Spring 2016

# Examining latent change classes: An application of factor mixture modeling to change scores

Thai Q. Ong
*James Madison University*

Examining Latent Change Classes:

An Application of Factor Mixture Modeling to Change Scores

Thai Quang Ong

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

May 2016

---

FACULTY COMMITTEE:

Committee Chair: Monica K. Erbacher

Committee Members/ Readers:

Deborah L. Bandalos

Sonia J. Host

Acknowledgements

I would like to thank my committee chair and advisor, Dr. Monica K. Erbacher. Without you, this thesis would have not been written or completed. I know you truly believe in me as a researcher and student and for that, I am honored. You accepted me for who I am and provided me with a safe environment to conduct research in. I don’t know what I did to deserve such an amazing mentor like you. Good luck in Arizona and I can’t wait to meet up with you under the hot Arizona sun.

I would also like to thank the other members of my thesis committee, Dr. Debbi Bandalos and Dr. Jeanne Horst. Both of you have provided me with tremendous support with my thesis and in a number of areas. Debbi, your feedback for my thesis was invaluable. I look forward to working with you more in the doctoral program. Jeanne, you have been there for me since day one. Even at times when I struggled in the program, you were there to help pick me back up.

I would like to thank my good friends, Kristen Smith, Liz Smith, Scott Strickman and, Courtney Sanders. I don’t know if I could have made it through this without the support from each and every one of you. Thank you for allowing me to be me.

I would like to thank Gregory Ryan Brown. We have been through so much together these last four years. Your support has helped me greatly, especially during the last three months. I can’t thank you enough for everything you have done for me. I look forward to spending the summer with you abroad. I love you GRB.

Last but not least, I would like to thank my parents and sister. Mom and dad, you are the hardest working people I know. You’ve built a life for me in America and allowed

me to pursue the American dream. Both of you have sacrificed so much for me and I can’t express in words how much that means to me.

Table of Contents

List of Tables

List of Figures

Abstract

Although change scores are used in a variety of statistical methods (e.g., analysis of variance and regression), there is a lack of application of latent variable modeling methods to change scores. This thesis provides a detailed description of two latent variable modeling methods applied to change scores: factor analysis of change scores and change score factor mixture modeling. To illustrate advantages of these methods, both were applied to change score data from undergraduates. Students responded to sense of identity items during a university-wide assessment day on two occasions, once as incoming freshmen and again as second-semester sophomores. Change scores were computed by subtracting sophomore item responses from freshmen item responses. Factor analysis results indicated sense of identity change scores were best represented by two factors, change in sense of self and purpose and development of morals and beliefs. Factor mixture modeling results suggested two latent classes underlying these factors. The classes differed in both factor means and factor variances, which implied two possible change patterns associated with development of sense of identity. One class contained students who were stable on the two change score factors (i.e. changed minimally on sense of self and purpose and morals and beliefs) and the other class contained students who were fluid on one of the two factors. Classes were somewhat replicated with a second, independent sample, in that two classes were detected, but class means and variances diverged from those in the first sample. Results across the two methods provided insightful information about change processes of sense of identity, particularly how development of sense of identity is not the same across students. The

applied example highlights the advantages of applying these methods to change scores. Implications of the two methods are further discussed throughout the thesis.

# CHAPTER ONE

## Introduction

### Overview

Given the field of psychology's focus on individual growth and development across time, longitudinal data have become an increasingly popular demand. While cross-sectional data provide information about a sample at one specific time point, only longitudinal data allow researchers to examine change or growth. For instance, with cross-sectional data, the researcher is only able to examine one level of sample means, variances, and covariances. In contrast, longitudinal data may have multiple levels of means, variances, and covariances (Biesanz, West, & Kwok, 2003), which allows researchers to compare estimated parameters of interest across three different levels: between groups, between individuals (interindividual), and/or within individuals (intraindividual, depending on the number of occasions). As such, it is not surprising longitudinal data can provide an abundance of additional information compared to cross-sectional data.

