squared *t*-test.

Entropy	0.8				1.0			
Sample Size	250		1,000		250		1,000	
	Chi	Ti	Chi	Ti	Chi	Ti	Chi	Ti
Class on Covariate = 0								
Value	0.724	0.006	0.704	0.003	0.902	0.006	0.896	0.003
S.D.	1	0.993	0.982	0.981	0.994	0.995	0.953	0.961
P-value of average stat	0.395	0.936	0.401	0.956	0.382	0.941	0.395	0.954
Average p-value	0.544	0.282	0.551	0.282	0.516	0.281	0.524	0.286
Number rep p < 0.05	13	23	12	20	18	18	14	17
Proportion rep p < 0.05	0.026	0.046	0.024	0.04	0.036	0.036	0.028	0.034
Class on Covariate = 0.5								
Value	9.52	11.4	35.6	45.1	3.8	15.4	58.4	60.1
S.D.	5.37	1.04	10.2	1.04	16.2	1.06	16.2	1.01
P-value of average stat	0.002	0.001	0.007	0	0	0	0	0
Average p-value	0.027	0.017	0	0	0.007	0.006	0	0
Number rep p < 0.05	437	453	500	500	473	486	500	500
Proportion rep p < 0.05	0.874	0.906	1	1	0.946	0.972	1	1

50

Table 2.6

Monte Carlo Covariate Regression Results: Logistic Regression Coefficient= 0

N	250	1,000	250	1000
Entropy	0.4	0.4	0.6	0.6
Most Likely Class Membership as Outcome				
Est.	0.003	0.002	0.004	0.003
S.D.	0.143	0.066	0.126	0.063
S.E.	0.141	0.065	0.129	0.064
MSE	0.02	0.004	0.016	0.004
Coverage	0.954	0.938	0.954	0.955
Est./S.E.	0.021	0.031	0.031	0.047
Prop Est/ SE > 1.96	0.046	0.062	0.046	0.045
Class Probability Weighted Regression				
Est.	0.0004	0.003	0.005	0.001
S.D.	0.105	0.043	0.106	0.052
S.E.	0.1	0.042	0.106	0.052
MSE	0.011	0.002	0.011	0.003
Coverage	0.946	0.932	0.956	0.95
Est./S.E.	0.004	0.071	0.047	0.019
Prop Est/ SE > 1.96	0.054	0.068	0.044	0.05
Pseudo-Class Regression				
Est.	0.0004	0.003	0.005	0.001
S.D.	0.105	0.043	0.106	0.052
S.E.	0.167	0.081	0.149	0.074
MSE	0.011	0.002	0.011	0.003
Coverage	0.999	0.999	0.994	0.998
Est./S.E.	0.002	0.037	0.034	0.014
Prop Est/ SE > 1.96	0	0	0.006	0.002
Covariates in Model				
Est.	0.01	0.007	0.008	0.003
S.D.	0.283	0.106	0.164	0.079
S.E.	0.301	0.107	0.172	0.079
MSE	0.079	0.011	0.027	0.006
Coverage	0.952	0.948	0.964	0.952
Est./S.E.	0.033	0.065	0.047	0.038
Prop Est/ SE > 1.96	0.048	0.052	0.036	0.048

	250	1000	250	1000
N	250	1000	250	1000
Entropy	0.8	0.8	1.0	1.0
Most Likely Class Membership as Outcome				
Est.	0.011	0.003	0.01	0.004
S.D.	0.128	0.063	0.128	0.061
S.E.	0.128	0.064	0.128	0.063
MSE	0.017	0.004	0.017	0.004
Coverage	0.954	0.95	0.958	0.958
Est./S.E.	0.086	0.047	0.076	0.0635
Prop Est/ SE > 1.96	0.046	0.05	0.042	0.042
Class Probability Weighted Regression				
Est.	0.008	0.002	0.01	0.004
S.D.	0.118	0.057	0.128	0.0612
S.E.	0.117	0.058	0.128	0.0634
MSE	0.014	0.003	0.017	0.004
Coverage	0.95	0.95	0.958	0.958
Est./S.E.	0.068	0.034	0.076	0.0635
Prop Est/ SE > 1.96	0.05	0.05	0.042	0.042
Pseudo-Class Regression				
Est.	0.008	0.002	0.01	0.004
S.D.	0.118	0.057	0.128	0.0612
S.E.	0.139	0.069	0.128	0.0634
MSE	0.014	0.003	0.017	0.004
Coverage	0.982	0.978	0.958	0.958
Est./S.E.	0.058	0.029	0.076	0.0635
Prop Est/ SE > 1.96	0.018	0.022	0.042	0.042
Covariates in Model				
Est.	0.009	0.002	0.01	0.004
S.D.	0.144	0.069	0.128	0.061
S.E.	0.144	0.07	0.128	0.063
MSE	0.021	0.005	0.017	0.004
Coverage	0.952	0.952	0.958	0.958
Est./S.E.	0.063	0.029	0.076	0.0635
Prop Est/ SE > 1.96	0.048	0.048	0.042	0.042

Table 2.7

Monte Carlo Covariate Regression Results: Logistic Regression Coefficient = 0.5

N	250	1,000	250	1000
Entropy	0.4	0.4	0.6	0.6
Most Likely Class Membership as Outcome				
Est.	0.245	0.274	0.371	0.371
S.D.	0.148	0.069	0.138	0.068
S.E.	0.143	0.066	0.135	0.067
MSE	0.088	0.056	0.036	0.021
Coverage	0.54	0.078	0.812	0.496
Est./S.E.	1.71	4.15	2.74	5.53
Prop Est/ SE > 1.96	0.402	0.99	0.774	1
Class Probability Weighted Regression				
Est.	0.197	0.2	0.326	0.321
S.D.	0.106	0.045	0.112	0.055
S.E.	0.099	0.042	0.108	0.053
MSE	0.103	0.092	0.043	0.035
Coverage	0.156	0	0.616	0.09
Est./S.E.	1.99	4.76	3.02	6.06
Prop Est/ SE > 1.96	0.518	0.998	0.86	1
Pseudo-Class Regression				
Est.	0.197	0.2	0.326	0.321
S.D.	0.106	0.045	0.112	0.055
S.E.	0.176	0.083	0.156	0.077
MSE	0.103	0.093	0.043	0.035
Coverage	0.58	0.2	0.856	0.304
Est./S.E.	1.12	2.41	2.09	4.17
Prop Est/ SE > 1.96	0.186	0.81	0.578	0.998
Covariates in Model				
Est.	0.536	0.516	0.521	0.506
S.D.	0.316	0.122	0.19	0.09
S.E.	0.346	0.121	0.19	0.088
MSE	0.101	0.015	0.037	0.008
Coverage	0.97	0.954	0.96	0.94
Est./S.E.	1.55	4.26	2.74	5.75
Prop Est/ SE > 1.96	0.362	0.996	0.81	1

| N	250	1000	250	1000
Entropy	0.8	0.8	1.0	1.0
Most Likely Class Membership as Outcome				
Est.	0.445	0.436	0.518	0.506
S.D.	0.142	0.07	0.148	0.069
S.E.	0.137	0.067	0.139	0.069
MSE	0.023	0.009	0.022	0.005
Coverage	0.918	0.828	0.952	0.956
Est./S.E.	3.25	6.51	3.72	7.33
Prop Est/ SE > 1.96	0.906	1	0.976	1
Class Probability Weighted Regression				
Est.	0.418	0.409	0.518	0.506
S.D.	0.13	0.062	0.148	0.069
S.E.	0.123	0.06	0.139	0.069
MSE	0.023	0.012	0.022	0.005
Coverage	0.87	0.664	0.952	0.956
Est./S.E.	3.39	6.81	3.72	7.33
Prop Est/ SE > 1.96	0.932	1	0.976	1
Pseudo-Class Regression				
Est.	0.418	0.409	0.518	0.506
S.D.	0.131	0.062	0.148	0.069
S.E.	0.148	0.074	0.139	0.069
MSE	0.024	0.012	0.022	0.005
Coverage	0.934	0.792	0.952	0.956
Est./S.E.	2.82	5.53	3.72	7.33
Prop Est/ SE > 1.96	0.864	1	0.976	1
Covariates in Model				
Est.	0.519	0.506	0.518	0.506
S.D.	0.166	0.079	0.148	0.069
S.E.	0.159	0.077	0.139	0.069
MSE	0.028	0.006	0.022	0.005
Coverage	0.952	0.944	0.952	0.956
Est./S.E.	3.26	6.57	3.72	7.33
Prop Est/ SE > 1.96	0.926	1	0.976	1
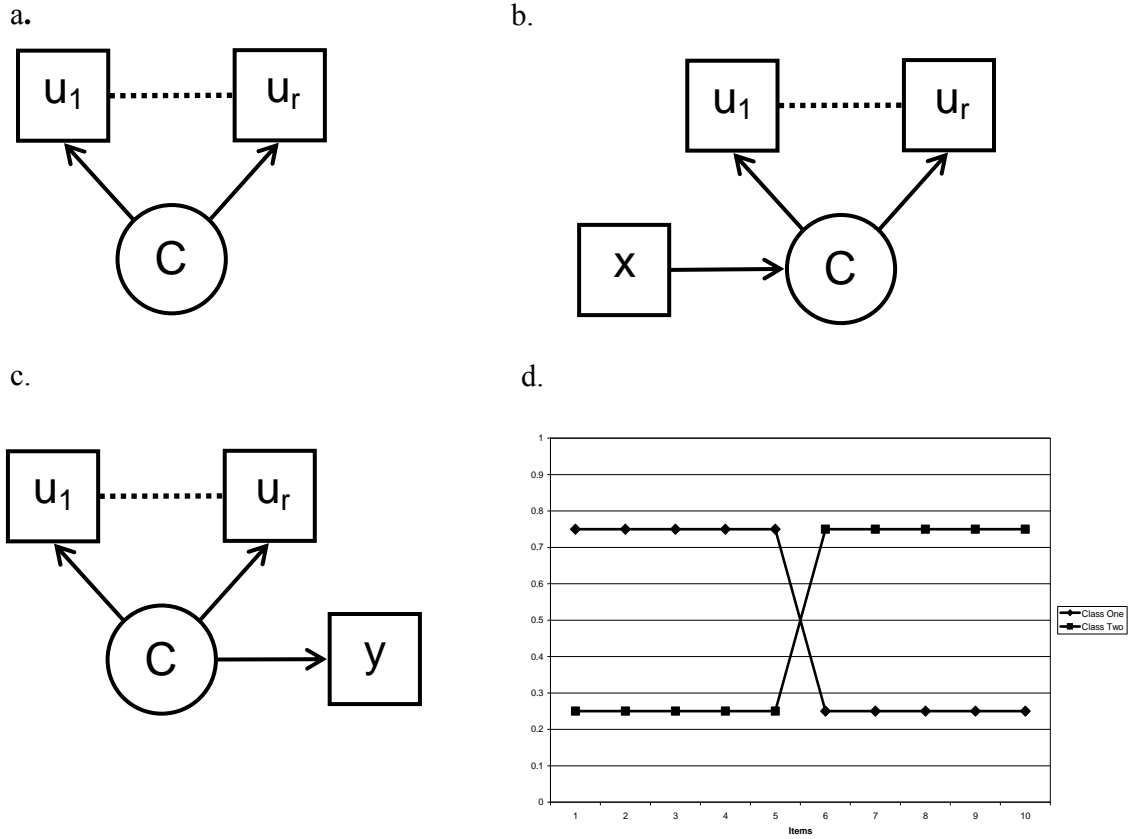
Figure 2.1. a. Latent class analysis model diagram. b. Latent class analysis with covariate model diagram. c. Latent class analysis with distal outcome model diagram. d. Crossed profile plot.
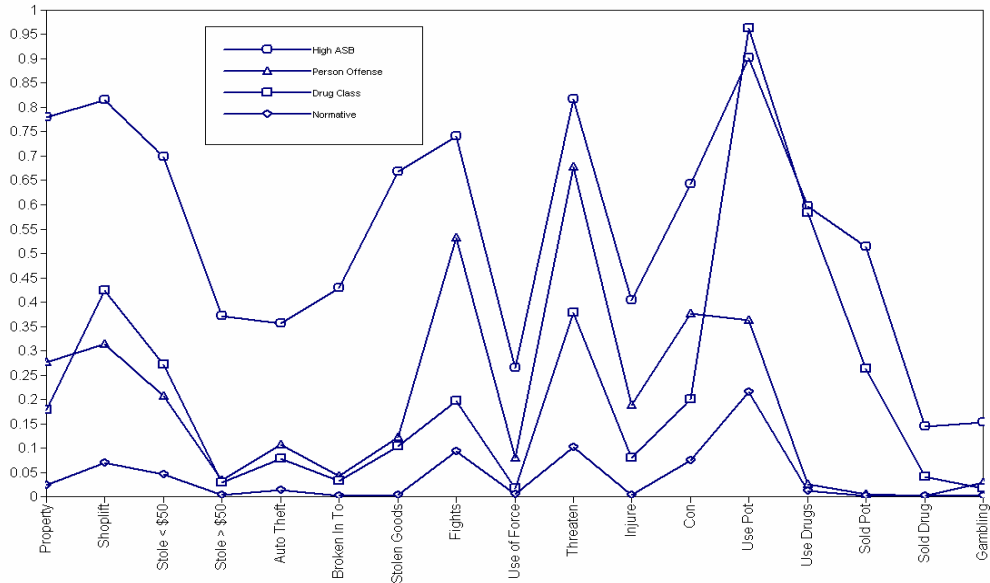
Figure 2.2. Antisocial behavior data profile plot.
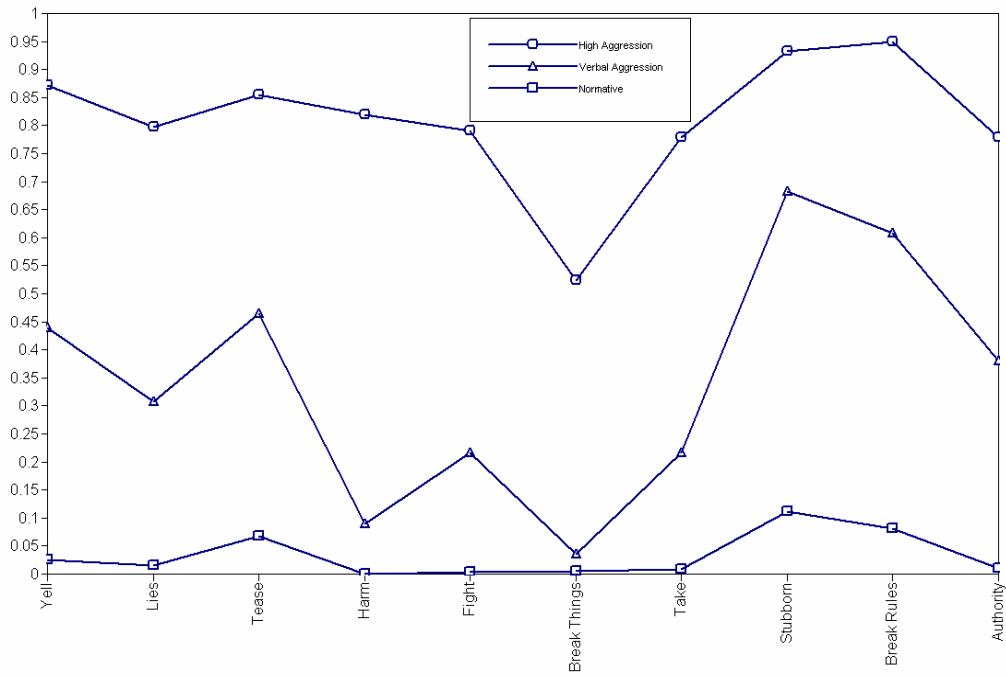
Figure 2.3. TOCA data profile plot.

Appendix A: Mplus Code for Chapter 2

Example Mplus code for the case in Table 2.7 where entropy is high,

and the sample size is large.

Input 1: Monte Carlo input to generate data under a latent class model with a covariate and analyze the generated data using the same model.

```
Title:    Monte Carlo Settings
          High entropy = 0.8, Sample size = 1,000,
          Class on covariate = 0.5
Montecarlo:
    Names are u1-u10 x; !Assigning names to generated variables
    Generate = u1-u10(1);
         ! Generating 10 categorical items with one threshold
    Categorical = u1-u10; ! Specifying which items are categorical
    Genclasses = c(2); ! Generating 2 classes named c
    Classes = c1(2); ! Analyzing data with 2 classes called c1
    Nobservations = 1000; !Sample size of each data set
    Seed = 86142; ! Specifying seed to be used for random draws
    Nrep = 500; !Number of data replications
    Repsave = ALL; !Specifies that all replications will be saved
    Save = sim*.dat;
         ! Naming the files to which the data will be saved. The
         ! asterisk is replaced by the replication number.
Analysis: Type = Mixture;
         !Specifies that a mixture model will be used
         Starts = 0; !Specifies the number of random starts
```

```
Model Population:

    !Specifying the model from which the data is generated

    %Overall%

    [x@0];  ! Setting covariate mean to zero

    x@1;  ! Setting covariate variance to zero

    [c#1*0];

    c#1 on x*0.5; !Regression of class on covariate

    %c#1% !Specifying the class specific item thresholds

    [u1$1*1 u2$1*1 u3$1*1 u4$1*1 u5$1*1 u6$1*-1 u7$1*-1 u8$1*-1

    u9$1*-1 u10$1*-1];

    %c#2%

    [u1$1*-1 u2$1*-1 u3$1*-1 u4$1*-1 u5$1*-1 u6$1*1 u7$1*1 u8$1*1

    u9$1*1 u10$1*1];

Model:  !Specifying the model by which the data is analyzed

    %Overall%

    [c1#1*0];

    c1#1 on x*0.5;

    %c1#1%

    [u1$1*1 u2$1*1 u3$1*1 u4$1*1 u5$1*1 u6$1*-1 u7$1*-1 u8$1*-1

    u9$1*-1 u10$1*-1];

    %c1#2%

    [u1$1*-1 u2$1*-1 u3$1*-1 u4$1*-1 u5$1*-1 u6$1*1 u7$1*1 u8$1*1

    u9$1*1 u10$1*1];

Output: Tech9;

    !Provides information about the convergence of each replication
```

Input 2: Analyzing data generated in Input 1 with a latent class model in order to be able to save class probabilities and most likely class membership.

```
Variable:

      Names are u1-u10 x c0;

      Usevariables are u1-u10;

            ! Specifies which variables are to be used in the analysis

      Categorical = u1-u10;

      Classes = c1(2);

      Auxiliary = x;

            ! Identifies auxiliary variables and includes them when

            ! saving data

Data: File is sim.dat;

Analysis:

      Type = Mixture;

      Starts = 0;

Model: %Overall%

      [c1#1*0];

      %c1#1%

      [u1$1*1 u2$1*1 u3$1*1 u4$1*1 u5$1*1 u6$1*-1 u7$1*-1 u8$1*-1

      u9$1*-1 u10$1*-1];

      %c1#2%

      [u1$1*-1 u2$1*-1 u3$1*-1 u4$1*-1 u5$1*-1 u6$1*1 u7$1*1 u8$1*1

      u9$1*1 u10$1*1];

Output: Tech9;

Savedata: File is simpp.dat; !Name of file to save data to

      Save = Cprob;

            !Saving class probabilities and membership information
```

Input 3: Taking most likely class membership saved from Input 2 and regressing class membership on the covariate x.

```
Variable:

     Names are u1-u10 x p1 p2 c0;

     Usevariables are c0 x;

     Nominal = c0; !Specify which dependent variable is nominal


Data: File is simpplist.dat;

          !Name of file that contains a list of names of the

          !generated data

     Type = Montecarlo;

          !Specifies that data are multiple data sets generated using

          !the Monte Carlo option

Analysis: Algorithm = Integration;

     !Specifies that numerical integration should be used

Model: c0#1 on x*0.5; ! Regressing class membership on the covariate
```

Input 4: Taking most likely class probabilities saved from Input 2, converting them to the logit scale and regressing them on the covariate x.

```
Variable:

     Names are u1-u10 x p1 p2 c0;

     Usevariables are x logit;

     !Uses logit variable created in Define statement below

Data: File is simpplist.dat;

     Type = Montecarlo;
```

```
Define:

    If(p1 lt .00001)then p1 =.00001;

    If(p1 gt .9999)then p1 = .9999;

    !Recoding p1 so that values of 0 and 1 will be able to be
converted to the logit scale

    logit = log(p1/(1-p1));

    !Creating a new variable called logit which converts p1 to logit

    scale

Model: logit on x*0.5; !Regressing logit on covariate
```

Input 5: Taking most likely class probabilities saved from Input 2 and using the

probabilities as weights in a regression.

```
Variable:

    Names are u1-u10 x p1 p2 c0;

    Usevariables are u1-u10 x p1 p2;

    Categorical = u1-u10;

    Classes = c1(2);

    Training = p1 p2 (probabilities);

        !Specifies which variables contain information about class

        !membership and that the information is the posterior

        !probabilities

Data: File is simpplist.dat;

    Type = Montecarlo;

Analysis: Type = Mixture;

    Starts = 0;

Model: %Overall%

    [c1#1*0];
```

```
 c1#1 on x*0.5;

%c1#1%

[u1$1*1 u2$1*1 u3$1*1 u4$1*1 u5$1*1 u6$1*-1 u7$1*-1 u8$1*-1

u9$1*-1 u10$1*-1];

%c1#2%

[u1$1*-1 u2$1*-1 u3$1*-1 u4$1*-1 u5$1*-1 u6$1*1 u7$1*1 u8$1*1

u9$1*1 u10$1*1];
```

Input 6: Taking data generated from Input 2 and analyzing using a latent class model with pseudo-class draws.

```
Variable:

     Names are u1-u10 x c0;

     Usevariables are u1-u10;

     Categorical = u1-u10;

     Classes = c1(2);

     Auxiliary = x(r);

          ! Specifies the auxiliary variable and asks for psuedo-
          !class regression

Data: File is simlist.dat;

Analysis: Type = Mixture;

     Starts = 0;

Model: %Overall%

     [c1#1*0];

     %c1#1%

     [u1$1*1 u2$1*1 u3$1*1 u4$1*1 u5$1*1 u6$1*-1 u7$1*-1 u8$1*-1

     u9$1*-1 u10$1*-1];

      %c1#2%
```

```
      [u1$1*-1 u2$1*-1 u3$1*-1 u4$1*-1 u5$1*-1 u6$1*1 u7$1*1 u8$1*1

      u9$1*1 u10$1*1];
```

Input 7: Taking data generated using Input 2 and analyzing with a latent class model with the covariates included.

```
Variable:

      Names are u1-u10 x c0;

      Usevariables are u1-u10 x;

      Categorical = u1-u10;

      Classes = c1(2);

Data: File is simlist.dat;

      Type = Montecarlo;

Analysis:

      Type = Mixture;

      Starts = 0;

Model: %Overall%

      [c1#1*0];

      c1#1 on x*0.5;

      %c1#1%

      [u1$1*1 u2$1*1 u3$1*1 u4$1*1 u5$1*1 u6$1*-1 u7$1*-1 u8$1*-1

      u9$1*-1 u10$1*-1];

      %c1#2%

      [u1$1*-1 u2$1*-1 u3$1*-1 u4$1*-1 u5$1*-1 u6$1*1 u7$1*1 u8$1*1

      u9$1*1 u10$1*1];
```

Input 8: Taking data saved in Input 2 and analyzing with a latent class model with pseudo class draws used to form Wald chi-square tests for mean comparisons.

```
Variable:

    Names are u1-u10 x c0;

    Usevariables = u1-u10;

    Categorical = u1-u10;

    Classes = c1(2);

    Auxiliary = x(e);

        ! Specifies the auxiliary variable and asks for psuedo-

        !class Wald chi-square test


Data: File = simlist.dat;

Analysis: Type = Mixture;

    Starts = 0;

Model: %Overall%

    %c1#1%

    [u1$1*1 u2$1*1 u3$1*1 u4$1*1 u5$1*1 u6$1*-1 u7$1*-1 u8$1*-1

    u9$1*-1 u10$1*-1];

    %c1#2%

    [u1$1*-1 u2$1*-1 u3$1*-1 u4$1*-1 u5$1*-1 u6$1*1 u7$1*1 u8$1*1

    u9$1*1 u10$1*1];
```

References

Asparouhouv, T. & Muthén, B.O. (2007). *Wald test of mean equality for potential latent class predictors in mixture modeling*. From: http://www.statmodel.com/download/MeanTest1.pdf

Bandeen- Roche, K., Miglioretti, D.L., Zeger, S.L., &Rathouz, P.J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92, 1375-1386.

Clogg, C.C. (1995). Latent class models: Recent developments and prospects for the future. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp.311-352). New York: Plenum.

Dayton, C.M., & Macready, G.B. (1998). Concomitant variable latent class analysis. *Journal of the American Statistical Association*, 83, 173-178.

Dolan, L., Kellam, S.G., Brown, C.H., Werthamer-Larsson, L., Rebok, G.W., Mayer, L.S., Laudolff, J., et al. (1993). The short-term impact of two classroom based preventive intervention trials on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology* 14, 317-345.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87, 476-486.

Hagenaars, J.A. (1993). *Loglinear models with latent variables*. London: Sage.

van der Heijden, P., Dessens, J., & Bockenholt, U. (1996). Estimating the concomitant-variable latent-class model with the EM algorithm. *Journal of Educational and Behavioral Statistical, 31(3), 215-229.*

Ialongo, L.N., Werthamer, S., Kellam, S.K., Brown, C.H., Wang, S., and Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression and antisocial behavior. *American Journal of Community Psychology,* 27, 599-641.

Kellam, S.G., Rebok, G.W., Ialongo, N., and Mayer, L.S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry*, 25, 359-382.

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

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

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

Muthén, L., & Muthén, B. (1998-2008). Mplus *User's Guide*. Fifth Edition. Los Angeles, CA: Muthén & Muthén.

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

Nagin, D.S., & Tremblay, R.E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods*, 6(1), 18-34.

Nylund, K.L. & Masyn, K. (2008). Covariates and Mixture Modeling: Results of a Simulation Study Exploring the Impact of Misspecified Covariate Effects. Paper presented at the annual meeting of the Society for Prevention Research.

Petras, H., & Masyn, K. (2009). General growth mixture analysis with antecedents and consequences of change. Forthcoming in Piquero, A., & Weisburd, D. (Eds.), *Handbook of Quantitative Criminology*.

Roeder K., Lynch, K.G., & Nagin D.S. (1999). Modeling uncertainty in latent class membership: a case study in criminology. *Journal of the American Statistical Association*, 94 (47), 766-7776.

Wang C.P., Brown, C.H., Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: Examining the impact of preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association*, 100 (3), 1054-1076.

Chapter Three: Models and strategies for factor mixture analysis:

Two examples concerning the structure underlying psychological disorders

In recent years there has been a debate in the psychological literature about whether the underlying structure of psychological disorders, such as conduct disorder and attention-deficit hyperactivity disorder, is categorical or dimensional. In the categorical view, psychological disorders are represented by diagnostic categories that indicate whether a person is affected or unaffected by a disorder and/or whether a person has a specific subtype. This has been the predominant view of psychopathology because it has the advantage of meeting clinical needs and the needs of reporting for health care planners and insurance companies (Muthén, 2006). Alternatively, psychological disorders are considered dimensional in nature and are represented as a continuous distribution, with each individual having some amount of the disorder. The advantage of the dimensional view is that each disorder can be represented as a quantitative score, or scores, which provides a more precise measure of functioning and more power for further statistical analyses than categorical outcomes (Muthén, 2006).

In the psychometric literature, each of these views has a counterpart. The categorical view can be represented by latent class analysis, which models unobserved heterogeneity in a sample through the use of categorical latent variables called latent classes. In this analysis, individuals are grouped into their most likely class based on their observed symptoms so that latent classes can then be interpreted as diagnostic categories or subtypes. The problem with latent class analysis, and the categorical approach to

psychopathology, is that the categories do not consider the range in severity and impairment within and across diagnostic classes.

The dimensional view of psychological disorders has its counterpart in factor analysis. Here, continuous latent variables, called factors, are used to model the correlations among the symptoms. Each of these factors represents an underlying dimension of the disorder. One drawback of this approach is that there is, generally, no easy way to classify individuals into groups, which, as stated earlier, is currently a clinical necessity and required by insurance companies and other reporting agencies.

One solution to the debate, proposed by Muthén (2006), is the factor mixture model (FMM). The FMM uses a hybrid of both categorical and continuous latent variables, which allows the underlying structure to be simultaneously categorical and dimensional. The structure is considered categorical because the model allows for the classification of individuals into diagnostic groups through the use of latent class variables. The structure is also considered dimensional because once individuals are classified into groups, the FMM allows for variation in the severity of the disorder through the use of continuous latent variables. This approach is useful because it does not have the limitations of the two conventional representations of psychopathology.

While the studies that have introduced the factor mixture model to the psychological literature have explained the conceptualization of the FMM, the use of the FMM is still not prevalent. One reason for this is that even though the concept of the FMM has been explained, there is little research about how these models should be

69

applied in practice and, once a well fitting model is obtained, how it should be interpreted. This paper seeks to remedy these shortcomings.

This paper aims to explain the different variations of the FMM including the various steps in building a FMM, and how to decide between a FMM and alternative models. The FMM will be explored in detail by studying two real data examples: conduct disorder and attention-deficit hyperactivity disorder.

The first section of this paper explains the latent class, factor analytic, and factor mixture models in more technical detail. The second section focuses on the model building process and on how to compare among the different types of models. The third section presents the real data examples. The final section discusses the utility and feasibility of the factor mixture model in practical settings. All analyses in this paper were carried out using Mplus V5.1 (Muthén & Muthén, 1998-2008). In order to elucidate the factor mixture model, sample inputs for each model variation are available in the Appendix B.

Background

Latent Class Analysis

The latent class analysis (LCA) model, introduced by Lazarsfeld and Henry (1968), is used to identify subgroups, or classes, of a study population. A diagram of an example of a latent class analysis model is shown in Figure 3.1. There are two major concepts depicted in Figure 3.1, the observed outcomes or items that define the class and latent class itself. These can be seen in Figure 3.1 as u_1-u_r, and C, respectively. The boxes, u_1 to u_r, represent the observed response items or outcomes. The outcomes in a

LCA model can be categorical, continuous, count, censored, or nominal, though this paper will specifically focus on dichotomous, categorical items. The circle with the letter C in the middle is the unordered, categorical latent class variable with K classes. The arrows pointing from the latent class variable to the boxes above indicate that those items are measuring the latent class variable. This means that class membership is based on the observed response pattern of items. An important assumption, called the conditional or local independence assumption, implies that the correlation among the observed outcomes is explained by the latent class variable. Because of this, there is no residual correlation between the items.

For a LCA model with categorical outcomes, there are two types of model parameters: conditional item probabilities and class probabilities. The conditional item probabilities are specific to a given class and provide information about the probability that an individual in that class will endorse a specific item. The class probabilities specify the relative size of each class, or the proportion of the population that is in a particular class.

Consider a LCA model with r observed binary items, u, and a categorical latent variable C with K classes ($C = k$; $k = 1, 2, \ldots,$ K). The marginal item probability for item $u_j = 1$ is

$$P(u_j = 1) = \sum_{k=1}^{K} P(C = k)P(u_j = 1 \mid C = k). \tag{1}$$

Assuming conditional independence, the joint probability of all the r observed items is

$$P(u_1, u_2, \ldots, u_r) = \sum_{k=1}^{K} P(C = k)P(u_1 \mid C = k)P(u_2 \mid C = k)\ldots P(u_r \mid C = k). \tag{2}$$

71

A product of LCA is the estimated class probabilities for each individual, called posterior probabilities, which are analogous to factor scores in a factor analysis (Muthén, 2001). These are estimates of

$$P(C = k \mid u_1, u_2, ..., u_r) = \frac{P(C = k)P(u_1 \mid C = k)P(u_2 \mid C = k)...P(u_r \mid C = k)}{P(u_1, u_2, ..., u_r)}. \quad (3)$$

Note that each individual is allowed fractional class membership and may have non-zero values for many classes. It is from these probabilities that class membership is assigned. An individual is assigned to be a member of a class based on their highest probability of being in a given class, even though an individual may have several classes to which they are a partial member. Once assigned to a class, an individual is assumed to be a part of that class 100%.

One potential shortcoming to using LCA to represent psychological disorders is the assumption of conditional independence. This implies that all the symptom items are statistically independent within a latent class. One way to satisfy the conditional independence assumption is to add additional latent classes until the assumption is met. There are two possible scenarios that can occur as a result of adding more classes. The first is that the added latent classes are reflecting only residual correlations between a small number of items, beyond what the latent classes explain. In this sense, the added latent classes are not genuine subgroups. FMM is a good alternative for this situation because the addition of a factor would account for the residual item correlations and reduce the need for a larger number of classes. Using the FMM in this scenario may lead to the factor not having all loadings be non-zero. For instance, if there is an extra

correlation, beyond what the classes explain, between only two items, then only those two item's loadings would be non-zero. This highlights the flexibility of the factor approach. The second scenario is that the additional classes are genuine subgroups. FMM might be useful in reducing the number of classes in this scenario if the classes are modeling differences in severity. In this case the factor will account for this variation in severity.

Another implication of the conditional independence assumption is that for each item in a class, individuals have the same conditional item probability. This implies that all individuals in a class will have the same probability of symptom presence. This is problematic because there is often within-class heterogeneity in the form of variation in severity which violates the conditional independence assumption (Muthén & Asparouhov, 2006). This implies that individuals do not have the same conditional item probability in a class. The FMM avoids this through the use of the latent factor. Within-class, individuals have different factor values which allows for people to vary in their conditional item probabilities.

Class enumeration in mixture models. A well-known problem with mixture models is that they tend to converge on local solutions, rather than a global solution (McLachlan & Peel, 2000). This is problematic because a local solution may differ dramatically from the global solution. The use of multiple sets of randomly generated starting values is one way to avoid converging on a local solution (McLachlan & Peel, 2000). Observing the same maximum likelihood for multiple sets of starting values makes it more likely that a global solution has been found. One of the advantages of

73

using Mplus V5.1 (Muthén & Muthén, 1998-2008) is that it has a random starts feature that generates a different number of random starting value sets to facilitate in finding the global solution.

Another known issue with mixture models is the difficulty in deciding on the appropriate number of classes, called class enumeration. Despite the various suggestions, there is no commonly accepted methodology on how to compare models with differing numbers of classes (Nylund, et al. 2007). When considering a plausible set of models it is wise to utilize a combination of statistical and substantive model checking (Muthén, 2003), which is the strategy that is used in this paper. There are two common ways to compare models. One is statistical tests of model fit, such as the likelihood ratio test of neighboring models. Another is using statistical indicators, such as information criteria. The standard chi-square difference test (likelihood ratio test, LRT) cannot be used with mixture models because the more restrictive model has its probability parameter for one class at zero, which is on the border of the admissible parameter space (Machlan & Peel, 2000). Two alternatives to using the chi-square difference test are the Vuong-Lo-Mendell-Rubin test (LMR-LRT; Lo, Mendell, & Rubin, 2001) and the parametric bootstrapped LRT (BLRT; McLachlan, 1987). Both of these tests approximate the difference in likelihood values between a k- and (k-1)-class model. Each of these tests provides a p-value, which indicates whether a (k-1)-class model can be rejected in favor of a k-class model. When comparing the p-values of a plausible set of models, one is looking for the first time the p-value is non-significant, or greater than 0.05.

Another way of choosing the number of classes is to use information criteria. Information criteria, such as the Bayesian Information Criterion (BIC; Schwarz, 1978), are based on the best maximum likelihood of a model and different criteria apply different penalties for the number of parameters and sample size. Because of the different penalties that are applied by each information criterion, they will often point to differing numbers of classes. For mixture models with different numbers of classes, the model with lowest value of an information criterion is deemed the best fitting model. Nylund et al. (2007), which considers the performance of statistics and indices described here, recommend using the BIC and LMR to narrow the number of models under consideration and then include the BLRT for a smaller set of model comparisons due to the computational demands of the BLRT.

Besides looking at the statistical comparisons of the models, it is important to also understand whether the models make substantive sense and to examine the residuals. One way to explore the interpretation of the models is to examine the mean class profile for different models. It is also worthwhile to examine the class size and proportion since an over-extraction of classes can result in small and non-distinct classes (Masyn et al., in press). Another way to explore the interpretation of a FMM is through a pattern response table that displays the item response pattern, its frequency, the total number of items endorsed, class probabilities, and factor scores broken down by each class. The residuals, or the difference between the observed and estimated values, provide further insight into whether the model fits the data well. A large number of significant residuals indicates that the model does not fit the data well.

Factor Analysis

In factor analysis (FA), the goal is to investigate common content among the items (Lubke & Muthén, 2005) by seeing if items group together on continuous latent variables called factors. An example of a FA model can be seen in Figure 3.2. As with the latent class model, there is a circle, this time with the letter F in the middle, representing the factor. But, the circle in a FA is not an unordered, categorical latent variable like in LCA, but instead, a continuous one. Because the latent variable is continuous, there is no assumption of different subpopulations of individuals, like in LCA. Instead, it is assumed that all individuals in the sample are from the same homogenous population and that differences among individuals arise because of differences on the factor. Similar to LCA, there are arrows emanating from the factor that point towards the items. This indicates that the factors are measured by those combinations of items that people tend to endorse together. For example, in the context of ADHD diagnosis, one factor might be measured by those symptom items relating Inattentiveness and another might be measured by those symptom items relating to Hyperactivity.

The factor analytic model with dichotomous items y takes on the following form for individual i:

$$y^*_i = \tau + \Lambda\,\eta_i + \varepsilon_i,$$

$$\eta_i = \alpha + \zeta_i,$$

and

$$y_{ij} = \begin{cases} 1 & \text{if } y*_{ij} > 0 \\ 0 & otherwise. \end{cases}$$

(4)

where y_i are individual i's responses, which is a p vector of observed outcomes; $y*_i$ is individual i's latent response vector; j is one of the items in a p vector; τ is a p vector of measurement thresholds; Λ is a $p \times m$ matrix of factor loadings, where m is the number of factors; η_i is a m vector of factor scores; ε is p vector of residuals; α is a m vector of the intercepts of the factors; ζ_i is a p vector of residuals which is assumed to be normally distributed with mean zero and variance Ψ.

An important piece of information that results from a FA is the estimated factor scores, η_i. The factor scores are estimated from equation 4, after the analysis has been conducted. The factor scores of all the individuals in a sample can be thought of as forming an approximation to the sample distribution of the factor. The factor score of an individual can be used as indication of where that individual is located on the factor, relative to the other individuals in the sample.

One disadvantage of FA is that it does not give a model-based classification of individuals, which can be problematic if this is the goal of the analysis. One option would be to plot the factor scores and see if there are any natural cut points or thresholds for classification. Natural cut points, however, can be hard to find in practice (Muthén, 2006), especially if there is more than one factor.

One advantage of using FA is that it is a widely used technique and so, there are many guidelines for how to do factor analysis in practice. For a technical description of FA with continuous items see Joreskog (1969) and for a more applied description and a guideline for how to do FA in practice see Brown (2006). For information on FA with categorical items see Bartholomew and Knott (1999) and Muthén (1989).

Factor Mixture Model

The factor mixture model (FMM) uses a hybrid of latent class and factor analysis (Muthén, 2008) and an example of a FMM can be seen in Figure 3.3. The figure shows that the FMM is a combination of LCA and FA because there is both a latent class variable, the circle with the C in the middle, and a latent continuous factor, the circle with the F in the middle. Through the use of these two types of latent variables the FMM simultaneously provides both a categorical and dimensional view of psychology. The latent class variable allows for the classification of individuals into groups while the factor models the severity of the disorder. The factor also yields quantitative scores of the disorder in the form of the factor scores.

Figure 3.3 shows an added flexibility of the FMM. The solid lines starting from the latent variables and ending at the symptom items indicate that the latent variables are measured by those symptoms. Additionally, there are dashed lines coming from the latent class variable and ending on the line that goes from the factor to the symptom items. This indicates that the factor structure of the model can be different in each class.

The formulation of the FMM is similar to what was seen in the FA model except that all the parameters have the potential to be different across the classes. A factor

mixture model for $k = 1, 2, \ldots, K$ latent classes with dichotomous items can be specified as follows:

$$y^*_{ik} = \tau_k + \Lambda_k \eta_{ik} + \varepsilon_{ik},$$

$$\eta_{ik} = \alpha_k + \zeta_{ik},$$

where,

$$\zeta_{ik} \sim N(0, \Psi_k)$$

and

$$y_{ij} = \begin{cases} 1 & \text{if } y^*_{ij} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$(5)$$

All of the parameters in the equation above have the same interpretation as the FA model presented in equation 4, but now they can be class specific as indicated by the subscript k. By allowing all of the parameters to have the potential to be class-specific, the model becomes more flexible and allows for several variations that differ in the amount of measurement invariance. These variations will be discussed in more detail in the next section.

The FMM has distinct advantages over LCA and FA that may make it a better model to represent psychological disorders. Unlike LCA, the FMM does not need a conditional independence assumption given latent class. This suggests that individual differences in the severity of a disorder can be explored within a given class. These variations within a class are represented by the continuous factors and can be quantified by using the factor score for each individual (Lubke & Muthén, 2005; Muthén et al.,

2006). One disadvantage to FA is that there is no model-based classification of individuals. In a FMM, classification is obtained through the latent class variable.

The latent class and factor analytic models can be seen as special cases of the FMM. LCA is a special case of FMM in which the factor covariance matrix is zero. The factor analytic model can be thought of as a FMM with one latent class to which every individual in the sample belongs.

While there has been little research on how to do the FMM in practice, several authors have made contributions toward the development of this model. One of the first articles is by Yung (1997), which considers a FMM where all of the parameters were class-varying. McLachlan and Peel (2000) discuss a FMM where the factor structure is exploratory rather than confirmatory. Muthén (2006) and Muthén and Asparouhov (2006) consider a FMM using dichotomous outcomes. Muthén (2008) provides an overview of the different types factor mixture models and breaks them down into four branches depending on the amount of measurement invariance present and whether the factor in each model is parametric or not. This paper shows how the FMM variations presented map onto the branches presented in the Muthén (2008) article. Muthén (2008) and Masyn et al. (in press) discuss longitudinal extensions of FMM. Masyn et al. also position the FMM in their dimensional-categorical spectrum which organizes latent variable models based on the types and combinations of latent variables used. Kim and Muthén (2008) explain how to do two-part FMM to be able to model data with strong floor effects.

There have been additional applications of FMMs in genetics and IRT analysis. FMM has been applied to genetic contexts by McLachlan, Do, and Ambroise (2004) to

80

micro-array gene expression data and by Muthén, Asparouhov, and Rebollo (2006) to a

twin heritability study. Mislevy and Verhelst (1990), Mislevy and Wilson (1996), Wilson

(1998), and Boeck, Wilson, and Acton (2005) applied FMM to IRT studies.

FMM: Model Variations

In this paper, several different variations of the FMM and their interpretations are

presented. Model presentation will start with the more restrictive models, since the

interpretations tend to be simpler, and move to less restrictive models, where the

interpretation can be more complex. The formulas in this section have no link function

relating y* to y in order to save space. The link function is the same as the one presented

in equation 5.

The first model variation is the latent class factor analytic (LCFA) model. It is

referred to as FMM-1 and can be formulated as:

$$y^*_{ik} = \tau + \Lambda \eta_{ik} + \varepsilon_{ik},$$

$$\eta_{ik} = \alpha_k , \qquad\qquad (6)$$

In this model, the only parameter that changes across classes is the factor mean, which is

indicated by the subscript k on α. The item thresholds and factor loadings are held

invariant across classes, suggesting that the disorder is being measured the same way

across all classes. The factor covariance matrix, Ψ, is fixed at zero, indicating that there

is no severity in the disorder. Figure 3.4 shows a diagram of this model on the left and on

the right, an example plot of the factor means versus the frequency of each class. On the

right, bars are used to mark the factor means instead of a distribution because the factors

have no variance, which suggests that this model has a non-parametric factor distribution.

The height of the bars in the plot represents the percentage of the sample in that class. This plot suggests that the only difference between individuals arises due to having different amounts of the disorder and that there are only a set number of amounts that individuals can have, which are the number of classes. In the model diagram on the left, the latent class variable points to the factor, and not directly to the items, indicating that class membership is based on each individual's location on the factor. Because this variation has measurement invariance and non-parametric factor distributions, FMM-1 is representative of Branch 2 in Muthén (2008).

The second model variation is the FMM-2, also called a mixture factor analysis. The FMM-2 is similar to the FMM-1 except that instead of setting the factor variances and covariances to zero, they are now freely estimated in each of the classes. The formulation for the second model is:

$$y^*_{ik} = \tau + \Lambda \eta_{ik} + \varepsilon_{ik},$$

$$\eta_{ik} = \alpha_k + \zeta_{ik} ,$$

where,

$$\zeta_{ik} \sim N(0, \Psi_k). \tag{7}$$

Figure 3.5, similar to Figure 3.4, shows a model diagram on the left and a plot of the factor means versus the frequency on the right. On the right, the classes are now represented as distributions, rather than bars as in FMM-1, because the factors have variance. This indicates that there are many possible amounts of the disorder an individual can have. The model diagram on the left is almost the same as for the FMM-1 except that there is now an arrow pointing into the latent factor which indicates a residual

so that the within-class factor variance is now being modeled. One variation of the FMM-2 is to allow the factor variances to be freely estimated, but held equal across classes. Because this model variation has measurement invariance and parametric factor distributions, FMM-2 is representative of Branch 1 in Muthén (2008).

In the third model variation, FMM-3, the factor loadings and factor covariance matrix are invariant, while the item thresholds are allowed to change across classes. The formulation for the FMM-3 is:

$$ \boldsymbol{y^*}_{ik} = \boldsymbol{\tau_k} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_{ik} + \boldsymbol{\varepsilon}_{ik}\,, $$

$$ \boldsymbol{\eta}_{ik} = \boldsymbol{\zeta}_i\,, $$

where,

$$ \boldsymbol{\zeta}_{ik} \sim \mathrm{N}(\boldsymbol{0},\,\boldsymbol{\Psi}). \tag{8} $$

Also, the factor mean is set to zero for identification purposes and so does not appear in the equations above. The factor mean is set to zero in the remaining model variations. Figure 3.6, which shows the model diagram for the FMM-3, is different than the previous two model diagrams because the arrows starting from the latent class variable now point to the items rather than the factor. This shows that the item thresholds are changing across classes, which implies that the classes are now based on the responses to those items rather than the factor, as was seen in the previous two models. In FMM-3, and the remaining FMM variations, some or all the measurement parameters are non-invariant and the factor distribution is parametric, so FMM-3 through -5 fit into Branch 3 in the Muthén (2008) paper.

The fourth model variation, FMM-4, is the same as the third model except that the factor covariance matrix is now allowed to change across the classes. Having different factor variation across the classes implies that there are different amounts of severity within each class. For example, a class that can be thought of as "Unaffected" might have less variance because individuals are showing none to almost no symptoms while a class that is "Affected" might have more variation because individuals might have a greater range of symptoms. The formulation for the fourth model is:

$$\boldsymbol{y^*}_{ik} = \boldsymbol{\tau_k} + \boldsymbol{\varLambda}\,\boldsymbol{\eta}_{ik} + \boldsymbol{\varepsilon}_{ik}\,,$$

$$\boldsymbol{\eta}_{ik} = \boldsymbol{\zeta}_{ik}\,,$$

where,

$$\boldsymbol{\zeta}_{ik} \sim \mathrm{N}(\boldsymbol{0},\,\boldsymbol{\varPsi}_k). \tag{9}$$

The fifth model, FMM-5, is the least restrictive model in terms of invariance of the parameters. The item thresholds, factor loadings, and factor covariance matrix are all allowed to vary across classes. The formulation for the fifth model is:

$$\boldsymbol{y^*}_{ik} = \boldsymbol{\tau_k} + \boldsymbol{\varLambda_k}\,\boldsymbol{\eta}_{ik} + \boldsymbol{\varepsilon}_{ik}\,,$$

$$\boldsymbol{\eta}_{ik} = \boldsymbol{\zeta}_{ik}\,,$$

where,

$$\boldsymbol{\zeta}_{ik} \sim \mathrm{N}(\boldsymbol{0},\,\boldsymbol{\varPsi}_k). \tag{10}$$

FMM and Measurement Invariance (MI)

The FMM variations presented above differ in terms of how much measurement invariance (MI) is present. The amount of MI can have important implications for how a

FMM is interpreted, specifically whether or not the same factors are being measured in each latent class. There are many definitions of MI, but this paper will use Meredith's (1993) definition of strong factorial invariance which requires equality of factor loadings, λ, and thresholds, τ, across classes because strong factorial invariance is often considered to be sufficient for the comparison of subpopulations (Little, 1997; Widaman & Reise, 1997). For a further discussion about different definitions of MI and how they apply to FMMs see Lubke and Muthén (2005).

In strong factorial invariance, the factor loadings and thresholds of an item must be held equal across latent classes. If the factor loadings, which are slopes in the regression of the observed variable on the factors, are non-invariant in the latent classes, then a unit increase in the factor score does not result in the same increase of the dependent variable in the different classes. This differential increase can suggest one of two things. The first is that if there are many items with large differences in their factor loadings between classes, then each class has a different underlying factor. The second is if there are relatively few items with differences in their factor loadings between classes, then those items with differential loadings function differently in each class. If the item thresholds, which are negative intercepts in the regression of observed variables on the factor, are not equal across latent classes then one latent class scores consistently higher or lower than the others, independent of scores on the factor. This suggests that observed differences between classes are not entirely due to differences in the factor.

In FMM-1 and -2, the factor loadings and item thresholds are invariant across classes, suggesting that the interpretation of the factor remains the same in each class or,

put another way, the factor retains the same meaning at both the low and high values in the population. In these two variations, the latent classes are used to model the non-normality of the factor in the population. For FMM-1, Figure 3.4, which plots the factor distribution on the right side of the figure, shows that the factor is categorical with each latent class representing a category. For FMM-2, Figure 3.5, which also plots the factor distribution, shows that the factor is continuous, but non-normal. The factor in a FMM-2 is a mixture of normal distributions located at different points on the factor distribution, and each normal distribution is a latent class.

In FMM-3 and -4, the item thresholds are non-invariant, while in FMM-5 both the item thresholds and the factor loadings are non-invariant. In these model variations, because there is a violation of strong factorial invariance, the same factor does not apply to the whole population. Instead, there is a different factor, with a different interpretation, in each latent class. The implication of this is that there are several populations, represented by the latent classes, each with their own distributions.

In practice, FMM-1 and -2 often do not fit real data well because the specification of invariant factor loadings and thresholds are likely to be too restrictive for certain items. In these two models, the factor loadings influence both the item mean, which changes over classes for an item j as a function of $\tau_j + \lambda_j^* \alpha_k$, and the item correlation between item j and item l within class which is $\lambda_j^* \psi^* \lambda_l$. Item means are not likely to change over classes as function of $\tau_j + \lambda_j^* \alpha_k$, but instead change as a function of τ_{jk}. It is unlikely that α_k is the only parameter that creates changes across classes in item means for these two models. Furthermore, it is because the factor loading has a connection between both

the item mean changes over classes and the item correlation within-class that FMM-1 and -2 seldom fit real data well.