Although a variety of models have been applied to longitudinal data, the concept of measuring and representing quantitative change has been a controversial topic in behavioral science (e.g., Bohmstedt, 1969; Burr & Nesselroade, 1990). Even the simplest form of change analyses (such as change scores) have been among the most controversial topics in terms of usefulness and reliability (Cronbach & Furby, 1970; Williams & Zimmerman, 1977). Despite this, change analyses have been useful in research fields such as psychology and higher education (e.g., Allison, 1990; Culpepper, 2014). Longitudinal researchers using change analyses are often interested in making statements about individuals in the population (e.g., how do people differ in their patterns of change across time?). The focus of these studies tends to be on how individuals' responses change over a period of time, rather than about the state of individuals at a single occasion. Change can be complex (i.e., affected by a variety of factors), and individuals may differ in patterns of change across occasions. Thus, individuals' patterns of responses should be examined holistically across occasions.

One way to explore this is through person-centered approaches, in which patterns of responses across variables at the individual level are examined. Person-centered approaches are useful for longitudinal data because research questions are often framed in terms of individuals and patterns of change rather than in terms of variables. Along with person-centered approaches, change scores can prove particularly useful in educational and psychological research fields.

## The Usefulness of Change Scores in Educational and Psychological Measurement

As a consequence of an increased focus on change and development, analyses on change scores have also become popular, particularly in exploring development across two occasions. Change scores represent the observed change of an individual on a construct or item. Each individual has a calculated change score, representing growth or decline in a construct. Change scores are easy to calculate, interpret, and provide ample information at the individual and group levels. For example, a single change score represents an individual's change on a construct across two occasions. Alternatively, the mean of change scores represents the average amount of change for a sample on a construct. Thus, even when examined alone, inferences about change at the individual and group level may be made.

The inclusion of change scores in statistical models (such as regression or ANOVA) has several implications. It enables researchers to explore how change predicts or relates to other variables. Researchers are able to empirically test whether other variables are related to change or predictive of change. In most models, change scores on one construct are treated as a single variable and are easily entered into the statistical model of interest. Practitioners unfamiliar with sophisticated latent variable longitudinal models use change scores in simpler models (e.g., regression) to make inferences about change. For these reasons, the calculation and application of change scores is useful for those interested in exploring change and its characteristics. For example, predicting change in GPA would allow a researcher to identify covariates of academic growth, which is distinct from achievement alone.

Despite an increased focus on change and development, additional methods of analyzing longitudinal data are still needed. For example, one method often under-utilized is direct analysis of change scores, particularly in latent variable models. Where factor analytic methods are used to explore underlying latent dimensions, the same factor analytic methods are applicable to change scores. Factor analyzing change scores can provide novel information about the change process of a construct. Importantly, the underlying dimensions of change may not be identical to dimensions underlying cross-sectional scores. Thus, dimensions of change should be examined independently rather than assumed. In doing so, researchers may uncover aspects of change associated with a construct that differ from cross-sectional dimensions, allowing for a more comprehensive understanding of the construct and associated developmental processes.

Although latent variable models have been proposed to examine latent change (e.g., latent change models and latent growth curve models), dimensions of change in many of these models are assumed to parallel the cross-sectional dimensions of a construct. Although this assumption may hold for certain constructs, it is not always true for complex constructs studied in educational and psychological research (e.g., see Nesselroad & Cable, 1974). For these constructs, the change processes may be particularly complicated. Directly analyzing change scores enables researchers to examine aspects of the change processes within a construct. Even if dimensions of change are the same as the cross-sectional dimensions, change score factors may provide unique information about change processes and their covariates that is otherwise unobtainable. For example, change score factors may relate to outcomes differently compared to cross-sectional factors. Moreover, whereas other latent change models assume dimensions of change to be the same as cross-sectional dimensions, factor analyzing change scores allows researchers to test this assumption. Thus, factor analysis of change scores may be used to help inform other latent change models.