Even though FMM-1 and -2 may not fit real data well, partial MI is possible. Items which violate measurement invariance may be pinpointed using one of three possible methods. To demonstrate these methods, a simulated data set with six continuous items, y1-y6, is used. The data generating model was specified so that one item's thresholds, y5, were non-invariant across classes. The remaining thresholds for the other five items were invariant. The data were specified so that a FMM-2 was the data generating model, but with the factor variance structure equal across classes. One way to possibly identify a non-invariant item is to compare the estimated and observed item means. If an item was non-invariant then the observed and estimated means would be different from one another. In Table 3.1, which shows the observed and estimated means for each item from an FMM-2 in the top two rows of the table, the observed and estimated means are equivalent for each item. This is because mixture models, like FMMs, do a good job of fitting first and second order moments (means and variances), making the comparison of observed and estimated means an unsuitable method for detecting threshold non-invariance. A second way to identify these items is to examine the within-class residuals for each item's mean from a FMM-1 or -2. An item that is non-invariant may have a larger residual than items that are invariant. In Table 3.1, all of the items have small residuals which are close to zero, including y5. This suggests that examining the within-class residuals for each item's mean is not the best method for identifying threshold non-invariance.

A third way to detect non-invariance is to conduct a series of analyses in which one item's thresholds are held invariant and the thresholds for the rest of the items freely vary across the classes[1]. Using the model constraint feature in Mplus, a test of whether the difference between the items whose thresholds are allowed to vary is significantly different from zero can be conducted (See Appendix B for example Mplus code). If the test is significant, those item thresholds should be considered non-invariant and if the test is non-significant, the item thresholds should be invariant. Once this is performed with each item's thresholds being held equal across classes, a tally can be made of the number of times each item's threshold difference test is significant. The more times an item has a significant difference test, the more likely it is that item should have freely varying item thresholds. For the simulated data, y5 had 5 significant difference tests while the remaining items had zero significant difference tests suggesting that item y5 should be non-invariant across classes and the other items should be invariant. Based on these results, a partial MI model with y5 non-invariant and the remaining items invariant should be compared to the original model, FMM-2. This strategy will be demonstrated in the attention-deficit hyperactivity disorder example.

FMM: Factor Measurement Structure

Similar to FA, FMMs can have two possible factor measurement structures: confirmatory and exploratory. As in confirmatory FA, confirmatory FMMs utilize substantive theory to define factors that are measured only by specific symptom items, with the factor having no influence on the remaining items. This results in what is

---

[1] We thank Tihomir Asparouhov for suggesting this approach.

referred to in the FA literature as a simple measurement structure. Confirmatory FMM has the advantage of encouraging researchers to formalize their measurement hypothesis by having latent variables that are grounded in substantive theory, which can often lead to more parsimonious models (Asparouhov & Muthén, 2009). There are, however, disadvantages to using a simple measurement structure including fixing small cross-loadings to zero when a simple structure is used. Ignoring these cross-loadings may force a researcher to specify a more parsimonious model for the data than is necessary, which can lead to poor model fit (Asparouhov & Muthén, 2009).

An alternative to a confirmatory factor structure is an exploratory structure, where the number of factors and the relationships between the factors and the symptom items do not need to be specified a priori. Asparouhov & Muthén (2009) suggest using an exploratory structure in structural equation models when limited measurement knowledge is available or a more complex measurement structure is needed. An exploratory structure can be applied to a FMM using the Type = Mixture EFA option in Mplus V5.1 (Muthén & Muthén, 1998-2008). Using this option not only provides rotated factor solutions, but also standard errors for all rotated parameters and overall tests of model fit.

In this paper, all of the FMMs presented used a confirmatory structure that was guided by a priori substantive reasoning as to the number of factors and relationships between the factors and the symptom items. Given that a mixture was found to be a reasonable fit to the data in the examples, an exploratory structure was also fit to the data to examine if the presence of mixtures would change the number of factors or the factor structure since the presence of mixtures may call for a simpler factor measurement

89

structure. In the examples presented in this paper, the presence of mixtures did not change the factor structure or the number of factors. In order to demonstrate the use of the exploratory structure in FMMs, however, it was applied to the second example.

FMM Model Building and Comparison

Because FMMs are relatively new to the literature there is a dearth of writing about the steps to build a FMM, but general guidelines for how do them are offered in several papers (Muthén, 2006; Muthén & Asparouhov, 2006; Muthén, Asparouhov, & Rebollo, 2006). In each of these articles, the authors analyze their data using LCA, FA, and FMMs and then compare the fit among these three types. None of these articles provide instruction on how to construct the FMMs. Specifically, no information is given on how to decide on the number of classes and factors or how to pick among the different variations of the FMM.

A suggested strategy is to first fit LCA models with increasing number of classes and FA models with increasing number of factors. The next step is to fit a FMM with two classes and one factor, and in subsequent models increase the number of classes. After this, the number of factors would be increased to two, and the classes would once again be increased in subsequent models. This pattern of increasing the number of classes and then the number of factors would be repeated. The question arises: at what point does one stop increasing the number of classes and factors? In several papers we have reviewed, the strategy to decide on the number of classes and factors has been to take the best fitting number of classes from a LCA and the best fitting number of factors from a FA and combine them to form a FMM. After this is complete, no further investigation is

90

conducted to see if there is a need for a smaller number of classes and factors. This practice is highly discouraged because having both factors and classes in the model may reduce the need for a greater number of classes and factors. For example, if a LCA solution has many classes that are parallel to one another, which is thought to model severity, then adding one factor will likely reduce the need for the same number of classes because the factor will now model the severity. Instead, it is suggested that the number of classes from the best fitting LCA model and the number of factors from the best fitting FA model be the end point of combinations of classes and factors in model building. The strategy described above should be applied to all of the FMM variations. Once this is complete, the best FMM model for the data will be selected. Each of the examples will apply the strategy outlined above for model building.

Once a best fitting FMM has been selected, it is important to compare this model to the best fitting LCA and FA models to see if a more parsimonious solution can provide a better fit and explanation of the data. This is similar to the "best" candidate approach outlined for the Dimensional-Categorical spectrum in Masyn et al. (in press). Similar to deciding on the number of classes in a mixture analysis, the LLRT cannot be used for comparing among the different model types. Model comparison will be based on information criteria, with the lowest value indicating the best fitting model type, and substantive interpretation.

Examples

In the following section, two examples are explored to illustrate the model building and comparison strategies described above and to aid in understanding the

substantive interpretation of the FMM. Each example will begin with a brief description of the data, followed by model building and comparison of latent class, factor analytic, and factor mixture models.

Example One: Conduct Disorder

The first example examines the underlying structure of conduct disorder (CD) in interview data from a population-based sample of Finnish twins modestly enriched for familial alcoholism risk. The study, called FinnTwin12 (FT12; Rose et al., 2004), is longitudinal, but data for this illustration are from the first follow-up at age 14. In total, there were 1786 children in the study with about an equal number of boys and girls. The analysis focused on interview reports of thirteen dichotomous symptom items measuring CD that are from the Child Semi-Structured Assessment for the Genetics of Alcoholism, Adolescent version (C-SSAGA-A). The C-SSAGA-A is a polydiagnostic instrument that was developed by the Collaborative Study on the Genetics of Alcoholism (COGA; Kuperman et al., 2001). Because data from co-twins cannot be considered independent observations, the relatedness of the twins was handled by correcting the standard errors using the Type = Complex option in Mplus. The Type = Complex option use a sandwich estimator to obtain standard errors which take into account the non-independence of the twins.

In order to obtain a diagnosis of CD, three or more symptoms must be present. In this sample, 56% of children reported no symptom presence while 12% of the sample reported having three or more symptoms present and, therefore, met criteria for a CD diagnosis. Symptom presence ranged from 0.1% (Has ever been suspended from school)

to 25% of the sample (Is physically cruel to animals). For a more extensive description of the interviewed sample and analysis of CD in these 14 year-old Finnish twins see Rose et al. (2004).

According to Table 3.2, which displays model fit and comparison statistics for this example, the LCA model with the lowest BIC was the two-class model, followed by the three-class model. The first time the LMR p-value was non-significant, or greater than 0.05, was for the four-class model, which suggested that the three-class model was the best fitting model.

Since, the BIC and LMR p-value have indicated that the two- and three-class models were potential candidates for the "best" LCA model, the BLRT p-value and substantive interpretation of these two solutions were examined to help decide between them. The zero p-value of the BLRT for the three-class model indicated that a three-class model should be chosen over a two-class model. Figure 3.7 shows the item profile plot for the two-class solution. Class One, 78% of the sample, has low symptom presence as indicated by low item endorsement probabilities. Because of the large size of the class and low symptom presence, this can be considered the "unaffected" or asymptomatic class. Class Two, 22% of the sample, is defined by the items relating to suspension, lying, and truancy; with the other items have relatively low probabilities of symptom endorsement, 0.30 or lower.

Figure 3.8 shows the item profile plot for the three-class solution. As in two-class solution, there is one class that has low item endorsement across all items. This class is slightly smaller than the asymptomatic class from the two-class solution, 69%, because

some of the individuals were absorbed into the two remaining classes. Supporting this idea are the lower endorsement probabilities of the suspension and truancy symptom items in the unaffected class. The class that was defined by the suspension, lying and truancy items in the two-class solution has now been split into two classes in the three-class solution. The first of these classes is similar to the one in the two-class solution with high probabilities of endorsing the suspension and lying items, but the probability of endorsing truancy item has drastically decreased. The probability of endorsing other symptoms items is relatively low in this class. Potentially, some individuals in this class may meet criteria for a CD diagnosis. The second of these classes, 5% of the sample, still has high probabilities of suspension and lying, but also has high probabilities of running away, property destruction, forcing someone to participate in sexual activities, and starting fights. This class can be thought of as those individuals that are likely to meet criteria for CD. Given that there is a substantive interpretation and reasoning for each of the three classes and the BLRT p-value points toward the three-class model, it was selected as the "best" LCA model for this data.

For the FA models, an exploratory FA was first conducted, then based on the results, a series of confirmatory FA models were explored. The results of the exploratory FA suggested two possible factor solutions. The first is a one-factor solution in which the factor is defined by all thirteen CD items. The second is a two-factor solution with one factor being defined by the symptom items relating to expulsion from school, cruelty to animals, and use of a weapon when fighting and the second factor being defined by the remaining items. Potential solutions with more than two factors had the additional factors

being solely defined by one symptom item which lead to the solution being deemed not appropriate. When comparing the fit results of the one-and two-factor confirmatory models in Table 3.2, the log-likelihood is identical, but the one-factor model has one less parameter than the two-factor model, making it more parsimonious. This means that the BIC is lower for the one-factor model than the two-factor model. The factor solution for the two-factor model had several items on the second factor that did not load highly suggesting that there was no need for an additional factor. Since the one-factor model is more parsimonious, has a lower BIC value, and there are low loadings on the second factor in the two-factor solution, the one-factor solution was chosen as the "best" FA model for the data.

The bottom part of Table 3.2 presents results for the FMMs. Models with one factor and one to three classes were fit to the data, based on the results of FA and LCA. In Table 3.2, the FMM with the lowest BIC value is the two-class, one-factor FMM-2, with a value of 9284. But, for this model the LMR p-value is greater than 0.05, which indicates that a one-class, one-factor model should be chosen over this model. The next lowest BIC value in the table is 9374, which is 90 points higher than the lowest BIC value. This BIC value occurs for two models, with additional models having BIC values that are not much higher. For these FMMs, the LMR p-value is less than 0.05, which indicated that the two-class FMM should be chosen over the one-class model. Because there was no agreement of which model was the "best" FMM based on the BIC and the LMR p-value, the BLRT p-value was examined in the four models with the lowest BICs. In these four FMMs, the BLRT p-value was lower than 0.05 indicating that a two-class

FMM should be chosen over a one-class solution. Based on these results, the two-class, one-factor FMM-2 is chosen as the "best" candidate FMM. It has the lowest BIC and the substantive interpretation of the model made the most sense.

In Figure 3.9, which shows the item profile plot of the two-class, one-factor FMM-2 at the mean of the factor, there are two classes. The first and largest class, 57% of the sample, is the lowest line on the profile plot, with almost all items having low probabilities of symptom endorsement. Individuals in this class are likely to never have any symptoms of CD. While this class is similar in its interpretation to the lowest class in the three-class LCA, the percentage of individuals in the low class differs between the two solutions suggesting that the latent classes from a LCA and a FMM are not the same. Also, the probabilities of item endorsement are lower, almost zero, in the FMM and the class proportion in the FMM is similar to the proportion of individuals in the sample that do not endorse any items, which suggests that the low class in the FMM solution may be a true asymptomatic class. The second class, 43% of the sample, has slightly elevated probabilities of endorsing the symptom items related to suspension, lying, and truancy, but these probabilities are still small. This class is similar to the second class in the three-class LCA solution, but with smaller item endorsement probabilities and a larger class proportion. Notice that there are bands around the second class showing what the item endorsement probabilities are if an individual is one standard deviation away from the factor mean. The first class does not have bands around the item profile because the factor variance in this class was non-significant. The bands around the factor mean of the second class give an idea of the variability within each class. For example, for the

symptom item relating to suspension there is a great range of endorsement (0.20 to 0.70) once an individual is one standard deviation away from the factor mean. But, for the symptom item relating to expulsion, there is little variation in this item's probability of endorsement as evidence by the bands being very close to the factor mean for that item.

Comparing the factor loadings of the FMM and FA solutions, shown in Table 3.3, suggests that while the factors are similar in interpretation, they are not equivalent. Table 3.3, presents the unstandardized and standardized factor loadings. The factor loadings were standardized to take into account differences in the factor variance across models in order to be able to compare the loadings. The FMM-2 had class invariant factor loadings, hence the unstandardized factor loadings are the same in each class. But, the factor variances are non-invariant in a FMM-2, so the standardized factor loadings are also different in each class. Comparing the loadings from each class in the FMM-2 to the loadings from the FA solution, the loadings in class two are lower than the loadings in the FA solution while the class one loadings are higher. This suggests that the factor, which is considered the severity, has more influence in class one, the asymptomatic class. The higher loadings in class one also indicate that the with-in class correlation among the items is higher in class one. Note also that the items relating to suspension from school, fire starting, and truancy have the highest loadings in both solutions suggesting that these items are important in both solutions.

Now that the "best" candidate model has been selected for each of the three model types, LCA, FA, and FMA, a comparison can be made among these "best" candidate models to see which one best describes the data. One way to decide between the models

is to examine the BIC for the lowest value. For this data, the two-class, one-factor FMM-2 has the lowest BIC. Another way to decide between the models is to examine what each model implies about the underlying structure of CD in this population. The three-class LCA model implies that the underlying structure of CD is categorical with three categories, which are the three-classes in the solution.

The one-factor FA solution implies that the underlying distribution of CD is continuous and normally distributed. The two-class, one-factor FMM-2 also implies that there is one underlying continuous factor of CD because of the class invariant item thresholds and factor loadings. But, unlike the one-factor FA, the distribution of the factor in FMM-2 is non-normal. The distribution of the factor in the FMM-2, shown in Figure 3.10, is a bi-modal distribution with a small peak at about 0 and a taller peak at about 6.5. It is these peaks and the distributions around them that form the mixtures in the FMM-2 solution. The first latent class is the small peak at the bottom of the plot with the wide distribution around it. The second class is the normal distribution further up in the plot. By having two latent classes, the FMM can model the non-normality of the factor more precisely than a FA can. Based on the results of the BIC and comparing what each model implies about the underlying structure of CD, the "best" model for this data is the two-class, one-factor FMM-2 because it models the non-normality of CD better than the other models.

Example Two: Attention-Deficit Hyperactivity Disorder (ADHD)

The second data set originated from the Los Angeles ADHD study (Smalley et al., 2000). Unlike the previous example, which was a population sample, the LA ADHD data

contain only children who have been diagnosed as being affected with ADHD. In total,

there were 994 individuals in the sample, with about seventy percent of the sample being

male. Ages in the sample ranged from 4 to 18 with a mean age of 10.47 ($SD = 3.18$).

Many of the children in the study had an affected sibling who also participated; the

relatedness of the siblings was handled by correcting the standard errors using the Type =

Complex option in Mplus, as described in the previous example.

In this sample, ADHD was measured using the Schedule for Affective Disorders

and Schizophrenia for School-Age Children – Present and Lifetime version (K-SADS-

PL; Kaufman et al., 1997). This diagnostic instrument has 18 symptom items measuring

ADHD, with nine symptom items measuring Inattention and nine symptom items

measuring Hyperactivity\Impulsivity. Symptom presence ranged from 54.3% (Difficulty

playing quietly) to 96.2% (Difficulty sustaining attention on tasks\play). There were three

possible diagnoses of ADHD: predominantly Inattentive, predominantly Hyperactive, or

Combined (both Inattentive and Hyperactive). A diagnosis of predominantly Inattentive

is given to individuals having six or more symptoms from the Inattentive scale. A

diagnosis of predominantly Hyperactive is given to individuals having six or more

symptoms from the Hyperactive scale. An individual is considered to have Combined

ADHD if he/she met criteria for both the predominantly Inattentive and predominantly

Hyperactive. In this sample, 31% of individuals were predominantly Inattentive, 5% were

predominantly Hyperactive, and 64% were Combined.

The LCA of this example was conducted by fitting one- through five-class models

to the data. The results for this example can be seen in Table 3.4. The lowest value of the

BIC occurred for the four-class model. Because the LMR p-values for the all of the models were less than 0.05, that LMR was inconclusive as to which is the best LCA model.

Since, the BIC indicated that the four-class model was the best model and the LMR was inconclusive, the BLRT p-values were examined for the three- through five-class solutions. In Table 3.4, the first time the BLRT p-value was non-significant, or greater than 0.05, was for the five-class model which suggested that a four-class model should be chosen. Since both the BIC and BLRT have indicated that the four-class solution is the best LCA model for the data, it was chosen as the best LCA model. The four-class model profile plot is presented in Figure 3.11. Class one had high probabilities of item endorsement for the inattentive items and low probabilities of endorsement for the hyperactivity items, suggesting that this class can be considered a predominantly inattentive class. The second and smallest class, 15% of the sample, had a high probability of endorsement for the Hyperactivity and Inattention items with the exception of items that were related to working memory. The second class can be considered a hyperactive class with good working memory. The third class, which comprised 44% of the sample and had a high probability of endorsement for both the Inattention and Hyperactivity items, can be considered a combined class. The fourth class also had high probabilities of item endorsement for all the Inattention items, but had moderate to high endorsement probabilities for only those Hyperactivity items related to Impulsiveness.

For the FA models, an exploratory FA was first conducted, then based on the results, a series of confirmatory FA models were explored. The results of the exploratory

FA suggested two possible solutions, which are shown in Table 3.5. The first was a two-factor solution with one factor defined by all nine symptom items on the Inattention scale and the second factor defined by all nine items of the Hyperactivity scale. The second potential solution had four factors, with the first being defined by five items on the Inattention scale related to dreaminess or not paying attention. The second was defined by three symptom items on the Inattention scale relating to working memory (ex: Often loses things, Forgetful in daily activities). The third factor was defined by five symptom items on the Hyperactivity scale and the fourth factor was defined by the remaining four symptom items on the Hyperactivity scale that were related to Impulsivity (ex: Blurts out answers). The model statistics for the two- and four-factor confirmatory solutions are found in Table 3.4. When comparing the two models, the four-factor model had the lowest BIC by 172 points. A LLRT also suggested that the four-factor model fit the data better than a two-factor model, $\chi^2(5) = 206$, $p < .0001$. The results of the BIC and the LLRT both suggested that a four-factor model fit the data better than a two-factor model, so the four-factor model was chosen as the "best" candidate model for the FA model.

For the FMM, the results of the LCA and FA suggested that FMMs with a maximum of four-classes and four-factors should be fit to the data. Because a large number of factors are computationally intensive to fit with the already complicated FMMs, the fitting of models stopped at models with two factors. For this example, FMMs with both a confirmatory and exploratory measurement structure of the factors were explored. The confirmatory FMMs are at the top of Table 3.4 and the exploratory FMMs are at the bottom of the table.

In Table 3.4, when comparing the FMMs with confirmatory measurement structures to those with exploratory ones, the models with a confirmatory structure outperformed those with an exploratory one. This occurred for two possible reasons. The first is that in Mplus, models with an exploratory structure are assumed to have non-invariant factor loadings and item thresholds, and a non-invariant factor covariance structure, like the FMM-5. In this example, the best fitting model was a FMM-3 which has invariant factor loadings. By forcing a FMM-5 onto data which has FMM-3 as the best fitting model, there were a large number of unnecessary parameters which were non-significant, including a large number of non-significant cross loadings on the factors. The second reason why the exploratory FMMs did not perform as well in this example is that there was a clean, simple structure in these particular items.

The results of the FMMs in Table 3.4 suggested that a two-class, one-factor, FMM-4 fit the data the best. This model had the lowest BIC value by 100 points and the LMR p-value and BLRT p-value were both close to zero, which indicated that a two-class model should be chosen over a one-class model. While this model fit the data "best" according to statistical indicators, a closer inspection of the model revealed that it was not appropriate. Even though the factors in the two classes do not have the same interpretation due to class-varying threshold values, the factor loadings, shown in Table 3.5, for the inattention items are small and non-significant when compared with the hyperactivity item loadings. This suggested that the model was only measuring the hyperactivity construct and not inattention, which is a vital part of ADHD. Further inspection of all the one-factor FMMs revealed this limitation. Because of this, the next

102

"best" fitting model, the two-class, two-factor FMM-3, was explored. In this model, the first factor was defined by the inattention items only and second factor was defined by the hyperactivity items only. All of the item loadings were significant, see Table 3.5. In Figure 3.12, which shows the item profile plot for this solution, the first class, 33% of the sample, had a high probability of endorsing the inattention items, and high probability of endorsing the hyperactivity items related to impulsiveness. The second class, 67% of the sample, is similar to a Combination class because of the high probability of endorsing both hyperactivity and inattention items.

One way to help ease the interpretation of FMMs is to use a pattern response table where the symptom response pattern, symptom endorsement total, factor scores, and class probabilities are displayed by latent class. In Table 3.6, which displays the pattern response table for the ten most frequent response patterns in each class for the two-class, two-factor FMM-3, both classes have a high symptom endorsement totals for the inattention items as indicated by the column labeled Inat. total, but class one generally has lower hyperactivity symptom endorsement than class two as indicated by the Hyperactivity symptom endorsement column. This supports the interpretation that the second class is similar to a Combined ADHD class because individuals in the second class have high endorsement of both inattentive and hyperactive items.

In the two-class, two-factor FMM-3, the thresholds are non-invariant, but there is a possibility for partial invariance. This example demonstrates the strategy, which was outlined in a previous section, of how to explore if there is a need for partial threshold invariance. A series of analyses were performed, one for each of the 18 items, in which

one item's thresholds were constrained to be equal across classes and the rest free. A tally was then made of the number of times each item's difference test was significant. The items with the highest number of times their threshold difference test was significant are: Motor (9), Seated (8), Fidgets (8), Runs (8), Attention (7), and Instruct (7). Based on the results of the tally, a new model was run where the items mentioned above had freely estimated thresholds in each class and the remaining items had thresholds that were held invariant. Table 3.4 shows the model fits results for this model in the last line of the table. The BIC for this partially invariant model is only one point lower than the non-invariant two-class, two-factor FMM-3, but the partially invariant model has 12 fewer parameters making it a more parsimonious model.

Comparing the BIC values of the "best" fitting LCA, FA, and FMM located in Table 3.4, the four-factor FA has the lowest BIC value by 100 points. This example highlights that though the FMM is a powerful and flexible model, it does not always provide the best fit for the data. One possible reason the FMM did not fit as well as the four-factor FA model is that this is a sample of only affected children so there was not much variation in symptom endorsement because the individuals had to endorse a high number of symptoms to be included in this sample. This is contrary to what was seen in the previous example where the FMM was the "best" model for the data. Lubke et al. (2007) applied LCA, FA, and FMM to a population sample in order to investigate the structure of ADHD and found that a FMM fit the data the best.

104

Discussion

This paper builds upon the work of previous research on factor mixture models in order to explicate how to conduct a factor mixture analysis in practice. This paper discusses several different variations of the FMM and what each suggests about the structure underlying psychological disorders. Steps for building a FMM, including how to decide on the number of classes and factors, and how to compare the factor mixture to other appropriate models, are also discussed.

The two examples examining the underlying structure of CD and ADHD shed light on some issues and challenges that arise when conducting a factor mixture analysis. The first example, which investigated the underlying structure of conduct disorder in a sample of Finnish twins, showed how difficult it can be to decide between a FMM and other appropriate models, such as the FA and LCA models. There is often no clear "winner" based on statistical tests and indicators. The decision between different types of models must be made based on what each model implies about the underlying structure and whether that can be justified substantively. This example also highlights how FMM can be used to model a non-normal, continuous factor.

In the second example, which explored the underlying structure of ADHD in a sample of affected children and adolescents, several issues arose. First, the "best" fitting FMM according to statistical tests and indicators may not always make substantive sense, which suggests models should always be checked for their interpretability. Second, despite the flexibility of the FMM, it is not always the best fitting or most appropriate model for the data. Researchers need to consider whether the FMM is appropriate for

their data and what the FMM they are using implies about the underlying structure. Third, despite the advances in computational technology in recent years, more complicated FMMs with large numbers of classes and factors can still be difficult to process. In this example, the results of the FA suggested that FMMs should be fit with up to four factors. But once the FMMs were increased to four factors, the computational demands of the model were too heavy to run the models. This example also demonstrated how to investigate if there is a need for partial measurement invariance in a factor mixture model.