**Two Approaches to Analyzing Longitudinal Data**

Variable-centered and person-centered approaches to analyzing longitudinal data have emerged alongside the development of longitudinal models. A variable-centered approach focuses on the aggregate relationships among variables rather than the individuals (Magnusson, 2003) and examines how specific variables relate to one another. Thus, research questions are framed in terms of variables, not individuals. For example, a research question about sense of identity as a predictor of GPA is a variable-centered research question. Researchers typically take the variable-centered approach because the variables collected are often the primary interests of the research study. Using this approach, it is often assumed relationships between variables are the same for everyone in the population. In contrast, a person-centered approach focuses on individuals and their patterns of responses. Thus, the goal of a person-centered approach is to identify different groups of individuals who share similar response patterns across variables. For example, a research question about different patterns of change in aspects of sense of identity is a person-centered research question. These approaches, however, should not be viewed as competing approaches to analyzing longitudinal data (Magnusson, 2003). Different longitudinal research questions often require a variety of approaches. The selection of which approach(es) to take should depend on the research question.

The variable- and person-centered approaches make different assumptions about the data. Whereas the variable-centered approach assumes the population of interest is homogenous, the person-centered approach assumes the population of interest is heterogeneous (Bergman & Magnusson, 1997; Laursen & Hoff, 2006). The latter is often the case with longitudinal research. Participants may appear similar in their levels on a construct at one occasion, but when examined holistically across two or three occasions, a variety of patterns of change may emerge for individuals in the population. Thus, the population is heterogeneous because individuals do not exhibit the same pattern of change across time. In this case, person-centered approaches are useful because they can help 1) identify patterns of change across time, and 2) classify individuals based on their patterns of change. In contrast to a variable-centered approach, in which individuals are assumed to have similar change trajectories over time, the person-centered approach allows for individual differences in patterns of change over time (Laursen & Hoff, 2006).

**A Person-Centered Approach to Analyzing Change Scores**

The application of person-centered analyses to change scores may be useful in identifying people with similar change patterns across time. One person-centered approach is known as mixture modeling. Mixture models are used to explore latent classes (i.e., unobserved groups) within a population. Mixture modeling assumes the population of interest consists of underlying sub-populations and is therefore heterogeneous (Pastor & Gagné, 2013). Although mixture modeling has been applied to cross-sectional scores, there has been limited research, if any, on the application of mixture modeling to factor analysis of change scores.

The application of mixture modeling to change score factors carries the implications of further understanding development and change in a construct. Whereas factor analysis of change scores allows one to examine underlying dimensions of the change process, the application of mixture modeling to change score factors (i.e., change score factor mixture modeling) allows one to examine latent change classes underlying change score factors. If change classes are identified, it provides evidence for individual differences in patterns of change across time in a construct. Further, the identification of change classes and their characteristics may be important for identifying correlates of change processes. For example, latent change classes may differ systematically on a number of characteristics. If that is the case, these characteristics may be influential to the change process of a construct and thus should be examined when studying the construct. Additionally, classes differing in their patterns of change may have different outcomes. A class with a particular pattern of change may have additional positive outcomes relative to a class with differing patterns of change. Exploring latent change classes can help inform researchers on the relationship between patterns of change across time and other variables or outcomes of interest.

**Purpose**

Given the usefulness of change scores, it is important for researchers to explore statistical models that can be applied to change scores. This paper will provide a detailed description of the application of mixture modeling (a person-centered approach) to factor analysis of change scores. In chapter two, I provide an overview of factor analysis of change scores, including its benefits and limitations. Additionally, literature pertaining to mixture modeling, specifically factor mixture modeling, will be examined. For both techniques, common practices will be mentioned and their application to change scores will be highlighted.

To provide the reader with a concrete example of this application, a series of factor mixture models will be conducted on change scores calculated from higher education data. Through the applied example and discussion (Chapter 5), I aim to demonstrate the usefulness of change score factor mixture modeling, particularly for constructs studied in educational and psychological research.