The two examples have shown that despite the work that has been done with the FMM, there are still avenues for further study. One area for further study is how and when to incorporate covariates or other auxiliary variables into a FMM. Lubke and Muthén (2005) describe different ways that a covariate can influence the latent classes and factors, but not at what point in the analysis covariates should be included or what method should be used to incorporate them. Another area for the further study is the performance of statistical tests and indicators in correctly identifying the correct combination of latent classes and factors. Nylund et al. (2007) have begun this work by investigating a single case with dichotomous items, but this paper only focused on the correct identification of the number of classes with no focus on identifying the number of factors. Lubke and Neale (2006, in press) have conducted simulation studies exploring the use of statistical indicators to identify the presence of measurement invariance in FMMs.

Even though there are some areas that need to be explored further, the FMM is still an important analytic tool for conceptualizing the structure underlying psychological disorders. Using the FMM, does not force the conceptualization to be either categorical or continuous, as previous notions have suggested. Instead the FMM allows for the underlying structure to be modeled as both so that one can simultaneously classify individuals into groups and have an estimate of the severity within those groups.

Table 3.1

Simulated MI Data: Observed vs. Estimated Mean and Within-Class Residuals

Item	y1	y2	y3	y4	y5	y6
Observed Mean	0.948	0.784	0.755	0.729	1.721	0.74
Estimated Mean	0.948	0.784	0.755	0.729	1.721	0.74
Class 1 Residual	0.00	-0.01	-0.02	-0.01	0.04	-0.01
Class 2 Residual	0.00	0.01	0.02	0.01	-0.04	0.02

Table 3.2

Conduct Disorder Example: Model Comparison Results (n =1758)

Model	Log-likelihood	Par.	BIC	LMR p-value	BLRT p-value
Latent Class Analysis					
One-Class	-5018	13	10062		
Two-Class	-4586	27	9374	0	0
Three-Class	-4540	41	9387	0.033	0
Four-Class	-4524	55	9460	0.095	
Five-Class	-4513	69	9543	0.132	
Factor Analysis					
One-Factor	-4561	26	9317		
Two-Factors	-4561	27	9324		
Factor Mixture Analysis					
2-Class, 1-Factor					
FMM-1	-4586	27	9374	0	0
FMM-2	-4537	28	9284	0.235	0.04
FMM-3	-4357	40	9374	0.006	0
FMM-4	-4536	41	9379	0.029	0
FMM-5	-4526	53	9449	1	
3-Class, 1-Factor					
FMM-1	-4540	41	9387	0.034	
FMM-2	*				
FMM-3	-4524	54	9452	0.527	
FMM-4	-4523	55	9481	0	
FMM-5	-4496*	81	9600	1	

Note. * Log-likelihood not replicated.

Table 3.3

Conduct Disorder Example: Unstandardized and Standardized Factor Loadings

Loadings	Unstandardized			Standardized		
Model	FA	FMM		FA	FMM	
Number of		2-Class, 1-Factor			2-Class, 1-Factor	
Factors	1 Factor	FMM-2		1 Factor	FMM-2	
		Class 1	Class 2		Class 1	Class 2
Items						
Suspend	0.90	0.90	0.90	1.81	4.41	1.07
Expelled	0.68	0.47	0.47	1.36	2.31	0.56
Stolen	0.49	0.36	0.36	0.99	1.75	0.42
Runaway	0.61	0.47	0.47	1.22	2.31	0.56
Lie	0.72	0.64	0.64	1.45	3.13	0.76
Fire	0.96	0.81	0.81	1.92	3.95	0.96
Truant	1.00	1.00	1.00	2.01	4.88	1.18
Property	0.59	0.44	0.44	1.18	2.17	0.53
Animal	0.67	0.42	0.42	1.34	2.06	0.50
Sexual	0.53	0.4	0.4	1.07	1.97	0.48
Fights	0.61	0.5	0.5	1.23	2.44	0.59
Weapon	0.36	0.26	0.26	0.72	1.25	0.30
Stolen	0.72	0.57	0.57	1.45	2.79	0.68
Factor Variance	4.03	23.8*	1.40			

Note. * Insignificant at the 5% level.

Table 3.4

ADHD Example: Model Comparison Results (n = 994)

Model	Log-likelihood	Par.	BIC	LMR p-value	BLRT p-value
Latent Class Analysis					
One-Class	-8378	18	16879		
Two-Class	-7534	37	15323	0	
Three-Class	-7320	56	15025	0	0
Four-Class	-7214	75	14944	0.016	0
Five-Class	-7155	94	14956	0.026	1
Factor Analysis					
One-Factor	-7448	36	15143		
Two-Factors	-7232	37	14719		
Four-Factors	-7129	42	14547		
Factor Mixture Analysis - Confirmatory					
2-Class, 1-Factor					
FMM-1	-7534	37	15323	0	
FMM-2	-7423	39	15115	0	
FMM-3	-7217	55	14813	0	
FMM-4	-7217	56	14547	0	0
FMM-5	-7161	73	14823	0	
3-Class, 1-Factor					
FMM-1	-7451	39	15170	0	
FMM-2	-7417	42	15123	0.002	
FMM-3	-7135	74	14780	0.083	
FMM-4	-7135	76	14792	0.164	
FMM-5	-7072	110	14901	0.461	
2-Class, 2-Factor					
FMM-1	-7534	38	15330	0	
FMM-2	-7191	45	14691	0.339	
FMM-3	-7131	56	14647	0.015	
FMM-4	-7124	59	14655	0.535	
FMM-5	-7090	75	14695	0.004	
3-Class, 2-Factor					
FMM-1	-7338	40	14952	0	
FMM-2	-7211	49	14659	0.004	
FMM-3	-7080	75	14675	1	
FMM-4	-7067	81	14691	0.015	
FMM-5	-7038	113	14852	0.011	
2-Class, 2-Factor Partial MI	-7171	44	14646		
Factor Mixture Analysis - Exploratory					
2-Class,1-Factor	-7165	73	14833		
2-Class, 2-Factor	-7044	107	14824		
2-Class, 3-Factor	-7001	139	14958		

2-Class, 4-Factor	**		
3-Class, 1-Factor	-7065	110	14886
3-Class, 2-Factor	-6979	161	15064
3-Class, 3-Factor	**		
3-Class, 4-Factor	**		

Note. ** Heywood case.

Table 3.5

ADHD Example: Factor Loadings

Model	FA					
Number of Factors	2		4			
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3	Factor 4
Items						
Attention	0.98		1			
Distract	0.65		0.68			
Mistake	0.53		0.36			
Listen	0.65		0.52			
Instruct	0.86		0.71			
Organization	0.73			0.34		
Avoid	0.41		0.27			
Loses	1			1		
Forget	0.95			0.46		
Seated		0.63			0.58	
Fidget		0.66			0.62	
Runs		0.91			0.84	
Quiet		0.46			0.38	
Blurt		0.53				0.80
Turn		0.59				0.78
Interrupt		0.59				1
Talk		0.38				0.51
Motor		1			1	

Model	FMM-4: 2 Class	FMM-3: 2 Class	
Number of Factors	1	2	
	Factor 1	Factor1	Factor2
Items			
Attention	-0.02*	0.97	
Distract	0.02*	0.62	
Mistake	-0.13*	0.40	
Listen	0.08*	0.52	
Instruct	-0.06*	0.75	
Organization	-0.12*	0.61	
Avoid	-0.01*	0.32	
Loses	-0.19*	1	
Forget	-0.22*	0.76	
Seated	0.61		0.61
Fidget	0.63		0.65
Runs	0.91		0.88
Quiet	0.45		0.48
Blurt	0.50		0.73
Turn	0.57		0.67
Interrupt	0.55		0.95
Talk	0.37		0.50
Motor	1		1

Note. * Insignificant at the 5% level.

Table 3.6

ADHD Example: Pattern Response Table for Two-Class, Two-Factor FMM-3.

Freq.	Inattention Items									
	Attn.	Distract	Mistake	Listen	Instruct	Organ.	Avoid	Loses	Forget	Inat. Total
Class One										
14	1	1	1	1	1	1	1	1	1	9
9	1	1	1	1	1	1	1	1	1	9
6	1	1	1	1	1	1	1	1	1	9
5	1	1	1	1	1	1	1	1	1	9
5	1	1	1	1	1	1	1	1	1	9
4	1	1	1	1	1	1	1	1	1	9
4	1	1	1	1	1	1	1	1	1	9
4	1	1	1	1	1	1	1	1	1	9
4	1	1	1	1	1	1	1	1	1	9
4	1	1	1	1	1	1	1	1	1	9
Class Two										
122	1	1	1	1	1	1	1	1	1	9
22	1	1	1	1	1	1	1	1	1	9
18	1	1	1	1	1	1	0	1	1	8
16	1	1	1	1	1	1	1	1	1	9
15	1	1	1	1	1	1	1	1	1	9
11	1	1	1	1	1	1	1	1	1	9
8	1	1	1	1	1	1	1	1	1	9
8	1	1	1	1	1	1	1	0	1	8
7	1	1	1	1	1	1	1	1	1	9
7	1	1	1	1	1	0	1	1	1	8

Freq.	Hyperactivity Items									
	Seated	Fidget	Runs	Quiet	Blurt	Turn	Inter.	Talk	Motor	Hyper. Total
Class One										
14	0	0	0	0	0	0	0	0	0	0
9	1	1	0	0	1	1	1	1	0	6
6	1	1	1	1	1	1	1	1	0	8
5	1	1	1	0	1	1	1	1	0	7
5	0	1	0	0	0	0	1	1	0	3
4	1	1	0	1	1	1	1	1	0	7
4	0	0	0	0	0	0	1	0	0	1
4	0	0	0	0	0	0	0	1	0	1
4	0	0	0	0	1	0	1	0	0	2
4	1	1	0	1	1	0	1	1	0	6
Class Two										
122	1	1	1	1	1	1	1	1	1	9
22	1	1	1	0	1	1	1	1	1	8
18	1	1	1	1	1	1	1	1	1	9
16	1	1	1	1	1	0	1	1	1	8
15	1	1	1	0	1	1	1	0	1	7
11	1	1	1	1	1	1	1	0	1	8
8	1	1	1	1	0	1	1	1	1	8
8	1	1	1	1	1	1	1	1	1	9
7	1	1	1	0	0	1	1	1	1	7
7	1	1	1	1	1	1	1	1	1	9

Freq.	Item Total	Factor Score		Class Probs.	
		Factor 1	Factor 2	Class1	Class 2
Class One					
14	9	1.22	-5.13	0.58	0.42
9	15	1.5	0.38	0.85	0.15
6	17	1.59	2.08	0.62	0.38
5	16	1.55	1.43	0.61	0.39
5	12	1.38	-1.91	0.77	0.23
4	16	1.53	0.94	0.85	0.15
4	10	1.3	-3.39	0.9	0.1
4	10	1.27	-4.1	0.77	0.23
4	11	1.35	-2.41	0.97	0.03
4	15	1.48	0.16	0.76	0.25
Class Two					
122	18	1.65	2.84	0.18	0.82
22	17	1.57	1.4	0.2	0.8
18	18	0.65	2.78	0.23	0.78
16	17	1.55	0.96	0.12	0.88
15	16	1.52	0.39	0.07	0.93
11	17	1.57	1.35	0.07	0.93
8	17	1.54	0.84	0.05	0.95
8	18	-0.92	2.7	0.03	0.97
7	16	1.5	0.02	0.05	0.95
7	17	-0.11	2.74	0.05	0.95
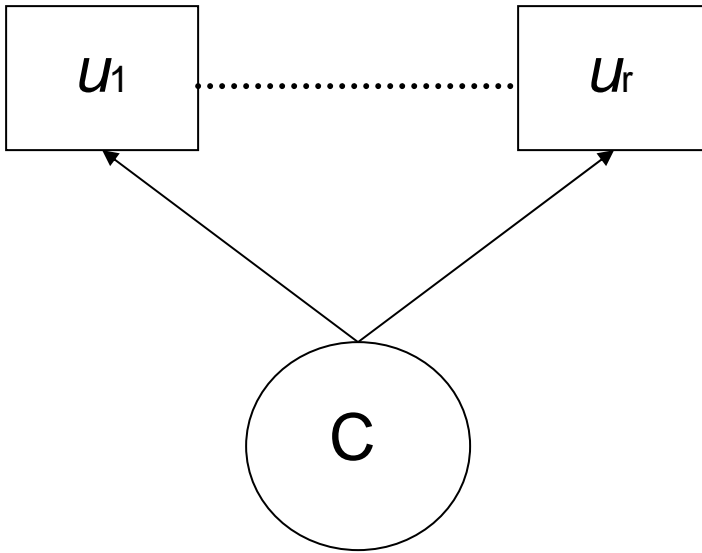
Figure 3.1. Latent Class Analysis Model Diagram.
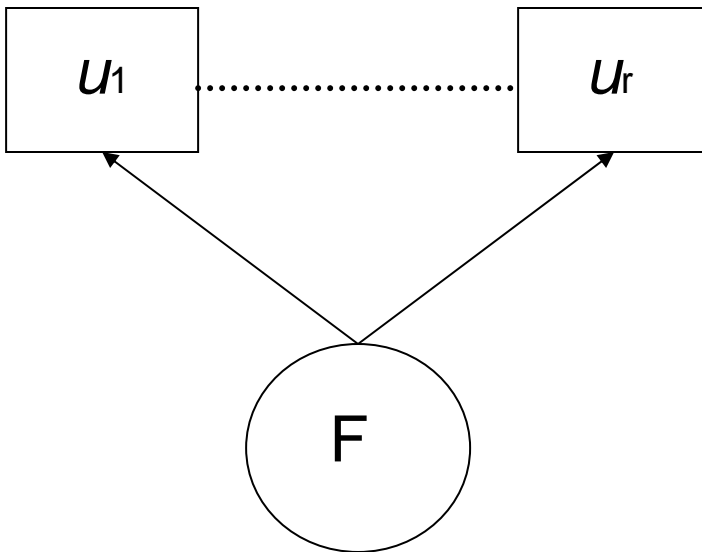
Figure 3.2. Factor Analysis Model Diagram.
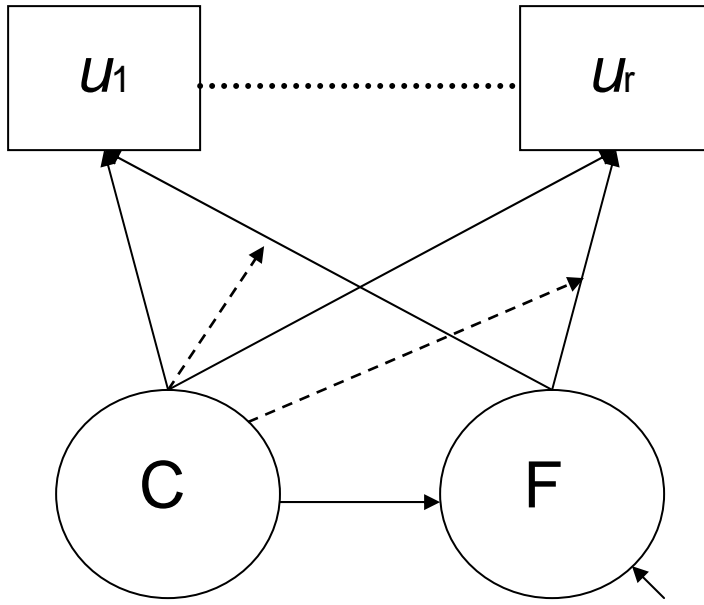
Figure 3.3. General Factor Mixture Model Diagram.
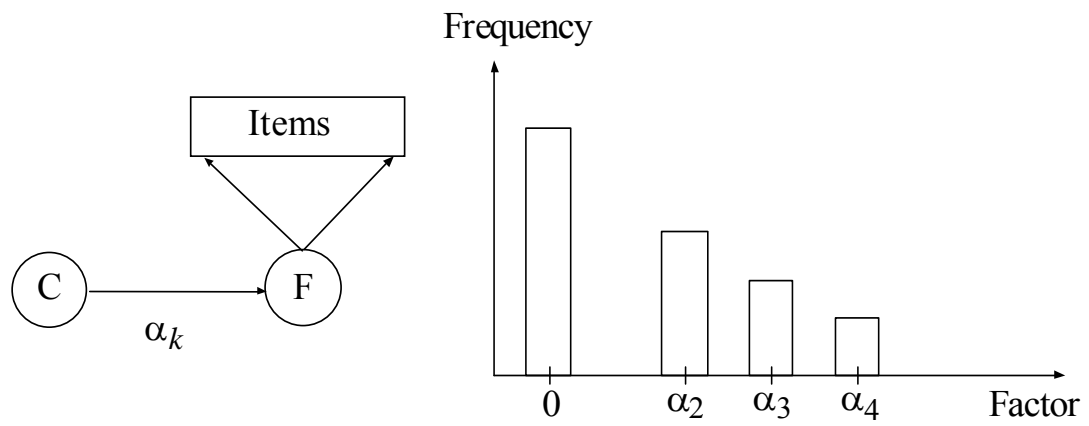
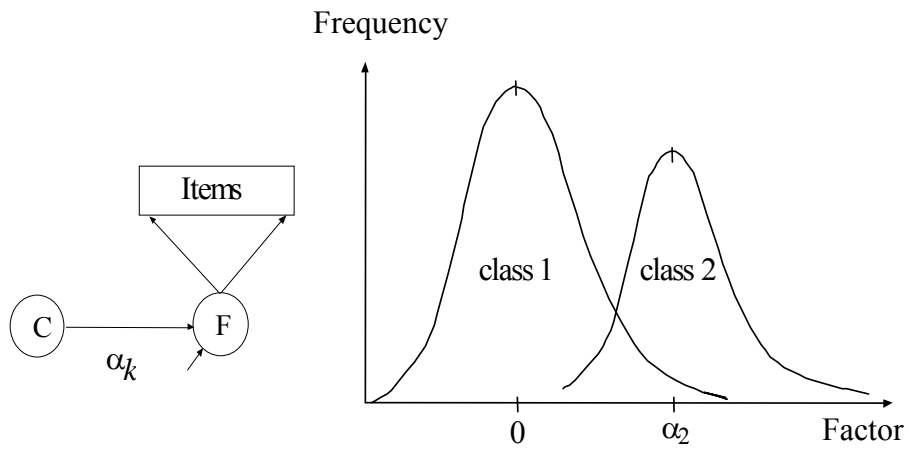*Figure 3.4.* Factor Mixture Model-1: Diagram and Factor Distribution Plot.

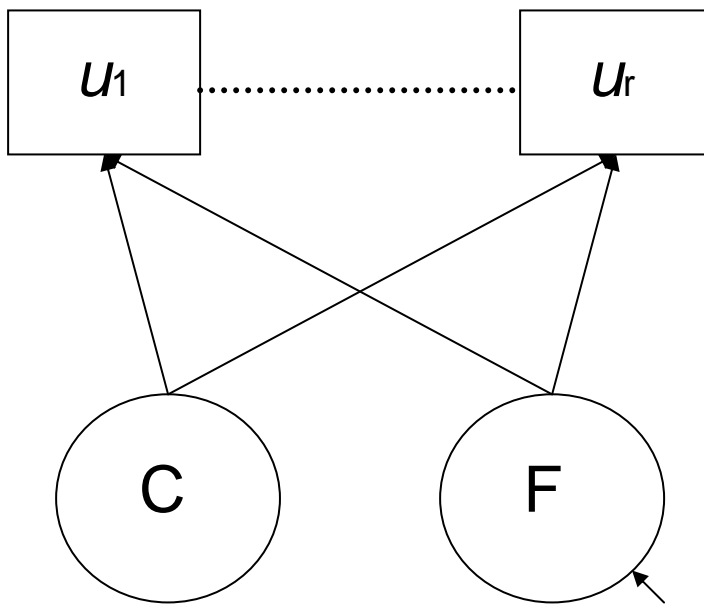*Figure 3.5.* Factor Mixture Model-2: Diagram and Factor Distribution Plot.

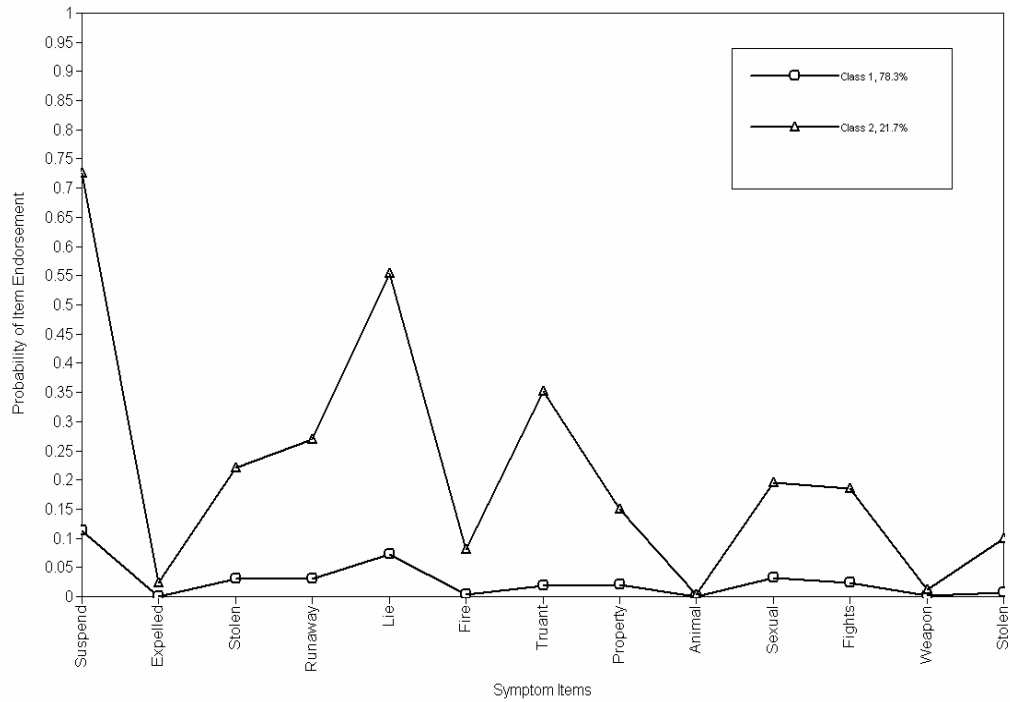*Figure 3.6.* Factor Mixture Model-3: Diagram.

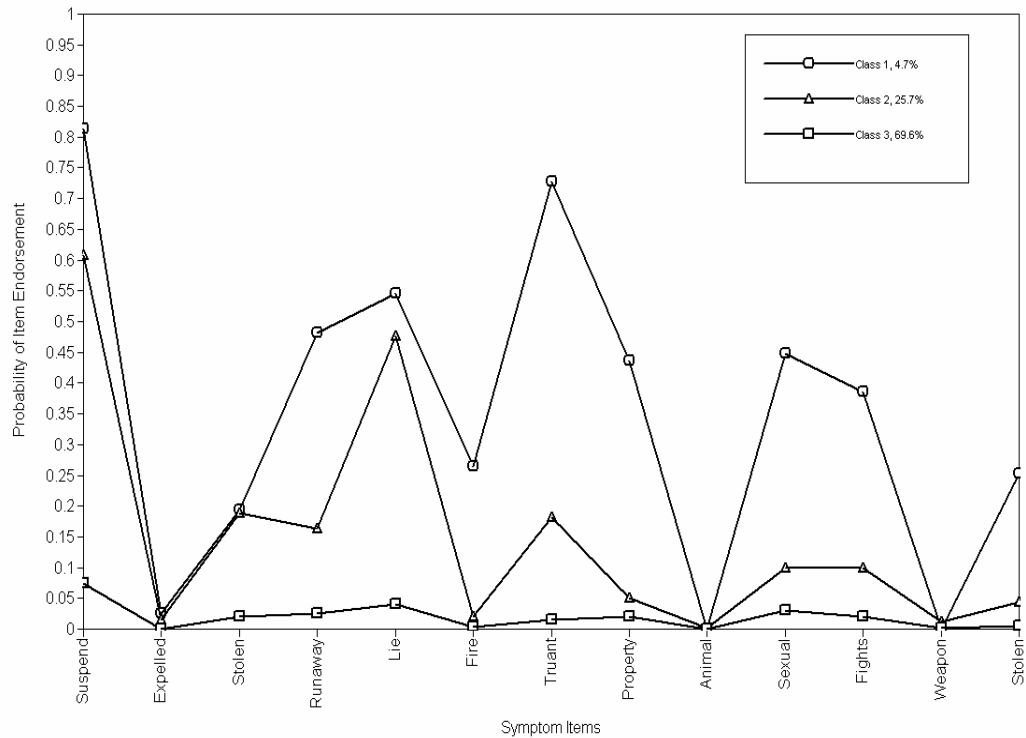*Figure 3 7.* Conduct Disorder Example: Two-class Latent Class Analysis Profile Plot.