# CHAPTER TWO

## Literature Review

### Change Scores

A change score (also known as a difference score or gain score) usually refers to the difference between two scores from the same person, on the same measure, at two different occasions (Bandalos, in prep):

$$D_i = X_{i2} - X_{i1} \tag{1}$$

In this change score formula, $D_i$ represents the change score and $X_{i1}$ and $X_{i2}$ are scores at times one and two for individual *i*. Often, convention dictates subtracting earlier occasions from later occasions so positive change scores represent growth and negative change scores represent decline. Assuming the same measure was used at both time points, a change score can conceptually be thought of as an individual's observed change or growth in a construct. Although change scores have been widely used in a number of contexts, researchers have debated the utility of these scores. The main argument against analyzing change scores is their reliability, as they are historically perceived as inherently unreliable (Cronbach & Furby, 1970). However, subsequent research supports the use of change scores, suggesting the reliability of change scores depends on a variety of factors (Williams & Zimmerman, 1977, 1996; Zimmerman & Williams, 1982a, 1982b).

### Reliability of Change Scores

Given change scores are used in a variety of statistical models and in making practical decisions, the reliability of change scores should be examined thoroughly. The reliability of change scores, similar to the reliability of scores at a single occasion, can be represented as the ratio of true change score variance over observed change score variance**.** Conceptually, it is the precision with which we can accurately rank order people by their change scores. There are a number of factors that contribute to the reliability of change scores (Williams & Zimmerman, 1996). To highlight these factors, I present an equation for the reliability of change scores from Williams and Zimmerman (1996, p. 60):

$$\rho_{DD'} = \frac{\lambda\rho_{X1X1'} + \lambda^{-1}\rho_{X2X2'} - 2\rho_{X1X2}}{\lambda + \lambda^{-1} - 2\rho_{X1X2}} \qquad 2$$

$X_1$ represents observed scores at the first occasion; $X_2$ represents observed scores at the second occasion; $\rho_{X1X2}$ is the correlation between $X_1$ and $X_2$; $\rho_{X1X1'}$ is the reliability of $X_1$; $\rho_{X2X2'}$ is the reliability of $X_2$; and $\lambda$ is a ratio of the two standard deviations. Specifically, $\lambda$ is computed by dividing the standard deviation of $X_1$ (first occasion) by the standard deviation of $X_2$ (second occasion) and $\lambda^{-1}$ is computed by taking the inverse of $\lambda$ (i.e., standard deviation of $X_2$ divided by the standard deviation of standard deviation of $X_1$). Note three things are taken into account when calculating the reliability of change scores with this equation: the reliability of the scores at each occasion, the correlation between the scores, and the standard deviation of the scores at each occasion.

Researchers who have argued against the reliability of change scores have not made their conclusions based on this equation. Instead, they have drawn their conclusions based on a more simplified version of the equation that makes two assumptions about the scores at each occasion. The first assumption is the standard deviations of the two sets of scores are the same, resulting in a $\lambda$ value of one. The second assumption is the reliabilities of $X_1$ and $X_2$ are equal (i.e., scores are equally reliable at both occasions of measurement). With these two assumptions, equation 2 can be reduced to (Williams & Zimmerman, 1996, p. 60):

$$\rho_{DD'} = \frac{\rho_{X1X1'} - \rho_{X1X2}}{1 - \rho_{X1X2}} \quad 3$$

Although these two assumptions may appear to be logical in Classical Test Theory where pre-test and post-test measures may be parallel, it does not hold in situations where the standard deviations and reliabilities are different at the two occasions. Thus, reliability of the individual scores, their standard deviations, and their correlation should be taken into account when calculating change score reliability. More importantly, it is critical to understand how these factors influence the reliability of change scores. Although it makes sense that two unreliable sets of scores will result in unreliable change scores, the impact of the correlation between the scores and their standard deviations on reliability of change scores may be unclear. Thus, I discuss these two characteristics in the following sections.

**Correlation between $X_1$ and $X_2$.** The reliability of change scores is affected by the degree to which the scores at two given occasions correlate with one another (Williams & Zimmerman, 1996). Change score reliability will increase as the correlation between the two occasions of scores decreases. For example, suppose a researcher measured a sample of students on life satisfaction at two occasions. If every student's score increases by five points from time one to time two, the correlation between the scores will be perfect at +1.00 and, consequently, the change scores will have zero variance. In this case, the reliability of change scores will equal zero because everyone has the same rank. In contrast, if some students' scores are positive (growth in life-satisfaction) and others' scores are negative (decline in life satisfaction), the correlation between scores on the two occasions will be lower than +1.00. As a result, the change scores will have variance. In this scenario, the reliability will be higher than zero because variance in the change scores allows students to be rank ordered.

**Standard Deviations of $X_1$ and $X_2$.** The standard deviations of $X_1$ and $X_2$ also influence the reliability of change scores. The effect of the correlation between $X_1$ and $X_2$ on reliability of change scores is most influential when the standard deviations of $X_1$ and $X_2$ are equal ($\lambda = 1$) and is less influential on reliability as $\lambda$ moves away from 1. In other words, when the standard deviations at each occasion are different from one another, correlated scores across time are less of a problem for reliability than when the standard deviations are similar (Sharma & Gupta, 1986). As the ratio of standard deviations moves further away from one, the reliability of change scores increases, holding all other factors constant.

## Scenarios in Which Change Scores are Reliable

It is common for longitudinal data in higher education and psychological research to possess characteristics that would result in high reliability of change scores. Experiments conducted by psychologists designed to assess the impact of an intervention are likely to produce scores not strongly correlated with one another at two occasions. Suppose an intervention study on alcohol use consists of two groups: an intervention group and a control group. Assuming the intervention is effective, the intervention group should report a greater decrease in alcohol use than the control group, which should report minimal change in alcohol use, if at all. The reliability of change score is not inherently a concern here because the correlation between the two occasions should be very low and the standard deviations would likely differ.

Additionally, in many instances, we would expect to find the spread of scores to be different across occasions. For example, course grades could have different standard deviations at two occasions of measurement. In an introductory statistics course, students likely come in with little knowledge of statistics. That is, students start out at approximately the same place in the course. A statistical knowledge test administered as a pre-test at the beginning of the course would likely yield scores with a small standard deviation. At the end of the course, some students are likely to have excelled and thus obtained near-perfect scores on a post-test. Others have mediocre scores, and some have failing scores. The standard deviation of the test scores on a post-test, administered at the end of the semester, would be much larger than the standard deviation of scores on the pre-test (resulting in $\lambda$ farther from 1.00 and boosting reliability of change scores). Under these and similar circumstances, researchers, especially those conducting higher education or psychology research, should feel comfortable using and analyzing change scores, although it is recommended researchers check the reliability of change scores. These scores can also be particularly useful in modeling latent constructs such as in factor analysis.

**Factor Analysis of Change Scores**

Factor analysis is a commonly used statistical technique in the fields of psychology and education. The purpose of factor analysis is to identify a reduced set of latent constructs (i.e., factors) that best explain the relationships among a set of observed variables (Benson & Nasser, 1998). There are two broad classes of factor analytic methods: exploratory factor analysis and confirmatory factor analysis. Exploratory factor analysis allows a researcher to explore and identify underlying latent constructs while confirmatory factor analysis provides a way to test factor structures that are driven by theory or prior research (Kline, 1998). Researchers traditionally factor analyze cross-sectional scores. However, change scores can also be factor analyzed; a method introduced by Cattell (1963; Nesselroade & Cable, 1974). Although factor analysis of change scores was introduced decades ago, researchers tend to neglect this approach. Factor analysis of change scores is particularly useful for better understanding constructs theorized to be fluid in nature over time, such as the growth and development of psychological phenomena and knowledge. This paper aims to build upon factor analysis of change scores.