Figure 3.8. Conduct Disorder Example: Three-class Latent Class Analysis Profile Plot.
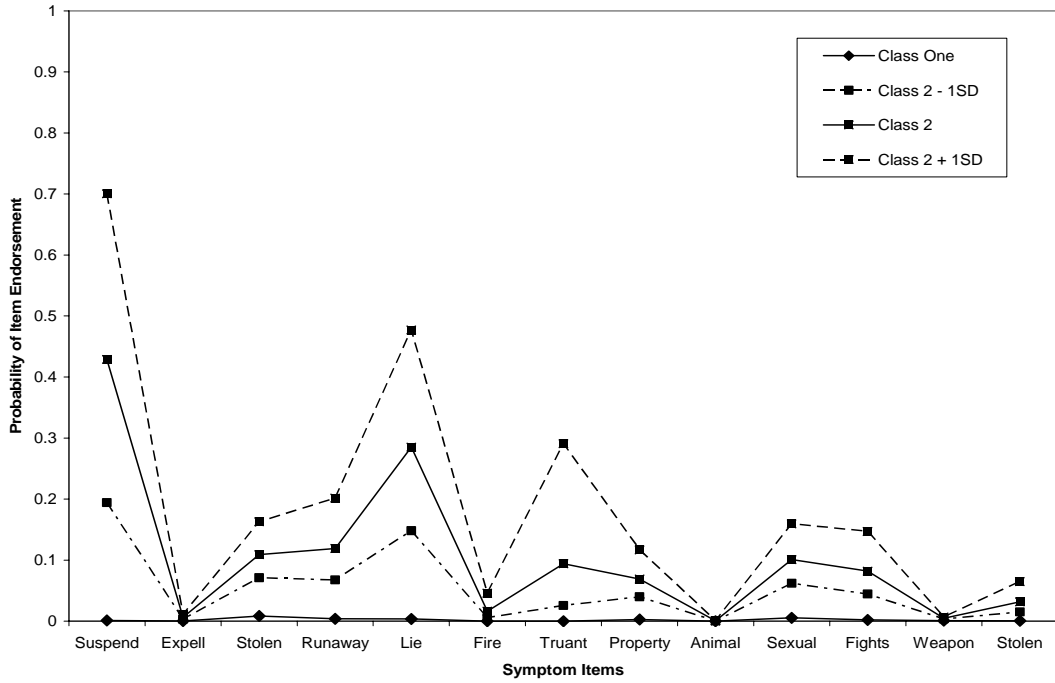
119

Figure 3.9. Conduct Disorder Example: Two-class, One-Factor FMM-2 Profile Plot.
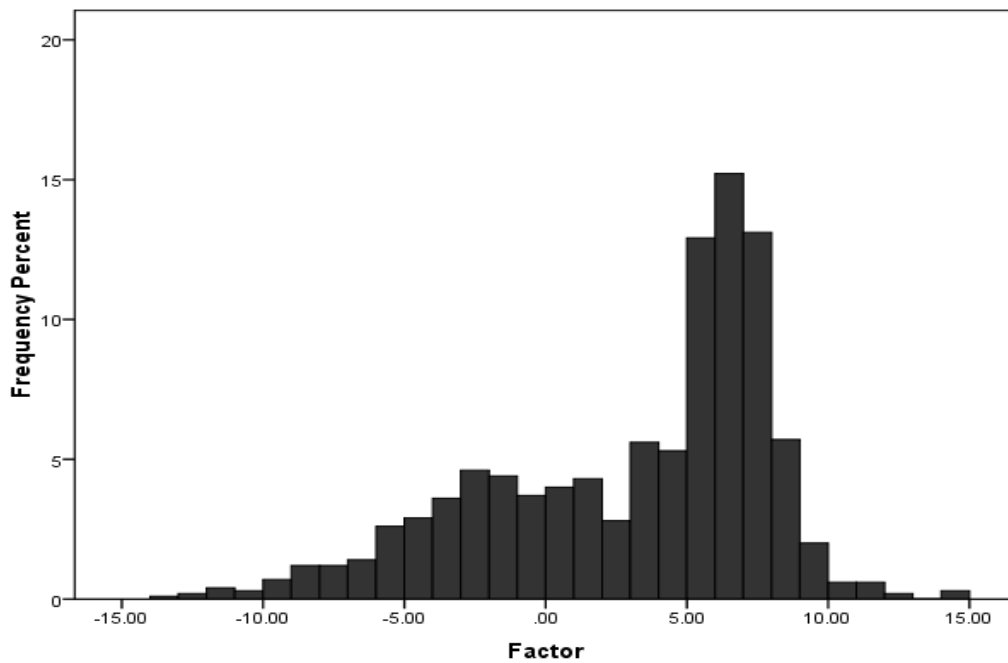

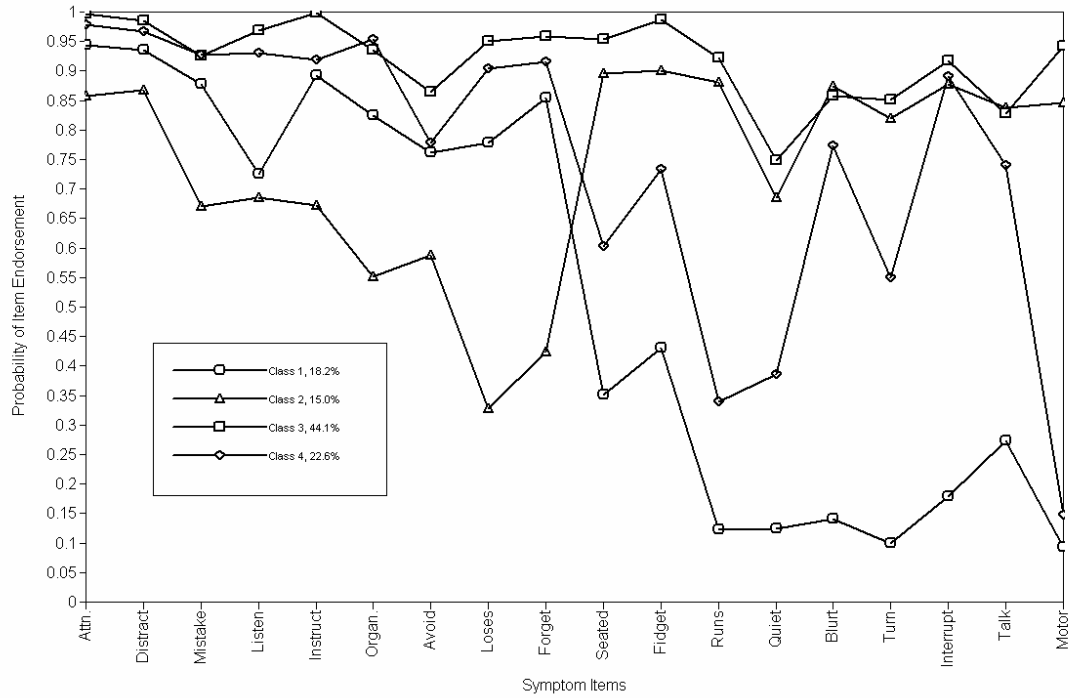*Figure 3.10.* CD Example: Two-class, One-factor FMM-2 Factor Distribution Plot.

120

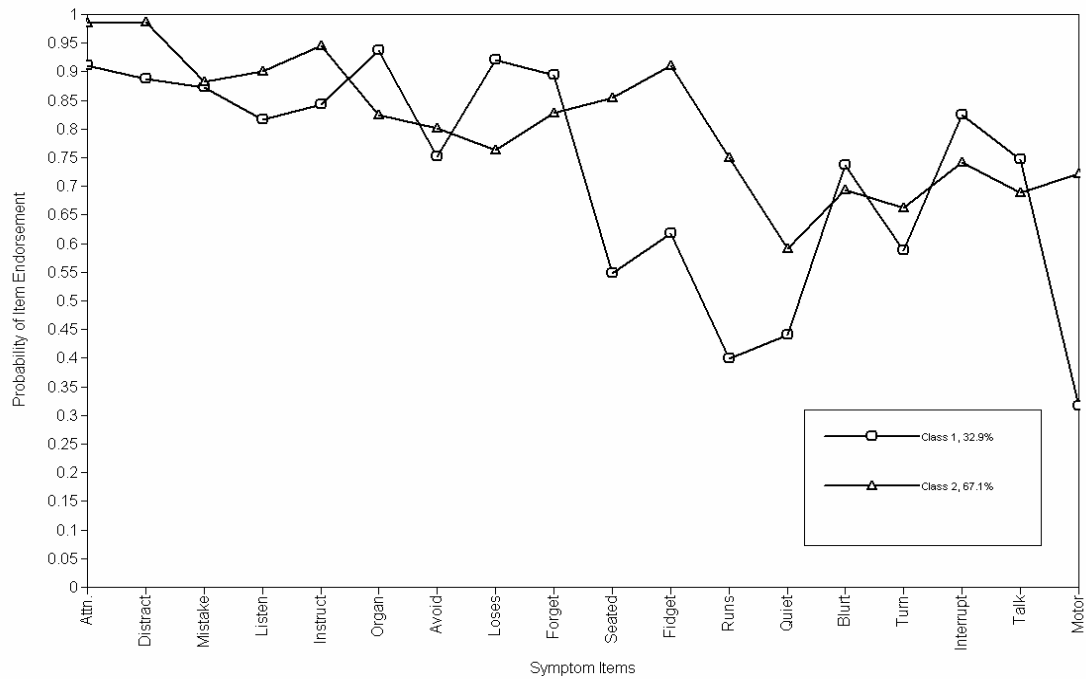*Figure 3.11*. ADHD Example: Four-class Latent Class Analysis Profile Plot.

Figure 3.12. ADHD Example: Two-class, Two-factor FMM-3 Profile Plot.

121

Appendix B: Mplus Code for Chapter 2

Example Mplus Input and Model Statements for Factor Mixture Model Variations

Input 1a: Factor Mixture Model 1 (FMM-1)

- Estimating both factor means in each class.

- Two-classes, two-factors.

- Class invariant thresholds (τ), class invariant factor loadings (λ), factor covariance matrix is equal to zero ($\psi = 0$), class varying factor means (α_k).

```
Data: File is example.dat;    ! This is how to comment out text.
Variable:
      Names are u1-u6; !List of variables in dataset.
      Usevariables are u1-u6;
            ! Specifies which variables are to be used in the analysis.
      Categorical = u1-u6;
            ! Specifies which dependent variable is categorical.
Classes = c(2); ! Defines a latent class variable called c with two
            !classes.
Analysis:
      Type = Mixture; !Specifies doing a mixture analysis.
      Starts = 100 10;
            ! Specifies total number of random starts and number of
            ! final stage iterations.
      Process = 2 (STARTS);
            ! Specifies that 2 processors are to be used in analysis
            ! and that random starts are to be spread out on the
```

```
                        ! multiple processors.
Model:

        %Overall% ! Specifies that model below will be applied to both

                ! latent classes.

        f1 by u1-u3; ! Defining a factor called f1 which is

                        !measured by items u1 to u3.

        f2 by u4-u6;

        f1-f2@0; ! Fixing factor variance at zero.

        f1 with f2 @0; ! Fixing covariance between f1 and f2 at

zero;

        %c#1% ! Indicates statements below apply to class 1 only.

        [u1$1-u6$1] (1-6);

        %c#2%

        [u1$1-u6$1] (1-6); ! Number in parentheses indicates fixing

                                !the threshold of this item to be the

                                !same in class 1 and class 2.
```

- By default, the last class will have the factor mean fixed to zero, and because there

is a factor loading fixed at one, a factor mean for a different class can be estimated as

well.

Input 1b: Factor Mixture Model 1 (FMM-1)

- Three-classes, one-factor.

- Estimating all factor loadings.

- Class invariant thresholds (τ), class invariant factor loadings (λ), factor

  covariance matrix is equal to zero ($\psi = 0$), class varying factor means (α_k).

123

```
Model: %Overall%

          f1 by u1-u3*;  !* indicates that all factor loadings are to

                         ! be freely estimated.

          f1@0;

     %c#1%

          [f1*]; !* indicates the factor mean is freely estimated.

          [u1$1-u6$1] (1-6);

     %c#2%

          [f1@1]; !Fixing factor mean at one.

          [u1$1-u6$1] (1-6);

     %c#3%

          [u1$1-u6$1] (1-6);
```

- By default the last class will have the factor mean fixed to zero. Because all of the

factor loadings are freely estimated, at least one of the factor means must be fixed for

model identification. Here, class 2 has the factor mean set to one.

Input 2: Factor Mixture Model 2 (FMM-2)

-   Two-classes, two-factors.

-   Class invariant thresholds (τ), class invariant factor loadings (λ), class varying

    factor covariance matrix (ψ_k), class varying factor means (α_k).

```
Model: %Overall%

          f1 by u1-u3;

          f2 by u4-u6;

          f1-f2@0;
```

124

```
        f1 with f2 @0;

%c#1%

        f1-f2;

        [f1-f2@0];

        [u1$1-u6$1] (1-6);

%c#2%

        f1-f2;

        [f1-f2*];

        [u1$1-u6$1] (1-6);
```

Input 3: Factor Mixture Model 3 (FMM-3)

- Two-classes, two-factors.

- Class varying thresholds (τ_k), class invariant factor loadings (λ), class

  invariant factor covariance matrix (ψ), factor means fixed at zero ($\alpha_k = 0$).

```
Model: %Overall%

        f1 by u1-u3;

        f2 by u4-u6;

        f1 - f2;

        f1 with f2;

        [f1-f2@0];

%c#1%

        [u1$1-u6$1];

%c#2%

        [u1$1-u6$1];
```

Input 4: Factor Mixture Model 4 (FMM-4)

- Two-classes, two-factors.

- Class varying thresholds (τ_k), class invariant factor loadings (λ), class varying

  factor covariance matrix (ψ_k), factor means fixed at zero ($\alpha_k = 0$).

```
Model: %Overall%

        f1 by u1-u3;

        f2 by u4-u6;

        [f1-f2@0];

    %c#1%

        f1-f2;

        f1 with f2;

        [u1$1-u6$1];

    %c#2%

        f1-f2*;

        f1 with f2;

        [u1$1-u6$1];
```

Input 5: Factor Mixture Model 5 (FMM-5)

- Two-classes, two-factors.

- Class varying thresholds (τ_k), class varying factor loadings (λ_k), class varying

  factor covariance matrix (ψ_k), factor means fixed at zero ($\alpha_k = 0$).

```
Model: %Overall%

        f1 by u1-u3;

        f2 by u4-u6;

        [f1-f2@0];

    %c#1%
```

```
            f1 by u2-u3;

            f2 by u3-u6;

            f1-f2;

            f1 with f2;

            [u1$1-u6$1];

    %c#2%

            f1 by u2-u3;

            f2 by u5-u6;

            f1-f2;

            f1 with f2;

            [u1$1-u6$1];
```

-The overall statement (%Overall%) defines the factors and in the class specific

statements (%c#1%, %c#2%) there is what looks like another definition of the factor, but

this is not defining the factors. This allows the loadings to be freely estimated in both

classes. Note that the in the class specific statement, the factor definition starts with the

second item and not the first. This is because in the overall statement where the factor is

first defined Mplus defaults to fixing the loading of the first item to one. Since the

loading of the first item is already fixed, it does not need to be set to be freely estimated

in the class specific statement so the class specific statements begins with the second

item.

Input 6: MI Investigation Example

- Two-classes, one-factor.

- Class invariant threshold anchor item: u1.

```
Model:      %Overall%

          f1 by u1 – u6;

          f1;

          [f1@0];

    %c#1%

          [u1$1]       (1);

          [u2$1]       (p1_2);

          [u3$1]       (p1_3);

          [u4$1]       (p1_4);

          [u5$1]       (p1_5);

          [u6$1]       (p1_6);

    %c#2%

          [u1$1]       (1);

          [u2$1]       (p2_2);

          [u3$1]       (p2_3);

          [u4$1]       (p2_4);

          [u5$1]       (p2_5);

          [u6$1]       (p2_6);


Model Constraint: New(difi1 difi2 difi3 difi4 difi5 difi6);

      difi1= p2_1 - p1_1;

      difi2= p2_2 - p1_2;

      difi3= p2_3 - p1_3;

      difi4= p2_4 - p1_4;

      difi5= p2_5 - p1_5;

      difi6= p2_6 - p1_6;
```

*Note***:** All of the example inputs and model statements above used dichotomous items, but factor mixture modeling can be conducted with other types of items as well with slight adjustments to the code above. Taking Input 4 as an example, the code can be adjusted for polytomous item by mentioning all the thresholds in the items. In the example below, the code has been adjusted for use with trichotomous items.

```
Model: %Overall%

        f1 by u1-u3;

        f2 by u4-u6;

        [f1-f2@0];

    %c#1%

        f1-f2;

        f1 with f2;

        [u1$1-u6$1];

        [u1$2-u6$2]; ! Mentioning second threshold of trichotomous

        item.

    %c#2%

        f1-f2*;

        f1 with f2;

        [u1$1-u6$1];

        [u1$2-u6$2];
```

Also using Input 4 as an example, the code can be adjusted for continuous items by mentioning the mean of the items, rather than threshold.

129

```
Model: %Overall%

          f1 by u1-u3;

          f2 by u4-u6;

          [f1-f2@0];

     %c#1%

          f1-f2;

          f1 with f2;

          [u1-u6]; ! Setting mean of items to be different across

          classes.

     %c#2%

          f1-f2*;

          f1 with f2;

          [u1-u6];
```

References

Bartholomew, D.J., and Knott, M. (1999). *Latent variable models and factor analysis*.
Kendall's Library of Statistics 7. Oxford University Press: New York.

de Boeck, P., Wilson, M., & Acton, G.S. (2005). A conceptual and psychometric
framework for distinguishing categories and dimensions. *Psychological Review*,
112, 129-158.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: The
Guilford Press.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor
analysis. *Pyschometrika*, 34.

Kaufman J., Birmaher B., Brent, D., Rao, U., Flynn, C., Moreci, P., et al. (1997).
Schedule for affective disorders and schizophrenia for school-age children –
present and liftetime version (KSADS-PL): Initial reliability and validity data.
Journal of the American Academy of Child and Adolescent Psychiatry, 36, 980-8.

Kim, Y.K. & Muthén, B. (2008). Two-part factor mixture modeling: Application to an
aggressive behavior measurement instrument. Forthcoming in *Structural Equation
Modeling*.

Kuperman, S., Schlosser, S. S., Kramer, J. R., Bucholz, K. K., Hesselbrock, V., Reich, T.,
et al. (2001). Developmental sequence from disruptive behavior diagnosis to
adolescent alcohol dependence. *American Journal of Psychiatry*, 158, 2022-2026.

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

Little, T.D. (1997) Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32, 53-76.

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

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

Lubke, G., Muthén, B., Moilanen, I., McGough, J., Loo, S., Swanson, J., Yang, M., Taanila, A., Hurtig, T., Jarvelin, M. & Smalley, S. (2007). Subtypes versus severity differences in the Attention-Deficit/Hyperactivity disorder in the northern Finnish birth cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 1584-1593.

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

Lubke, G. H., & Neale, M.C. (In press). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research*.

Masyn, K., Henderson, C., & Greenbaum, P. (In press). Exploring the latent structures of psychological constructs in social development using the Dimensional-Categorical Spectrum. *Social Development*.

McLachlan, G.J. (1987). On bootstrapping the likelihood ratio test statistic for the

number of components in a normal mixture. *Applied Statistics*, 36, 318-324.

McLachlan, G.J., Do, K-A., Ambroise, C. (2004). *Analyzing microarray gene expression data*. New York: Wiley.

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

Meredith, W. (1993) Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.

Mislevy, R.J., & Verhulst, N. (1990). Modeling item response when different subjects employ different solution strategies. *Psychometrika*, 55, 195-215.

Mislevy, R.J., & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika*, 61, 41-71.

Muthén, B. (1989). Dichotomous factor analysis of symptom data. In Eaton, & Bohrnstedt (Eds.), Latent Variable Models for Dichotomous Outcomes: Analysis of Data from the Epidemiological Catchment Area Program (pp. 19-65), a special issue of *Sociological Methods & Research*, 18, 19-65.

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

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, B. (2006). Should substance use disorders be considered as categorical or

dimensional? *Addiction*, 101 (Suppl. 1), 6-16.

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

Muthén, B. & Asparouhov, T. (2006). Item response mixture modeling: Application to
tobacco dependence criteria. *Addictive Behaviors,* 31, 1050-1066.

Muthén, B., Asparouhov, T. & Rebollo, I. (2006). Advances in behavioral genetics
modeling using Mplus: Applications of factor mixture modeling to twin data.
Twin Research and Human Genetics, 9, 313-324.

Muthén, L.K., & Muthén, B.O. (2008). Mplus (Version 5.1). Los Angeles, CA.

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.

Rose, R.J., Dick, D.M., Viken, R.J., Pulkkinen, L., Nurnberger, J.I., Jr., & Kaprio, J.
(2004). Genetic and environmental effects on conduct disorder, alcohol
dependence symptoms, and their covariation at age 14. *Alcoholism: Clinical and Experimental Research*, 28, 1541-1548.

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

Smalley, S.L., McGough, J.J., Del'Homme, M., NewDelman, J., Gordon, E., Kim, T.,
Liu, A., & McCracken, J.T. (2000). Familial clustering of symptoms and

disruptive behaviors in multiplex families with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39: 9.

Widamon, K.F., & Reise, S.P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K.J. Bryant, M. Windle, & S.G. West (Eds.), *The science of preventions: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, 105, 276-289.

Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62, 297-330.

Chapter Four: ACE Analysis for Latent Variable Phenotypes

Two common phenotypes used in behavioral genetics are diagnostic status and a sum score of symptom items. While these phenotypes are easy to obtain, they are not without limitations. Using diagnostic status may impose an arbitrary categorization and often results in misclassification (Dwyer, 1996). Using a sum score implies that each item has equal importance in phenotype definition, which may not always be true. For example, in conduct disorder, the presence of those items pertaining to more severe behaviors, such as physical violence against animals or sexual violence against other individuals, might be more important in defining the conduct disorder phenotype than more commonly occurring symptoms. And while differential weights can be used, an empirical basis for such weights may be difficult to obtain.

An alternative to using diagnostic status or sum scores as phenotypes are models that use latent variables, which have distinct advantages over commonly observed phenotypes. One key advantage is that more pieces of information can be used to inform phenotype definition. For example, all the symptom items in a diagnostic instrument can be used to form the latent variable rather than having to rely on a single sum score of the symptoms. A second advantage is that because multiple pieces of information are used, measurement error associated with the phenotype is reduced.

The use of latent variables as phenotypes in behavior genetic analyses is not new. There is a long standing tradition of using the factor, or latent continuous variable, from a factor analysis (FA) as a phenotype. But using a factor from a FA model can be

problematic if the construct of interest is thought to be categorical or if the phenotype is non-normal, which may occur when the outcome of interest is rare or infrequent.

This paper explores two latent variable phenotypes that overcome the limitations of observed phenotypes. The first latent variable phenotype is the latent class analysis (LCA) model, which uses a categorical latent variable to account for the unobserved heterogeneity in a sample. Unlike typical observed categorical phenotypes that use arbitrary cut-points to divide a construct into categories (e.g., using the number of symptoms present to define diagnostic status), LCA uses a model-based classification system. This paper explores a combination of the LCA and ACE models which uses the latent class variable as the phenotype, as suggested, but not explored in Muthén and Muthén (2005) and Muthén et al. (2006).

The second latent variable phenotype is the factor mixture model (FMM) with measurement invariance. The FMM uses a hybrid of continuous and categorical latent variables. In a FMM, the latent categorical variable is used to model the non-normality in the latent factor by accounting for different heterogeneous sub-populations in a sample that may have different values of the factor. This paper explores different options for the placement of the ACE model and discusses what impact the placement of the ACE model within the FMM has on the interpretation and heritability of the phenotype.

This paper aims to elucidate the LCA and FMM as phenotypes and to show how they can be combined with an ACE model to explore genetic variance. The application of these models as phenotypes is demonstrated through an example concerning conduct disorder in a sample of Finnish adolescent twins.

This paper begins with an introduction to LCA and FMM, and the combination of these models with an ACE model. The next section introduces the data used in the example and discusses how the LCA and FMMs were fit to the data. The final sections discuss the heritability results and further extensions of these models.

Background

Latent Class Analysis (LCA)

Instead of using an observed categorical phenotype, which has the limitations mentioned above, an alternative is to use a latent categorical phenotype, specifically the LCA model. LCA, introduced by Lazarsfeld and Henry (1968), is used to identify subgroups, or classes, of a study population. A model diagram of a twin LCA model is shown in Figure 4.1. There are two major concepts depicted in Figure 4.1, the observed outcomes or items that define the class and latent class itself. These can be seen in Figure 4.1 as ua_1-ua_r and ub_1-ub_r, and Ca and Cb, respectively. The boxes, ua_1-ua_r and ub_1-ub_r, represent the observed response items or outcomes for twin A and twin B, respectively. The outcomes in a LCA model can be categorical, continuous, count, censored, or nominal; this paper will focus specifically on dichotomous, categorical items. The circles with the letter C in the middle are the unordered, categorical latent class variables with K classes. Each twin has an individualized latent class variable. The arrows pointing from the latent class variable to the boxes below indicate that those items are measuring the latent class variable. This means that class membership is based on the observed response pattern of items. An important assumption, called the conditional or local independence

138

assumption, implies that the correlation among the observed outcomes is explained by the latent class variable. Because of this, there is no residual correlation between the items.

For a LCA model with categorical outcomes, there are two types of model parameters: conditional item probabilities and class probabilities. The conditional item probabilities are specific to a given class and provide information about the probability that an individual in that class will endorse a specific item. The class probabilities specify the relative size of each class, or the proportion of the population that is in a particular class. For more information on LCA see Hagenaars and McCutcheon (2002) and Muthén (2001).

Heritability Model: ADCE

Figure 4.2 shows an example of the univariate genetic model that is applied in this paper: the phenotype for each twin is represented by a box with the label p_A for twin A and p_B for twin B. The variance of the phenotype is partitioned into the additive genetic variance (A), dominant genetic factors (interaction of genetic effects at the same locus) (D), common environmental variance (C), and unique environmental variance (E) (Neale & Cardon, 1992).

The double-headed arrow between the A components for the twins implies that the A component is correlated in twins. In MZ twins, their gene influences (A and D) correlate 1.0 because MZ twins are genetically identical. In DZ twins, the additive gene influences correlate 0.5 because, like ordinary siblings, they share half of their segregating genes, on average. The genetic effects due to dominance are correlated 0.25 in DZ twins.

The common environment component (C) refers to all environmental influences that make the twins more similar to one another irrespective of zygosity. Examples of common environment are attending the same school and growing up in the same household. Common environment effects are suggested by DZ twin correlations that exceed half of the MZ twin correlation. In the model shown, only C is given, but this could be replaced by D. Dominant genetic effects are suggested by DZ twin correlations that are smaller than half of the MZ twin correlation. Models with both C and D cannot be estimated in data with only MZ and DZ twins. Usually, an assumption is made that the common environment correlates 1.0 between twin pairs of both types, which is shown in Figure 4.2, with the additional assumption that mating is random for the trait in question.

Unique environmental effects (E) refer to all environmental influences that make the twins less similar. Examples of this include friends or activities that are not shared by the twins. The E variance component is uncorrelated between twins and also includes measurement error.

Heritability, or the proportion of variation in the phenotype attributable to genetics, can be calculated as:

$$h^2 = a^2 / (a^2 + c^2 + e^2),$$

where a, c, and e are the path values that appear in Figure 4.1.

In order to determine heritability for phenotypes that are binary or ordinal, one can use the liability threshold model (Neale and Cardon, 1992). This model assumes that there is an underlying, latent continuous variable, called the liability, which models how susceptible individuals are to developing a disorder (Thurstone, 1927; Wright, 1934;

Falconer, 1965). The distribution of the liability can be divided based on the thresholds in the categorical variable so that, in the case of a diagnostic status phenotype for example, once an individual has more liability than the threshold, that individual is considered to be affected by the disorder. Prescott (2004) presented a version of this model where the categorical phenotype was transformed into a latent, continuous variable through the use of a probit link function. When using a categorical phenotype, the E variance component generally cannot be estimated directly from the model. In Prescott's case, the total variance of the phenotype was specified to be one, so the E component could be calculated as one minus the sum of the other sources of variation.

Combining LCA and ACE. The idea of using a latent class model in behavioral genetics is not new. Eaves et al. (1993) and Rasmussen et al. (2002), among many others, use LCA to identify categorical phenotypes. But, this paper is among the first to take steps towards combining the LCA and ACE models.

One option for combining the LCA and ACE models, which is most commonly used in practice by behavioral geneticists, is a three-step ad hoc method. Described by Muthén et al. (2006), the first step is to estimate the LCA model parameters, and the second step is to assign individuals into their most likely latent class based on the posterior probabilities of class membership. The third, and final step, is to use the observed, categorical variable of assigned class membership as the phenotype in a liability threshold model version of ACE analysis. But using this method will likely result in biased parameter estimates and underestimated standard errors because the classification of individuals into their most likely latent class ignores the fractional class

membership suggested by the posterior probabilities and also treats the classification as free of any sampling error (Clark and Muthén, 2009). The method this paper proposes avoids this issue by performing a single-step analysis which uses the latent class variable as the phenotype instead of most likely class membership.

In this paper, the LCA and ACE models are combined using a liability threshold model, which assumes that class membership is due to one underlying liability. Figure 4.3, which displays the model diagram for the LCA and ACE combination model, shows that there are two latent class variables, *Ca* and *Cb*, one for each twin. Note that all parameters in the LCA are held equal across twins to ensure that the latent classes are being measured the same way and to ensure that there is no difference between twin A and twin B. This includes the conditional item and class probabilities. The class probabilities are restricted so that the table of joint latent class probabilities for the two twins is symmetric and have the same marginal probabilities (Muthén et al., 2006). Also note that *Cb* does not influence twin A's items and vice versa. The latent class variables, *Ca* and *Cb*, are transformed into continuous latent variables, *C*a* and *C*b*, using equation 3 below. The variance of *C*a* and *C*b* is then decomposed.

The formulation of this model is composed of two pieces: the measurement or LCA model and the liability threshold model. Consider a LCA model with *r* observed binary items, *u*, and a categorical latent variable *C* with K classes ($c = k$; $k = 1, 2, \ldots, K$). The marginal item probability for item *q*, individual *i*, and twin pair *j* for $u_{ijq} = 1$ is:

$$P(u_{ijq} = 1) = \sum_{k=1}^{K} P(C_{ij} = k) P(u_{ijq} = 1 \mid C_{ij} = k). \qquad (1)$$

142

Assuming conditional independence, the joint probability of all the *r* observed items is

$$P(u_{ij1}, u_{ij2}, ..., u_{ijr}) = \sum_{k=1}^{K} P(C_{ij} = k) P(u_{ij1} | C_{ij} = k) P(u_{ij2} | C_{ij} = k) ... P(u_{ijr} | C_{ij} = k). \quad (2)$$

In order to connect the liability threshold model to the LCA model, the latent class variable is related to the liability through the function below such that:

$$C_{ijq} = \begin{cases} 1 & \text{if } C^*_{ijq} > \tau_{iq} \\ 0 & otherwise, \end{cases} \quad (3)$$

where τ is the threshold on the liability above which an individual is considered affected by the disorder. The liability is then represented as the following sum, as in the classic ACE model:

$$C^*_{ij} = aA_{ij} + cC_{ij} + eE_{ij}, \quad (4)$$

Where C^*_{ij} is the transformed latent continuous variable version of the categorical latent variable C, and, A_{ij}, C_{ij}, and E_{ij} are the latent additive genetic, shared environment, and unshared environment components. Note that both the latent class variable and the shared environment component are denoted by the letter C by convention in the mixture modeling and behavioral genetics literatures, respectively. In this paper, the latent class variable, *C*, is italicized while the share environment component, C, is not. Similar to the Prescott model, the E component cannot be directly estimated, but because equation 3 is used it is considered to be equal to $\pi^2 / 3$.

Because a liability threshold model is applied to the latent class variable, only solutions with two classes or three or more ordered classes can be used. One way to

examine whether the classes are ordered is to plot the mean item endorsement probability

for each class and examine if the profiles are parallel to one another. If they are parallel,

then the classes can be considered ordered.

Factor Mixture Model

A categorical outcome is not always the best representation of a phenotype

because it implies that within a category, everyone is phenotypically equivalent, which

may not always be true. For example, if the categories are affected or not affected by a

disorder, the label of "affected" can mask a great range of severity and impairment in

those individuals within that category. So, behavioral geneticists often use quantitative

phenotypes to model these differences. One quantitative phenotype that is commonly

used is the latent factor from a factor analysis (FA).

But, the FA model makes a strong normality assumption which may be violated

when studying rare or infrequent outcomes. When the factor is non-normal, one

alternative is a non-parametric representation of the factor which uses latent categorical

variables to represent the non-normality of the factor. Called a factor mixture model

(FMM; Muthén, 2008), an example of this model for twins is shown in Figure 4.4. The

boxes at the bottom of the model diagram are the r dichotomous symptom items for each

twin. Above this are circles that represent the latent variables: the latent class variable,

the circle with the C in the middle, and a latent continuous factor, the circle with the η in

the middle. Similar to twin LCA, each twin has individualized latent continuous and

latent categorical variables as indicated by the subscripts a and b. In this variation of the

FMM, the factor loadings and item thresholds are invariant or equal across classes. This

suggests that the interpretation of the factor remains the same in each class or, put another way, the same factor runs through the entire population. Because the factor loadings and item thresholds are invariant, differences in the latent classes arise due to differences on the factor in each class, specifically the factor mean and variance. It is these differences in factor mean and variance in each class that give rise to the non-parametric representation of the factor.

There are two variations of the FMM shown in Figure 4.5 which differ based on whether the factor variance is zero or variant across classes. For the first variation, called FMM-1 (Clark and Muthén et al., 2009), Figure 4.5a shows that the factor distribution is categorical, because the factor variance is set to zero. In Figure 4.5a, each latent class represents a category of the factor. For the second variation, called FMM-2 (Clark and Muthén et al., 2009), Figure 4.5b shows that the factor distribution is continuous, but non-normal. The factor in the second variation is a mixture of normal distributions located at different points on the factor distribution. Each normal distribution is a latent class. For more information about these variations and the FMM see Clark & Muthén et al. (2009).

One way to combine the FMM and ACE model is shown in Figure 4.6. Because the measurement invariance ensures that the factor is being measured the same way throughout the population, the ACE model is on the factor. Additionally, the FMM is restricted so that the latent class variable is measured the same way across both twins. Similar to the LCA and ACE combination, C_b and η_B do not influence twin A's items and vice versa. The formulation of this model for binary outcomes can be broken down into

two parts: the measurement model (FMM) and the ACE model. A general representation of a FMM with measurement invariant factor loadings and item thresholds for $k = 1, 2, . . ., K$ latent classes with binary items can be specified as follows:

$$u^*_{ijk} = \tau + \Lambda \eta_{ijk} + \varepsilon_{ik},$$

$$\eta_{ijk} = \alpha_{jk} + \zeta_{ijk},$$

where,

$$\zeta_{ik} \sim N(\mathbf{0}, \Psi_{jk}), \tag{5}$$

And

$$u_{ijq} = \begin{cases} 1 & \text{if } u^*_{ijq} > 0 \\ 0 & otherwise. \end{cases}$$

$$\tag{6}$$

where u_i are individual i's responses in twin pair j, which is a p vector of observed outcomes; u^*_i is individual i's latent response vector; q is one of the items in a p vector; τ is a p vector of measurement thresholds; Λ is a $p \times m$ matrix of factor loadings, where m is the number of factors; η_i is a m vector of factor scores; ε is p vector of residuals; α is a m vector of the intercepts of the factors; ζ_i is a p vector of residuals which is assumed to be normally distributed with mean zero and variance Ψ. In the formulation above the only parameters that are varying across twins are the factor mean, α, and factor covariance matrix, Ψ, as indicated by the subscript j on these parameters in the equations above.

The ACE can be formulized as

146

$$\eta_{ij} = \mathrm{a}A_{ij} + \mathrm{c}C_{ij} + \mathrm{e}E_{ij}, \qquad\qquad (7)$$

where A_{ij}, C_{ij}, and E_{ij} are the latent additive genetic, shared environment, and unshared environment components. Alternative methods for combining the FMM and ACE when there is no measurement invariance are presented in the discussion section below.

Example

This example explores the heritability of conduct disorder in sample of adolescent Finnish twins. Also, this example demonstrates the use of latent variables as phenotypes in an ACE analysis. LCA, FA, and FMM interpretations are presented without making a distinction between MZ and DZ twins. All analyses were conducted in Mplus (Muthén & Muthén, 1998-2008) using special features which will be available in the forthcoming version six of the program. Example Mplus inputs for these models are available in the Appendix C.

Sample Description

This example considers conduct disorder (CD) in a subsample of a population-based study of Finnish twins, called the FinnTwin12 (FT12; Kaprio et al., 2002; Rose et al., 2001). The original study consisted of five consecutive birth cohorts (1983-1987) of twins identified in Finland's Central Population Registry. From this epidemiological sample, a subset of families of twins was selected for an intensive study. This subsample was moderately enriched to include more twins with familial alcoholism risk and is described in detail elsewhere (Rose et al., 2004).

In the subsample, there were 893 twin pairs in the study with about an equal number of boys and girls. Zygosity was determined using a well-validated questionnaire

completed by both twins at base-line, as described elsewhere (Kaprio et al., 1995) and confirmed by genetic marker data in may pairs. Thirty-eight percent of the twin pairs were monozygotic, while the other sixty-two percent were dizygotic. Same-sex twin pairs comprised seventy percent of the sample, while the remaining thirty were opposite-sex. For the purposes of introducing these new and already complicated methods, this paper will focus on sex-invariant models, but the methods shown can be applied to sex-limitation models.

These analyses focused on interview reports of thirteen dichotomous symptom items measuring CD which come from the Child Semi-Structured Assessment for the Genetics of Alcoholism, Adolescent version (C-SSAGA-A). The C-SSAGA-A is a polydiagnostic instrument that was developed by the Collaborative Study on the Genetics of Alcoholism (COGA; Kuperman et al., 2001). The data for this illustration are from the first follow-up when the twins were 14 years old. In the analyses presented below, one item pertaining to whether the individual had ever been physically cruel to animals was excluded from the analysis because the lack of individuals endorsing the item led to there being zero variance for the item.

In order to obtain a diagnosis of CD, three or more symptoms must be present. In this sample, 56% of children reported no symptom presence while 12% of the sample reported having three or more symptoms present and, therefore, met criteria for a CD diagnosis. Symptom presence ranged from 0.1% (Is physically cruel to animals) to 25% (Has ever been suspended from school). For a more extensive description of the interviewed sample and analysis of CD in these Finnish twins see Rose et al. (2004).

148

Clark and Muthén et al. (2009) fit LCA, FA, and FMM models to these data ignoring the dependency of the twins.

Latent Class Analysis

A three-class twin LCA was chosen based on the log likelihood, number of parameters, and Bayesian Information Criteria (BIC) (See Table 4.1). The item profile plots, which display the estimated item probabilities for a given class, are shown in Figure 4.7a and 4.7b for the two- and three-class solutions. The item profile plot for the three-class solution (Figure 4.7b) shows that the profiles of the classes are not parallel, which suggests that the classes are not ordered. This means that the three-class solution cannot be used in an ACE analysis, and instead the two-class solution should be used. For pedagogical purposes, both the two- and three-class solution will be interpreted.

The item profile plot for the two-class solution is shown in Figure 4.7a. Class One, 74%, has low symptom presence as indicated by low item endorsement probabilities. Because of the large size of the class and low symptom presence, this can be considered the "unaffected" or asymptomatic class. Class Two, 26%, is defined by the items relating to suspension, lying, and truancy; with the other items have relatively low probabilities of symptom endorsement, 0.30 or lower. Table 4.2a shows the percentage of the sample in each class by twin for the two-class solution. Note that concordance, twins in the same pair being in the same class, is high at about 98% in MZ twins and 80% in DZ twins.

Figure 4.8b shows the item profile plot for the three-class solution. As in the two-class solution, there is one class that has low item endorsement across all items. This

class is slightly smaller than the asymptomatic class from the two-class solution, 63%, because some of the individuals were absorbed into the two remaining classes. Supporting this idea are the lower endorsement probabilities of the suspension and truancy symptom items in the unaffected class. The class that was defined by the suspension, lying and truancy items in the two-class solution has now been split into two classes in the three-class solution. The first of these classes is similar to the one in the two-class solution with high probabilities of endorsing the suspension and lying items, but the probability of endorsing the truancy item has drastically decreased. The probability of endorsing other symptoms items is relatively low in this class. Potentially, some individuals in this class may meet criteria for a CD diagnosis. The second of these classes, 4.7% of the sample, still has high probabilities of suspension and lying, but also has high probabilities of running away, property destruction, forcing someone to participate in sexual activities, and starting fights. This class can be thought of as those individuals that are likely to meet criteria for CD. Table 4.2b shows the percentage of the sample in each class by twin for the three-class solution. As in the two-class solution, class agreement is high with about 97% of MZ twins and 72% of DZ twins being in the same class.

An ACE model with two classes, similar to the model presented in Figure 4.3, was fit to the CD data. The results are presented in Table 4.3. Both the A and C parameters are significant and the E parameter is $\pi^2/3$ because the latent class variable is transformed into a latent factor using equation 3. The heritability estimate is 0.79, with a

95% confidence interval of 0.783 to 0.797, which indicates that about seventy-nine percent of the variation in the phenotype is due to genetic influences.

Factor Analysis

Twin FA models were fit to the CD data. It was found that one factor was sufficient to fit the data. Table 4.1 shows the resulting log likelihood for this model. When compared to the twin LCA models, the one-factor FA model has a lower BIC value which suggests that the twin FA model outperforms the twin LCA solutions. Table 4.4 displays the factor loadings for the one-factor model. All of the items have significant loadings on the factor with the exception of the item pertaining to use of a weapon when fighting.

A common factor ACE model was also fit to this data. The results of this model are shown in Table 4.3. Compared to the LCA and ACE model, the FA and ACE model has a higher log likelihood and a lower BIC value. In the ACE portion of the FA and ACE model, both the A and E estimates were significant, while the C estimate was not. The heritability was estimated at 0.84, with a 95% confidence interval of 0.831 to 0.849, which suggests that over four-fifths of the variance in the phenotype can be explained by genetic variance.

Factor Mixture Model

A twin FMM was applied to the CD data. It was found that two classes and one factor were sufficient to fit the data. Table 4.1 shows the resulting log likelihood for this model. The FMM outperforms the two-class LCA model because both the log likelihood is higher and the BIC is lower for adding relatively few parameters, but when compared

to the three-class LCA model only the BIC was better in the FMM solution. A comparison of the FMM solution to the FA solution shows that the FMM solution has higher log likelihood, but the FA solution has a lower BIC value because it has six fewer parameters.

Figure 4.8 shows the item profile plot for the two-class, one-factor FMM solution. Note that the profile plot displays the item endorsement probabilities at the mean of the factor. If an individual is not at the mean of the factor, the item endorsement probabilities will vary from what is shown in the plot. The first and largest class, 92.7%, is the lowest line on the profile plot, with almost all items having low probabilities of symptom endorsement. The second class, 7.3% of the sample, has slightly elevated probabilities of endorsing the symptom items related to suspension, lying, and truancy, but these probabilities are still small.

Comparing the factor loadings of the twin FMM and FA solutions, shown in Table 4.4, suggests that while the factors are similar in interpretation, they are not equivalent. Table 4.4, presents the unstandardized and standardized factor loadings. The factor loadings were standardized to take into account differences in the factor variance across models in order to be able to compare the loadings. The FMM had class invariant factor loadings, hence the unstandardized factor loadings are the same in each class. But, the factor variances are non-invariant, so the standardized factor loadings are different in each class. Comparing the loadings from each class in the FMM solution to the loadings from the FA solution, the loadings in class two are lower than the loadings in the FA solution while the class one loadings are higher. This suggests that the factor, which is

considered the severity, has more influence in class one, the asymptomatic class. The higher loadings in class one also indicate that the with-in class correlation among the items is higher in class one. Note also that the items relating to suspension from school, fire starting, and truancy have the highest loadings in both solutions suggesting that these items are important in both the FA and FMM solutions.

A two-class, one-factor FMM with an ACE model on the latent factor, similar to Figure 4.6, was fit to the CD data. The results of this model are shown in Table 4.3. Comparing the FMM solution to the FA and LCA solutions, the FMM and FA solutions have the same log likelihood, which are both higher than the LCA solution. When comparing the log likelihood of different solutions, one is looking for the solution with the highest value. But, the FA solution has a lower BIC value because it has three fewer parameters than the FMM solution. When comparing the BIC values of different solution, the solution with the lowest BIC is regarded as the best model. In the FMM solution, both the A and C parameters were significantly different from zero. The E parameter, however, was not significantly different from zero and was estimated to be close to zero. This may be due to the high item agreement between twins in the same pair. The heritability was estimated at 0.85, with a 95% confidence interval of 0.842 to 0.858, which indicates that about eighty-five percent of the variation in the phenotype is attributable to genetic variation.

Discussion

This article presented two latent variable models for modeling phenotypic variation. The first model applied a liability threshold model to the classes in a LCA. The

second model applied an ACE model to the latent factor in a FMM with invariance of the measurement parameters to model a non-normal latent factor. Both of these models, and a FA and ACE model were applied to CD data in a sample of adolescent Finnish twins and all models yielded heritability estimates greater than 0.70.

In the FMM and ACE combination model presented in this paper, it was assumed that the measurement parameters, the factor loadings and item thresholds, were invariant across classes, assuring that the factor was being measured the same way throughout the sample. But, the requirement of invariant measurement parameters is often too restrictive for certain items. One can relax the measurement invariance requirement to allow factor loadings or item thresholds that are specific to each class, similar to the FMM-3 through - 5 models presented in Clark and Muthén et al. (2009). But once this restriction is relaxed, the factor is no longer being measured the same way throughout the population, and so the factor cannot be used as the phenotype in an ACE analysis. In FMMs that do not have measurement invariance, another option for combining the FMM and ACE model is to have the ACE model on the latent classes as shown in Figure 4.9. Similar to the combination of the LCA and ACE models presented, the liability threshold model is used to transform the latent class variables, C_a and C_b, into latent continuous factors, C^*_a and C^*_b. The variance of the factors C^*_a and C^*_b is then decomposed. But, this model can be difficult to estimate because of the intense computation required because of the four latent factors: C^*_a , C^*_b, η_a, and η_b. In this paper, the FMM with an ACE model on the classes was not fit to the CD data because a FMM with measurement invariance was indicated for this data.

Presented in Muthén et al. (2006), a second option for combining a FMM and ACE model when there is no measurement invariance, is to have the ACE on the factors. But, because there is no measurement invariance, the factors are not being measured the same way across the population or, said another way, there is a different factor in each class. Because there is a different factor in each class, there is a different ACE model in each class as well. This model yields a different heritability for each latent class which implies that there are potentially different phenotypes with different heritabilities for the same construct of interest. For example, when exploring attention-deficit hyperactivity disorder (ADHD), two latent classes related to having only inattention symptoms and only having hyperactive symptoms are often found. Using the Muthén et al. (2006) model in the ADHD context allows for the possibility of having different heritiabilities in the two types of ADHD classes. This would help answer the question of whether predominantly inattentive ADHD is more heritable than predominantly hyperactive ADHD.

The models presented in this paper can aid in further understanding CD because their implied underlying structures of CD are much closer than previous statistical models to the way that developmental psychopathologists view the etiology of CD. For example, Moffit (1993, 2006) proposed a "developmental taxonomy" that distinguishes between those individuals with early- and late-starting delinquency. But, Lahey and Waldman (2003, 2005) have proposed that the "developmental taxonomy" of early- and late-starting delinquency are not qualitatively-distinct classes, but instead represent an artificial dichotomization of a continuum of differences among antisocial trajectories

155

(D'Onofrio et al., 2009). The models presented can be seen as blending the Moffitt, and Lahey and Waldman views of CD. The LCA and ACE model and the FMM with an ACE model on the latent classes imply that there are distinct classes of CD, similar to Moffitt's theory, but the classes arise because of differences on an underlying continuous distribution, similar to Lahey and Waldman's theory. Additionally, competing theories of the etiology of CD can be tested using the models here.

Table 4.1

Model Fit for Latent Variable Twin Models

Model	LL	par	AIC	BIC
Two-Class LCA	-4510	26	9072	9196
Three-Class LCA	-4436	41	8955	9152
One-Factor FA	-4460	25	8971	9091
Two-Class, One-Factor FMM	-4452	31	8967	9116

Table 4.2

Pairwise Distribution (%) of Twins in Each Latent Class by Zygosity. a. Two-Class Twin

LCA solution. b. Three-Class Twin LCA solution.

a.

MZ	Twin B	
Class	1	2
Twin 1	69.1	0.5
A 2	0.5	29.6

DZ	Twin B	
Class	1	2
Twin 1	61.0	9.8
A 2	8.9	19.6

b.

MZ	Twin B		
Class	1	2	3
Twin 1	5.6	1.2	0
A 2	1.2	26.9	0
3	0	0	65.1

DZ	Twin B		
Class	1	2	3
Twin 1	1.2	2.1	0.5
A 2	2.1	19.3	11.4
3	0.5	11.4	51.6

Table 4.3

ACE Estimates for Latent Variable Models

Phenotype	Model	LL	par	BIC
LCA: 2-Class	AC	-5088	28	10366
FA:1-Factor	ACE	-5035	27	10253
FMM: 2-Class, 1-Factor	ACE	-5035	30	10274

Phenotype	A	C	E	H^2
LCA: 2-Class	45.3 (7.57)	35.1 (9.64)	$\sqrt{(pi^2/3)}$	0.79 (0.11)
FA:1-Factor	1.91 (0.24)	0.82 (0.35)*	0.23 (0.29)	0.84 (0.14)
FMM: 2-Class, 1-Factor	1.91 (0.24)	0.79 (0.33)	0.03 (0.29)*	0.85 (0.12)

Note. *Insignificant at the 5% level. A, C, and E estimates in this table are standard deviations.

Table 4.4

Factor Loadings for Twin One-Factor Factor Analytic (FA) and Two-Class, One-Factor

Factor Mixture Model (FMM) Solutions

Loadings	Unstandardized			Standardized		
Model	FA	FMM		FA	FMM	
Number of Factors	1-Factor	2-Class, 1-Factor		1-Factor	2-Class, 1-Factor	
		Class 1	Class 2		Class 1	Class 2
Items						
Suspend	1.00	1.00	1.00	2.04	2.19	0.61
Expelled	0.80	0.84	0.84	1.64	1.84	0.51
Stolen	0.51	0.52	0.52	1.04	1.14	0.32
Runaway	0.70	0.72	0.72	1.43	1.58	0.44
Lie	0.87	0.91	0.91	1.78	1.99	0.55
Fire	1.08	1.22	1.22	2.21	2.67	0.74
Truant	1.08	1.20	1.20	2.21	2.63	0.73
Property	0.66	0.68	0.68	1.35	1.49	0.41
Sexual	0.59	0.61	0.61	1.21	1.34	0.37
Fights	0.65	0.67	0.67	1.33	1.47	0.41
Weapon	0.35*	0.36*	0.36*	0.72*	0.79*	0.22*
Stolen	0.75	0.80	0.80	1.53	1.75	0.49
Factor Variance	4.18	4.79	0.37*			

Note. * Insignificant at the 5% level.
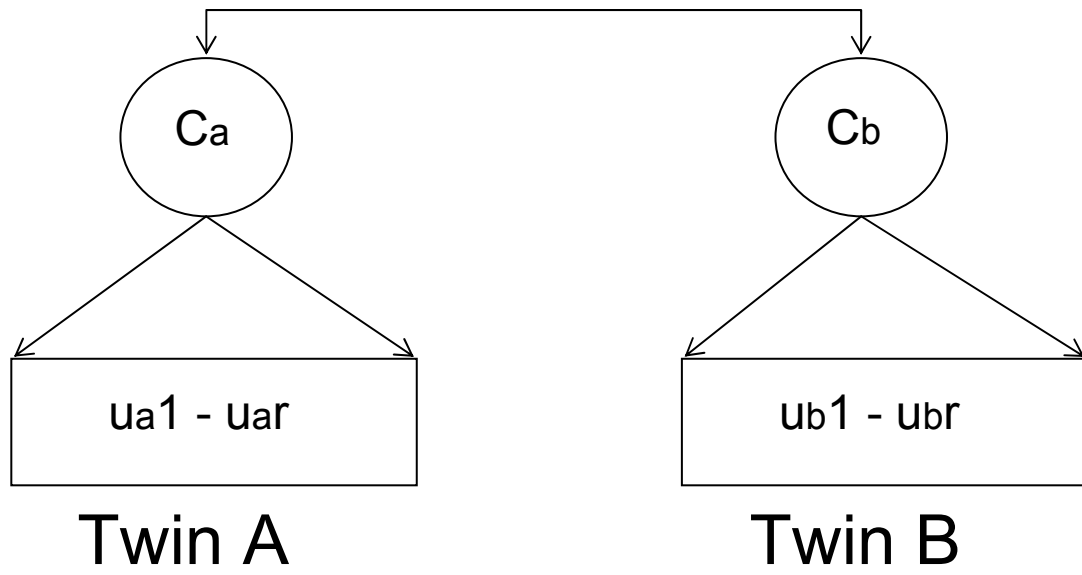
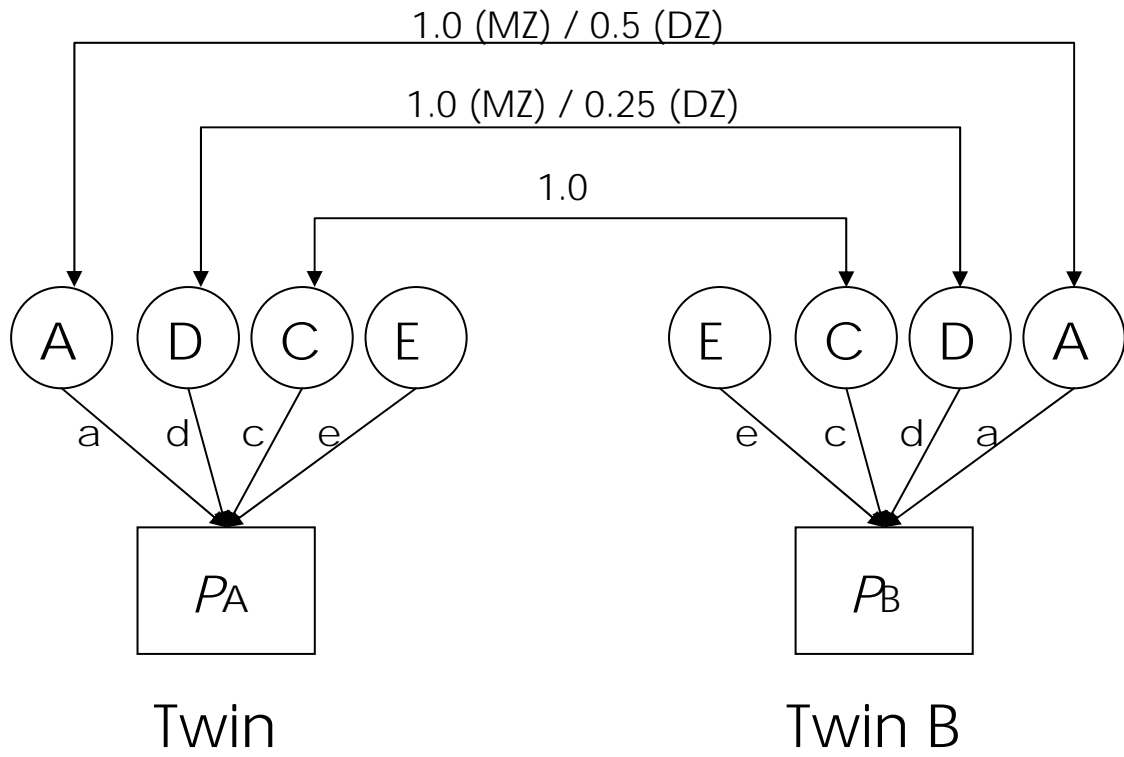*Figure 4.1.* Twin Latent Class Analysis Model Diagram.
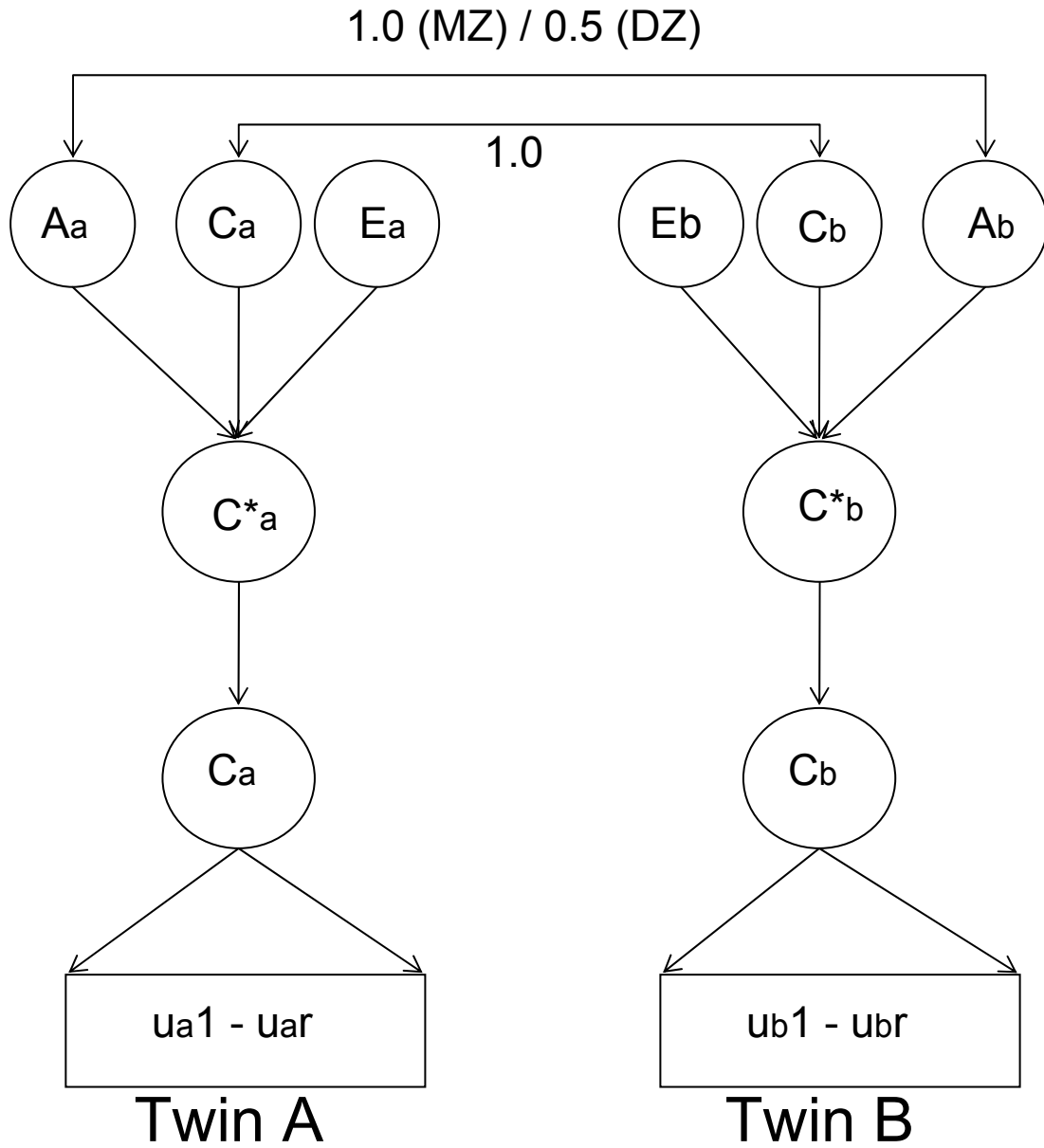
Figure 4.2. ADCE Model Diagram.

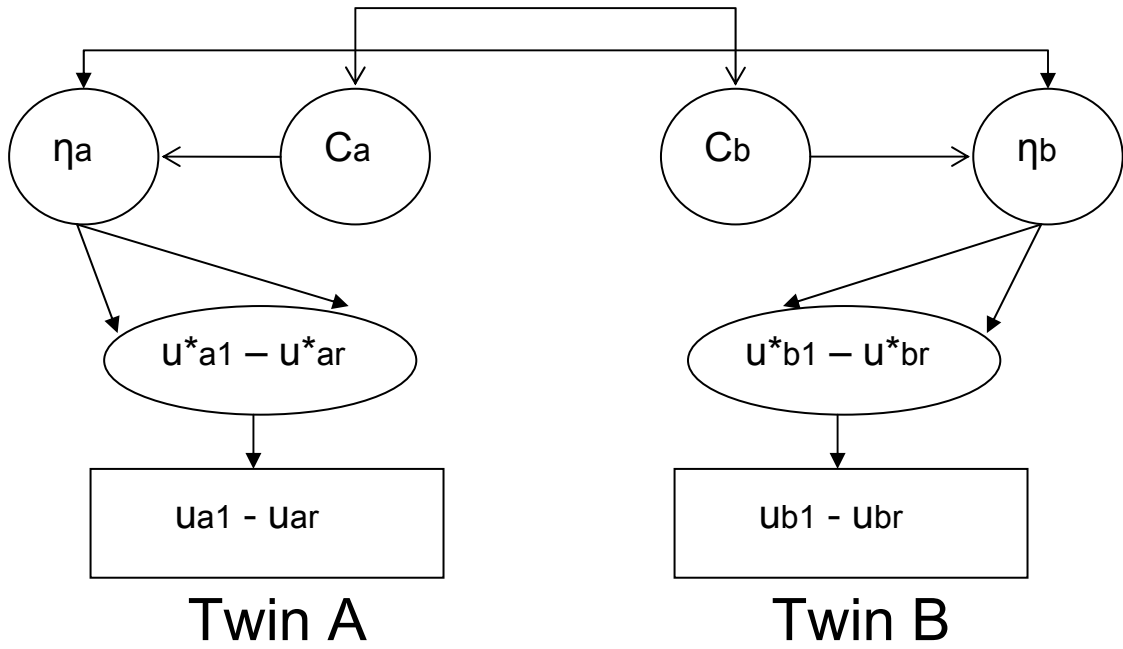*Figure 4.3.* Latent Class Analysis and ACE Model Diagram.

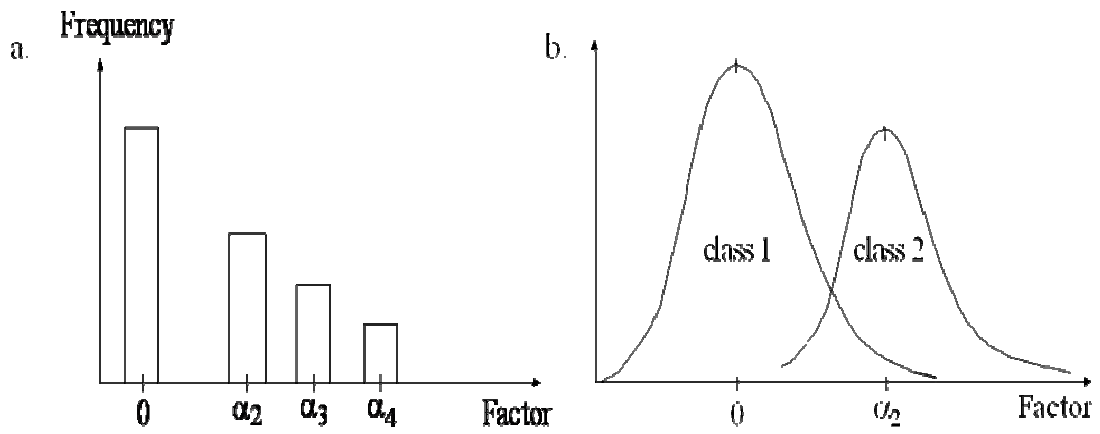*Figure 4.4.* Twin Factor Mixture Model Diagram.

Figure 4.5. FMM Variation Factor Distributions. A. Factor distribution with factor

variance set to zero. B. Factor distribution with factor variance varying across latent
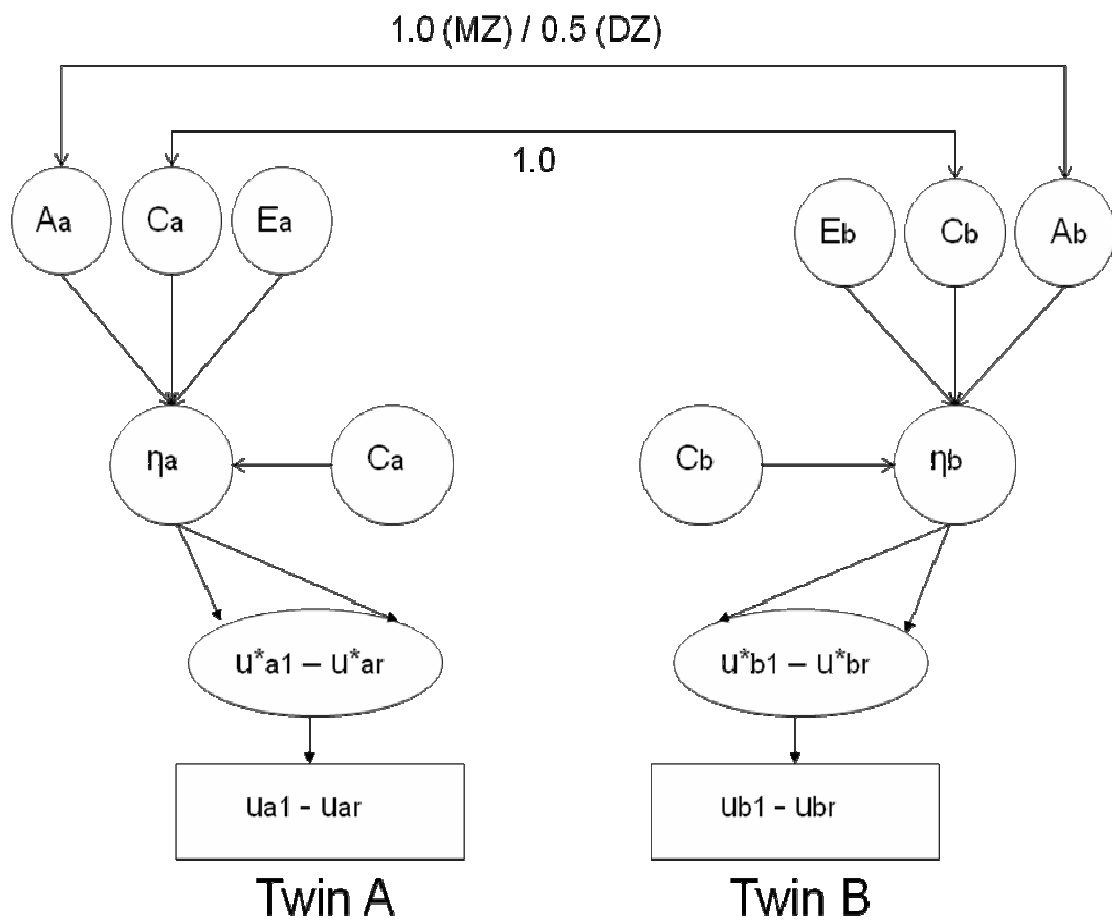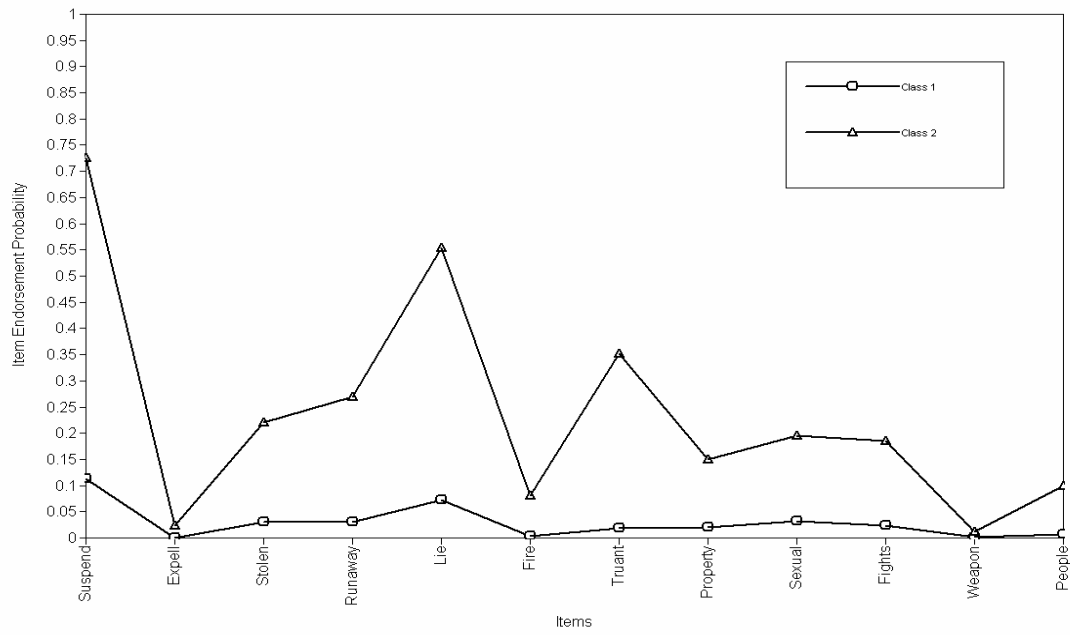
classes.

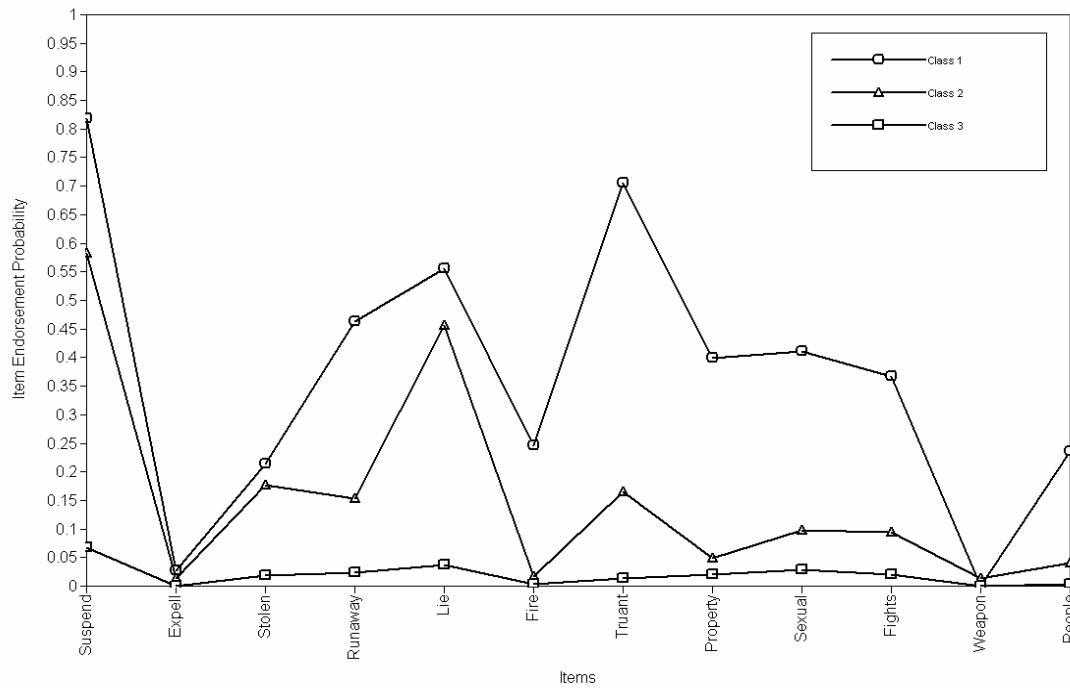*Figure 4.6.* Factor Mixture Model & ACE on Factors Diagram.

a.

b.

Figure 4.7. a. Two-Class LCA Profile Plot. b. Three-Class LCA Profile Plot.
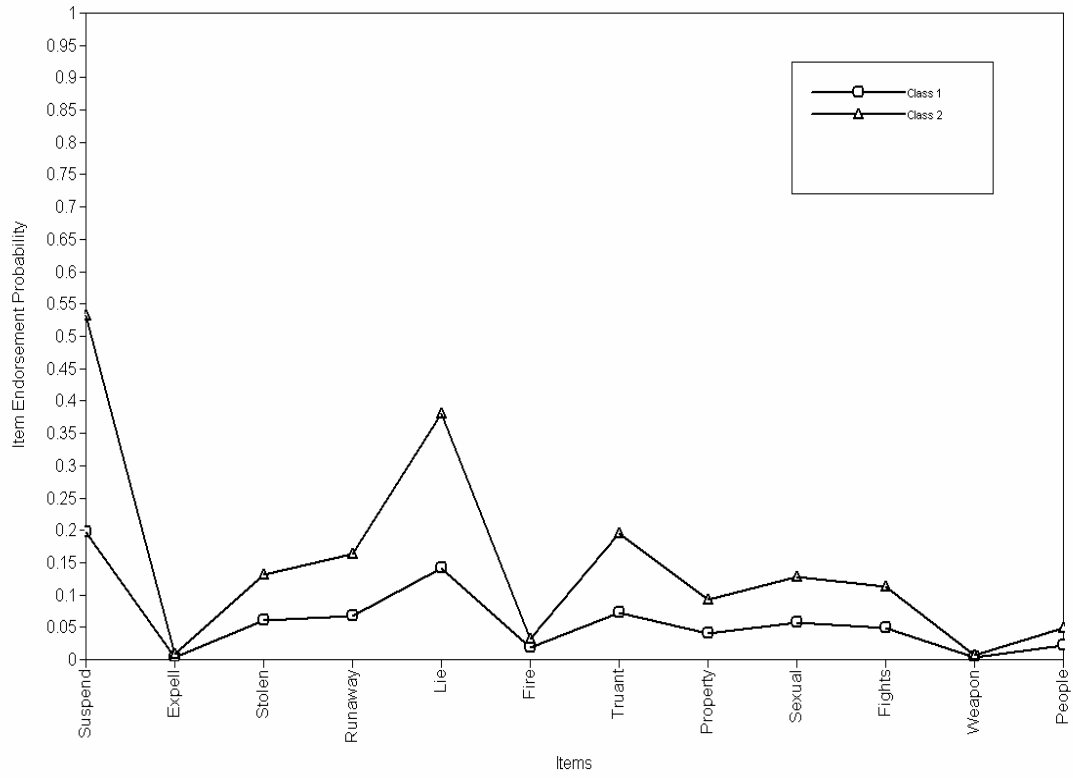
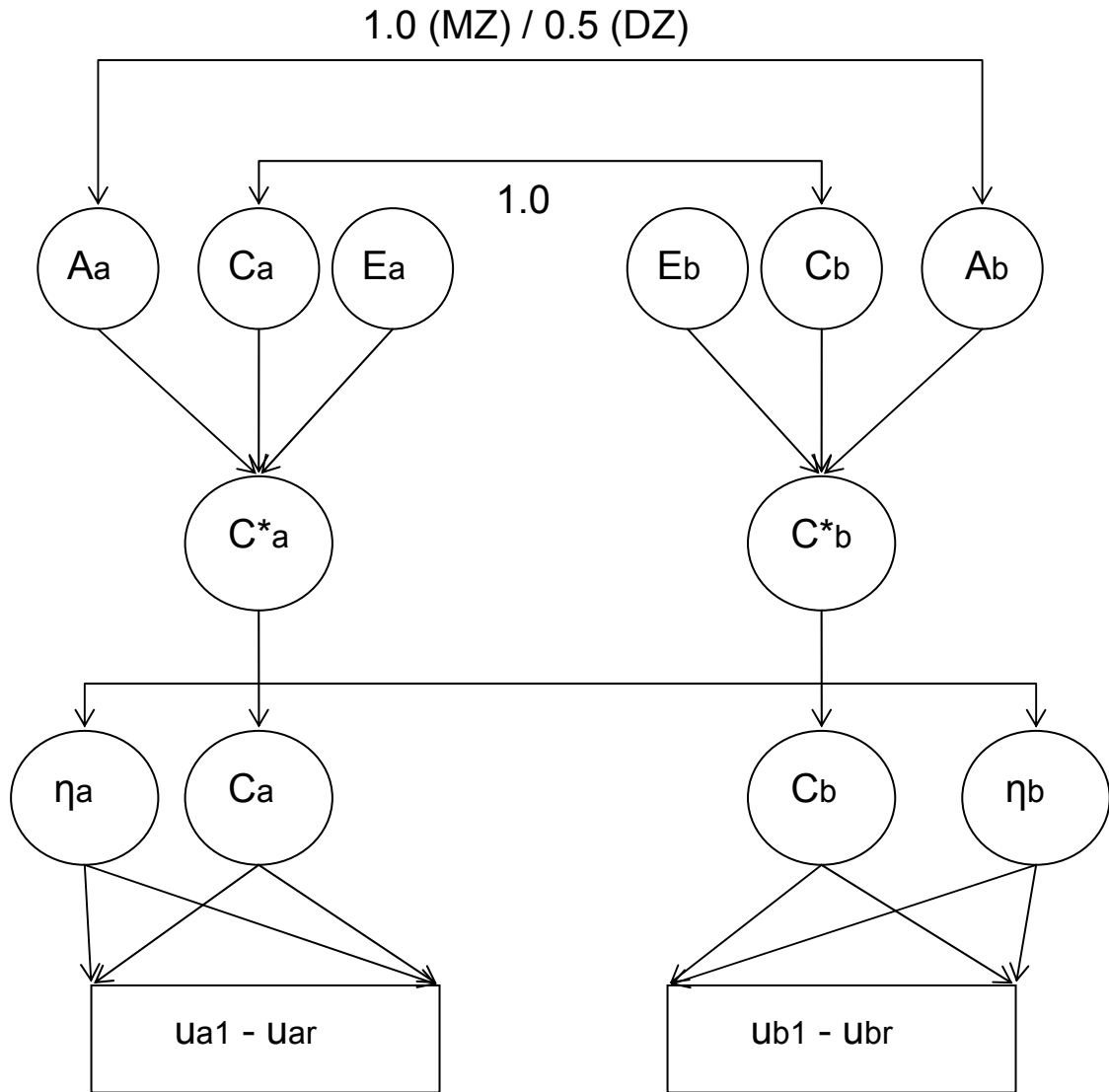*Figure 4.8.* Two-Class, One-Factor FMM Profile Plot

Figure 4.9. FMM & ACE on the Latent Classes Model Diagram.

Appendix C: Mplus code for Chapter 4

Example Mplu*s* Input and Model Statements

Input 1: Twin latent class analysis (LCA) model

VARIABLE: Names are ua1-ua6 ub1-ub6;

      Usevariable are ua1- ub6;

      Categorical are ua1- ub6;

      Classes = ca(2) cb(2);

ANALYSIS: Type = Mixture;

      Starts = 50 5;

      Parameterization = Loglinear;

MODEL: %Overall%

    ca#1 WITH cb#1;

    [ca#1] (901);

    [cb#1] (901);

MODEL ca: %ca#1%

      [ua1$1- ua6$1] (101-106);

     %ca#2%

      [ua1$1- ua6$1] (107-112);

MODEL cb: %cb#1%

      [ub1$1- ub6$1] (101-106);

%cb#2%

    [ub1$1- ub6$1] (107-112);

Input 2: Twin factor analysis (FA) model

ANALYSIS:

    Estimator = ML;

MODEL:

    eta_a BY ua1@1

        ua2-ua6 (2-6);

    eta_b BY ub1@1

        ub2-ub6 (2-6);

    [eta_a - eta_b@0];

    eta_a - eta_b *1 (100);

    eta_a WITH eta_b;

    [ua1$1- ua6$1] (101-106);

    [ub1$1- ub6$1] (101-106);

Input 3: Twin factor mixture model (FMM) with class invariant item thresholds and factor loadings

VARIABLE: Classes = ca(2) cb(2);

ANALYSIS: Type = Mixture;

    Algorithm = Integration;

Starts = 500 50;

Parameterization = Loglinear;

MODEL: %Overall%

ca#1 WITH cb#1;

[ca#1] (801);

[cb#1] (801);

eta_a BY ua1-ua6;

eta_b BY ub1-ub6;

eta_a - eta_b;

%ca#1.cb#1%

eta_a WITH eta_b;

%ca#1.cb#2%

eta_a WITH eta_b (999);

%ca#2.cb#1%

eta_a WITH eta_b (999);

%ca#2.cb#2%

eta_a WITH eta_b;

MODEL ca: %ca#1%

eta_a BY ua1@1

ua2-ua6 (2-6);

[eta_a @0];

eta_a (901);

[ua1$1- ua6$1] (101-106);

%ca#2%

eta_a BY ua1@1

ua2-ua6 (2-6);

[eta_a] (900);

eta_a (902);

[ua1$1- ua6$1] (101-106);

MODEL cb: %cb#1%

eta_b BY ub1@1

ub2-ub6 (2-6);

[eta_b @0];

eta_b (901);

[ub1$1- ub6$1] (101-106);

%cb#2%

eta_b BY ub1@1

ub2-ub6 (2-6);

[eta_b *1] (900);

eta_b *1 (902);

[ub1$1- ub6$1] (101-106);

Input 4: Twin LCA & ACE model

VARIABLE: Names are ua1-ua6 ub1-ub6 zyg;

      Usevariable are ua1- ub6;

      Categorical are ua1- ub6;

      Classes = cg(2) ca(2) cb(2);

      Knownclass = cg (zyg=2 zyg=1);

ANALYSIS: Type = Mixture;

      Algorithm = Integration;

      Starts = 500 50;

      Reciprocalinteraction = Nocheck[1];

MODEL: %Overall%

      Cstar_a BY ;

      Cstar_b BY ;

      ca#1 ON Cstar_a @0 Cstar_b @0;

      cb#1 ON Cstar_a @0 Cstar_b @0;

      [Cstar_a - Cstar_b @0];

      [ca#1] (801);

      [cb#1] (801);

      Cstar_a - Cstar_b (v);

MODEL cg: %cg#1%

      ca#1 ON Cstar_a @1 Cstar_b @0;

---

[1] Special feature available in the forthcoming Mplu*s* V6.0.

cb#1 ON Cstar_a @0 Cstar_b @1;

Cstar_a WITH Cstar_b (cdz);

%cg#2%

ca#1 ON Cstar_a @1 Cstar_b @0;

cb#1 ON Cstar_a @1 Cstar_b @0;

Cstar_a WITH Cstar_b @0;

MODEL ca: %ca#1%

    [ua1$1- ua6$1] (101-106);

 %ca#2%

    [ua1$1- ua6$1] (101-106);

MODEL cb: %cb#1%

    [ub1$1- ub6$1] (101-106);

 %cb#2%

    [ub1$1- ub6$1] (101-106);

 MODEL CONSTRAINT: New(a c h);

    v= a**2 + c**2;

    cdz = 0.5*a**2 + c**2;

    h = a**2 / (a**2 + c**2 + (3.14**2/3));

Note. In the first two lines of the class specific statements in the model cg section,

%cg#1% and %cg#2%, is a way to reduce the number of dimensions of integration in

order to avoid having two factors correlated at one, which can be problematic when using

numerical integration (Algorithm = Integration). In DZ twins, %cg#1%, only the A

component is correlated at one across twins, but in MZ twins, %cg#2% both the A and C

components are correlated at one across twins. In the MZ twin group, both latent classes,

Ca and *Cb*, are regressed on only *C*a*. Because *C*a* and *C*b* are statistically equivalent

because of the imposed measurement invariance, this avoids having a factor correlated at

one.

Input 5: Twin FA and ACE model

VARIABLE:

     Classes = cg(2);

     Knownclass = cg (zyg=2 zyg =1);

ANALYSIS: Type = Mixture;

     Algorithm = Integration;

MODEL: %Overall%

     eta_a BY ua1@1

         ua2-ua6 (2-6);

     eta_b BY ub1@1

         ub2-ub6 (2-6);

     [eta_a - eta_b @0];

     eta_a - eta_b (var);

     [ua1$1- ua6$1] (101-106);

     [ub1$1- ub6$1] (101-106);

%cg#1%

  eta_a WITH eta_b (covdz);

  %cg#2%

  eta_a WITH eta_b (covmz);

MODEL CONSTRAINT: NEW(a c e h);

  var = a**2 + c**2 + e**2;

  covmz = a**2 + c**2;

  covdz = 0.5*a**2 + c**2;

  h = a**2 / (a**2 + c**2 + e**2);

Input 6: Twin FMM with class invariant item thresholds and factor loadings & ACE
model on factors

VARIABLE: Classes = cg(2) ca(2) cb(2);

        Knownclass = cg (zyg=2 zyg=1);

ANALYSIS: Type = Mixture;

        Algorithm = Integration;

        Integration = 50;

        Starts = 500 50;

        Parameterization = Loglinear;

MODEL: %Overall%

        eta_a by ua1@1

                ua2-ua6 (2-6);

eta_b by ub1@1

ub2-ub6 (2-6);

eta_a - eta_b (v);

[ca#1] (801);

[cb#1] (801);

cg#1 WITH ca#1 (802);

cg#1 WITH cb#1 (802);

cb#1 WITH ca#1@0;

%cg#1.ca#1.cb#1%

eta_a WITH eta_b (cdz);

[eta_a - eta_b @0];

[ua1$1- ua6$1] (101-106);

[ub1$1- ub6$1] (101-106);

%cg#1.ca#1.cb#2%

eta_a WITH eta_b (cdz);

[eta_a @0];

[eta_b *] (903);

[ua1$1- ua6$1] (101-106);

[ub1$1- ub6$1] (101-106);

%cg#1.ca#2.cb#1%

eta_a WITH eta_b (cdz);

[eta_a *] (903);

[eta_b @0];

[ua1$1- ua6$1] (101-106);

[ub1$1- ub6$1] (101-106);

%cg#1.ca#2.cb#2%

eta_a WITH eta_b (cdz);

[eta_a - eta_b *] (903);

[ua1$1- ua6$1] (101-106);

[ub1$1- ub6$1] (101-106);

%cg#2.ca#1.cb#1%

eta_a WITH eta_b (cmz);

[eta_a - eta_b @0];

[ua1$1- ua6$1] (101-106);

[ub1$1- ub6$1] (101-106);

%cg#2.ca#1.cb#2%

eta_a WITH eta_b (cmz);

[eta_a @0];

[eta_b *] (903);

[ua1$1- ua6$1] (101-106);

[ub1$1- ub6$1] (101-106);

%cg#2.ca#2.cb#1%

eta_a WITH eta_b (cmz);

[eta_a *] (903);

[eta_b @0];

[ua1$1- ua6$1] (101-106);

[ub1$1- ub6$1] (101-106);

%cg#2.ca#2.cb#2%

eta_a WITH eta_b (cmz);

[eta_a - eta_b *] (903);

[ua1$1- ua6$1] (101-106);

[ub1$1- ub6$1] (101-106);

MODEL CONSTRAINT: New(a c e h);

v= a**2 + c**2 + e**2;

cdz = 0.5*a**2 + c**2;

cmz = a**2 + c**2;

h = a**2 / (a**2 + c**2 + e**2);

References

Clark, S.L., Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. Manuscript submitted for publication.

Clark, S.L., Muthén, B., Kaprio, J., D'Onofrio, B., Viken, R., Rose, R., & Smalley, S.L. (2009). Models and strategies for factor mixture analysis: Two examples concerning the structure underlying psychological disorders. Manuscript submitted for publication.

D'Onofrio, B.M., Rathouz, P.J., & Lahey, B.B. (2009). The importance of understanding gene-environment correlations in the development of antisocial behavior. In K. Kendler, S. Jaffee, D. Romer (Eds.), *The dynamic genome and mental health: The role of genes and environments in development.* Forthcoming.

Dwyer, C.A. (1996). Cut scores and testing: Statistics, judgment, truth and error. *PsychologicalAssessment*, 8(4), 360-362.

Eaves, L.J., Silberg, J.L., Hewitt, J.K., Ruter, M., Meyer, J.M., Neale, M.C., & Pickles, A. (1993). Analyzing twin resemblance in multisymptom data: Genetic applications of latent class model for symptoms of conduct disorder in juvenile boys. *Behavior Genetics*, 23, 5-19.

Falconer, D.S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics*, 29, 51-76.

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

Kaprio, J., Pulkkinen, L., & Rose, R.J. (2002). Genetic and environmental factors in

health-related behaviors: Studies on Finnish twins and twin families. *Twin Research*, 5, 366-371.

Kaprio, J., Rimpela, A., Winter, T., Viken, R.J., Rimpela, M., & Rose, R. (1995). Common genetic influences on BMI and age at menarche. *Human Biology*, 67, 739-753.

Kuperman, S., Schlosser, S. S., Kramer, J. R., Bucholz, K. K., Hesselbrock, V., Reich, T., et al. (2001). Developmental sequence from disruptive behavior diagnosis to adolescent alcohol dependence. *American Journal of Psychiatry*, 158, 2022-2026.

Lahey, B. B., & Waldman, I. D. (2003). A developmental propensity model of the origins of conduct problems during childhood and adolescence. In B. B. Lahey, T. E. Moffitt & A. Caspi (Eds.), *Causes of conduct disorder and juvenile delinquency* (pp. 76-117). New York, NY: Guilford Press Guilford Press.

Lahey, B. B., & Waldman, I. D. (2005). A developmental model of the propensity to offend during childhood and adolescence. In D. P. Farrington (Ed.), *Advances in criminological theory, Volume 13* (pp. 15-50). Piscataway, NJ: Transaction Publishers.

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

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674-701.

Moffitt, T. E. (2006). Life-course persistent and adolescence-limited antisocial behavior. In D. Cicchetti & D. Cohen (Eds.), *Developmental Psychopathology, 2nd Edition* (pp. 570-598). New York: Wiley.

Muthén, B. (2001) Latent variable mixture modeling. In G.A. Marcoulides & R.E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp.1-33). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Muthén, B., Asparouhov, T., & Rebollo, I. (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. *Twin Research and Human Genetics*, 9(3), 313-324.

Muthén, L.K., & Muthén, B.O. (2005). Statistical analysis with latent variables using Mplus workshop. Alexandria, Va.

Muthén, L.K., & Muthén, B.O. (2008). Mplus (Version 5.21). Los Angeles, CA.

Neale, M.C., & Cardon, L.R. (1992) *Methodology for genetic studies of twins and families.* Kluwer Academic Publishers, Dordrecht.

Prescott, C.A. (2004). Using the Mplus computer program to estimate models for continuous and categorical data from twins. *Behavior Genetics*, 34, 17-40.

Rasmussen, E.R., Neuman, R.J., Heath, A. C., Levy, F., Hay, D.A., & Todd, R.D. (2002).

Replication of the latent class structure of Attention-Deficit/Hyperactivity

Disorder (ADHD) subtypes in a sample of Australian twin. *Journal of Child

Psychology and Psychiatry*, 43, 1018-1028.

Rose, R.J., Dick, D.M., Viken, R.J., Pulkkinen, L., & Kaprio, J. (2001). Drinking or

abstaining at age 14? A genetic epidemiological study. *Alcoholism: Clinical and

ExperimentalResearch*, 25, 1594-1604.

Rose, R.J., Dick, D.M., Viken, R.J., Pulkkinen, L., Nurnberger, J.I., Jr., & Kaprio, J.

(2004). Genetic and environmental effects on conduct disorder, alcohol

dependence symptoms, and their covariation at age 14. *Alcoholism: Clinical and

Experimental Research*, 28, 1541-1548.

Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review*, 34, 278-

286.

Wright, S. (1934). An analysis of variability in number of digits in an inbred strain of

guinea pigs. *Genetics*, 19, 506-536.

Chapter Five: Discussion and Future Directions

This section contains a summary of the three papers presented in this dissertation and discusses the application of the methods utilized in each paper in randomized control trials (RCTs). It also provides a roadmap for future methodological work in combining latent variable and genetic models.

Chapter Two: Relating latent class analysis results to variables not included in the analysis

The first paper investigated how the method chosen to incorporate covariates into a mixture model can impact the estimates and standard errors in a regression or the results of a mean comparison test. One issue examined in this paper that is related to the incorporation of covariates into a mixture model is whether the latent class variable can be treated as an observed, exact variable. The first part of the study explored two real data examples to demonstrate the problem of treating class membership as an observed variable and to show how incorrect the estimates and standard errors can be when including many auxiliary variables. The second part of the study consisted of Monte Carlo simulations to explore under what conditions the results from the real data examples occur. The real data examples showed no differences among the methods used in the mean comparison tests. For the regression results, most likely class membership regression, probability-weighted regression, and pseudo-class regression performed similarly; but in comparison to including covariates while forming the latent classes, the other methods tended to underestimate the covariate effects and the standard errors.

Similar to the real data examples, the mean comparison Monte Carlo simulations showed no differences between the methods used, even when the strength of the relationship between the latent class and covariate was examined. The regression simulation results for when there was no relationship between the latent class variable and the covariate suggested that all the regression methods were able to recover the true effect of zero. But, when compared to including the covariate while forming the latent classes, the other regression methods tended to have underestimated standard errors. The regression simulation results for when there was a strong, positive relationship between the latent class and covariate showed that the different regression methods produced estimates of the covariate effect that were smaller than the true value, with the exception of incorporating the covariate while forming the latent classes. As the entropy of the latent classes decreased, the covariate effect estimates of the different methods got further away from the true value.

Based on the results of the real data examples and the simulation study, recommendations were made about when it is appropriate to use which regression method. When it is an option, incorporating the covariates while forming the latent classes is best. But, many times this is not feasible because as the number of classes increases, the computation time of the model also increases. Additionally, covariates can impact the formation and interpretation of the latent classes. If incorporating the covariates while forming the latent classes is not an option, then one alternative is to use most likely class membership, but only when the entropy is high. Additional

recommendations were made about how to select which covariates to include in an analysis when there are a large number from which to choose.

The issue of how to incorporate covariates into mixture models is important when evaluating the effectiveness of a RCT. By including covariates into an analysis, rival hypotheses, which might also explain why a RCT is effective, can be ruled out. As this first paper has shown, the method chosen to incorporate the covariate into a mixture model can affect the estimation of the covariate effect, which in turn, can impact the ability to correctly rule out alternative explanations for the results in an RCT.

Chapter Three: Models and strategies for factor mixture analysis: Two examples concerning the structure underlying psychological disorders

The second paper used a factor mixture model (FMM) to explore the underlying structure of psychological outcomes and explicates the FMM in more detail than has been given in previous research. Because the FMM has both latent categorical and continuous variables, the FMM has several advantages over other models used to represent the underlying structure of psychopathology. Unlike the categorical representation of psychopathology, which can be modeled using a latent class analysis (LCA) model, the FMM is able to account for the severity within categories through the use of a latent continuous variable. And unlike the dimensional view of psychopathology, which can be modeled using a factor analytic (FA) model, the FMM has a latent categorical variable that provides a model-based classification system to group individuals.

This paper introduced five variations of the FMM which differ based on the amount of measurement invariance present and discussed the impact measurement

invariance has on the interpretation of the different FMM variations. Also, an ad hoc method for detecting measurement invariance of the item thresholds was presented. This paper also discussed more practical issues of how to build a FMM in practice including a discussion on factor measurement structure options and a suggested strategy for determining the combination of latent categorical and latent continuous variables.

The second paper also explored the underlying structure of conduct disorder (CD) and attention-deficit hyperactivity disorder (ADHD) by fitting LCA, FA, and FMM models to the example data. Each of these examples also elucidated some practical issues that can occur when fitting FMM to data. In the CD example, a FMM-2, which implies that the underlying structure of CD is continuous, but non-normal, was found to be best fitting model for the data. This example highlighted how challenging it can be to compare among different model types, with each implying something different about the underlying structure. In the ADHD example, a four-factor FA solution was the best fitting model for the data. This suggests that even though the FMM is flexible enough to allow for the underlying structure to be simultaneously categorical and dimensional there are still cases where there may not be a need for this flexibility.

The use of FMMs in RCT analysis is not prevalent. As mentioned in the second paper, this may because there is a dearth in the literature about how FMMs should be applied in practice and once a well-fitting model is obtained, how it should be interpreted. Through the demonstration of fitting FMMs to the two example datasets and interpreting the solutions, this paper sought to remedy this gap. It is important the FMM be explored as a possible representation for the underlying structure of behavioral outcomes in RCTs

187

because the flexibility of the FMM allows for the underlying structure to be simultaneously categorical and dimensional. Using a FMM as the outcome in a RCT might aid in determining whether the RCT has a bigger impact on qualitative differences in the outcome, represented by the latent classes, or on the severity, represented by the latent factors.

Chapter Four: ACE Analysis for Latent Variable Phenotypes

The third paper explored the use of two latent variable models as phenotypes in an ACE analysis. The first was the LCA model, which uses a categorical latent variable to account for the unobserved heterogeneity in a sample. This paper proposed a single-step method for combining LCA and ACE analysis by using the latent class variable as the phenotype rather than a three-step ad hoc method with most likely class membership as the phenotype. In the LCA and ACE combination, a liability threshold model is used on the latent class variable.

The second was the FMM with measurement invariance which uses a combination of both latent categorical and continuous variables. The presence of measurement invariance in the FMM ensures that the factor is measured the same way throughout the entire sample. This makes the factor an ideal candidate for the placement of the ACE model in the FMM. Alternative methods for combining the FMM and ACE when there is no measurement invariance were also presented, but not applied to the data.

The use of LCA, FMM, and, for comparison, FA as phenotypes was demonstrated through an example concerning conduct disorder (CD) in a sample of Finnish twins. The LCA solution had two classes: a large class consisting of likely asymptomatic individuals

and a smaller class consisting of likely symptomatic individuals. The FA solution had one factor with all items except one, loading significantly on the factor suggesting that it was a general CD factor. The FMM solution had two classes and one factor. Similar to LCA solution, there was one large class consisting of individuals that had a small probability of having any CD symptoms and a smaller class of individuals with a moderate probability of endorsing some of the CD symptoms. When an ACE model was applied to each of these latent variable models, the heritability was above 0.70 in each model suggesting that more than 70% of the variation in the phenotype is attributable to genetic variation.

This third paper reflects what should be the first steps in incorporating genetics into RCTs: defining the phenotype of interest and determining if it is heritable. The latent variable models discussed in this paper can be used as an alternative to traditional phenotypes for behaviors, such as affection status or a sum of the number of symptoms present, because they provide advantages. One advantage is that more pieces of information can be used to inform phenotype definition. A second advantage is that because multiple pieces of information are used there is less measurement error associated with the phenotype. This paper shows how these latent variable phenotypes can be combined with an ACE analysis to determine if the behavior of interest is heritable.

Future Research

Each of the individual papers concluded with a discussion about future research that should be conducted in order to advance the understanding of the methods presented

189

in the papers. This section discusses future research possibilities for the application of these methods to issues specific to RCTs and the incorporation of genetic information.

The papers in this dissertation focused specifically on methods for cross-sectional models, but the goal of RCTs is often to evaluate the long-term impact of an intervention. If the goal is to evaluate the effect of an intervention at a later time point, longitudinal models need to be considered. The two most common types of longitudinal models that can be used in RCT analysis are growth curve and auto regressive models. Growth curve models describe the general rate at which an outcome of interest changes. Auto regressive models describe the probability of transitioning from one state to another. For example, an individual can transition between being affected or not affected by depression over time.

Both of these types of longitudinal models have latent variable counterparts that can be used as phenotypes in a genetic analysis. In growth curve modeling, the intercept and slope can be latent continuous variables, called growth factors. In an ACE model, for example, the latent growth factors from a growth curve model can then be used as the phenotype, as was done in Boomsma et al. (2007). Growth mixture modeling (GMM) adds a latent class variable, indicated by the latent growth factors, that groups individuals based on their observed growth trajectories. In the RCT context, GMM aids in exploring differential treatment effects in classes of people that differ on an outcome of interest (Muthén et al., 2002). In GMM there are two possible phenotypes, the latent growth factor or the latent classes of trajectories. Similar to LCA, if the latent classes are used as the phenotype in an ACE ananlysis, the latent class variable needs to be converted into a

190

latent continuous variable using the liabitlity threshold model. In Kerner and Muthén (2009), the classes from a GMM are used in a genome wide association study of systolic blood pressure.

Auto regressive models also have latent variable counterparts. Latent transition ananlysis (LTA) uses a latent class measurement model at each time point and a stuctural component that models change across groups over time (Nylund, 2007). Another latent variable auto regressive model is the FMM-LTA. Suggested by Muthén (2008), the FMM-LTA has a FMM as the measurement model and, as in LTA, a structural component is used to model change over time. When using these tranisition models as phenotypes in an ACE analysis, there are several options for where to place the ACE variance components and, for simplicity, are discussed in the LTA context. One place for variance components is on the latent class variable at each time point, which would provide estimates of the heritability at each time point, but not for the overall period of interest. Another option would be to have one ACE model influence all the latent class variables so that there is only one heritability. A third option is to combine the previous two so that there is an overall ACE model and one on each latent class variable, similar to an independent pathways model. One issue to keep in mind if the ACE model is placed on the latent class variable at each time point is that each latent class variable would need to be converted into a latent continuous one by the liability threshold model. If there are a large number of times points, this can be problematic because each time point would require its own latent variable. And with each latent variable the model would become more complex and require more computation time.

Another extension to the models in these papers would be to include gene by environment (GxE) interactions. In RCTs, the intervention can serve as a moderator for gene expression. In behavioral genetic analysis, GxE interaction models with latent variable phenotypes commonly use a FA model as the phenotype. For example, Harden et al. (2006) use a GxE model with a factor measuring cognitive ability as the phenotype to assess impact of parental education and income. An extension would be to use mixture models, such as the LCA and FMM, as phenotypes in a GxE analysis. The ACE models presented in paper three can easily be extended to include GxE interactions. In the LCA case, for example, equation 4 is extended to include a gene by treatment interaction by the following equation:

$$C^*_{ij} = t\text{Treatment}_{ij} + (a + a'\,\text{Treatment}_{ij})A_{ij} + (c + c'\,\text{Treatment}_{ij})C_{ij} + eE_{ij}\,, \qquad (1)$$

where C^*_{ij} is the latent continuous variable transformation of the latent categorical variable C, t is the effect of treatment status, Treatment_{ij}, and A_{ij}, C_{ij}, and E_{ij} are the latent additive genetic, shared environment, and unshared environment components.

There are many other modeling possibilities than what is presented here. This dissertation establishes a foundation for the conceptualization of how to incorporate mixture models and genetic information in randomized control trials.

References

Boomsma, D., Cacioppo, J., Muthén, B., Asparouhov, T. & Clark, S. (2007). Longitudinal genetic analysis for loneliness in Dutch twins. *Twin Research and Human Genetics*, 10, 267-273.

Harden, K.P., Turkheimer, E. & Loehlin, J.C. (2006). Genotype by environment interaction in adolescents' cognitive aptitude. *Behavioral Genetics*. Kerner, B. & Muthén, B. (2009). Growth mixture modeling in families of the Framingham Heart Study. Forthcoming in special Supplement edition of *BioMed Central*.

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

Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S., Carlin, J., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, 3, 459-475.

Nylund, K. (2007). Latent transition analysis: Modeling extensions and an application to peer victimization. Doctoral dissertation, University of California, Los Angeles.