Factor analytic methods conducted on change scores and cross-sectional scores are identical. Only the analyzed scores and interpretations of the factors extracted are different. Cross-sectional factors represent latent constructs influencing observed responses at a single occasion, whereas change score factors represent latent growth or change processes influencing change in observed responses. Thus, the factor structure of change scores and cross-sectional scores may be completely different from one another, even for scores from the same scale. Despite this claim, researchers have expressed concern and skepticism towards the independence of change score factors and cross-sectional factors. To aid in the understanding of how the two structures can be mostly independent, it may be helpful to consider Lord's paradox.

**Connection to Lord's Paradox.** In 1967, Lord presented a scenario in which two researchers, both interested in examining change between groups, found different results when analyzing the same dataset. The research question was whether the groups differ on some measure at time two accounting for baseline scores (scores at the first occasion;

Holland, 2005; London & Wright, 2012; Lord 1967; Schafer, 1992; Wainer, 1991). The two statistical approaches used were analysis of variance (ANOVA) on change scores and analysis of covariance (ANCOVA) on post-test scores, controlling for pre-test scores. Although the two approaches seemed to answer the same question, they produced contradictory results. This later became known as Lord's paradox. To help demonstrate how a researcher could obtain opposing results when using these two approaches, I compare the regression equations associated with each of these models. The ANCOVA model can be expressed in regression form as:

$$\text{Post} = b_0 + b_1(\text{Group}) + b_2(\text{Pre}) + \text{e} \qquad 4$$

In this equation, $b_0$ is the intercept, $b_1$ is the effect of Group on post-test scores independent of pre-test scores, and $b_2$ is the effect of pre-test scores on post-test scores independent of Group. Note that the ANOVA equation can be derived from the ANCOVA equation by fixing the $b_2$ parameter to one and subtracting pre-test scores from each side:

$$\text{Post-Pre} = b_0 + b_1(\text{Group}) + ((1*\text{Pre})\text{-Pre}) + \text{e} \qquad 5$$

$$\text{Change} = b_0 + b_1(\text{Group}) + \text{e} \qquad 6$$

Here, Change is the difference between pre-test and post-test scores, $b_0$ is the intercept, and $b_1$ is the effect of Group on Change. By examining the regression equations of the two approaches, it is clear the only difference between them is the conceptualization of the $b_2$ parameter. With the ANCOVA approach, the $b_2$ parameter is estimated, whereas for the ANOVA, the value is fixed to one. Conceptually, the ANCOVA examines how groups differ on their post-test scores, accounting for pre-test scores. Thus, we are mainly interested in the difference between groups at post-test. The ANOVA model tests whether the groups differ in their amount of change; taking pre-test scores into account in a different way.

London and Wright (2012) provided an applied example of Lord's paradox by comparing the use of ANOVA and ANCOVA to examine the effects of a child's age on words recalled after a 10-month delay. The children in this study were separated into two groups: one older group (ages seven to nine) and one younger (ages four to six). The children were tested on the number of words recalled at two occasions, once at the start of the experiment (time one) and again after a 10-month delay (time two). Researchers found what appear to be opposing results. The ANCOVA model indicated, after controlling for scores at time one, that older children remembered more words at time two ($b_{\text{Age-Group}} =$ 0.78). In contrast, the ANOVA model suggested the older children exhibited a steeper decline in words remembered across time than the younger children ($b_{\text{Age-Group}} =$ -4.07). In other words, "according to this approach the older children forget more" (London & Wright, 2012, p. 284). Although these results seem contradictory, they are not. The two models actually answer very different research questions. The ANCOVA answers the question: How do groups differ on where they end up, controlling for where they started? The ANOVA answers: How do groups differ in growth, change, or development? Although the dependent variables in these analyses contain overlapping information (i.e., they both contain post-test scores), they are not the same and can yield independent results.

This same idea of competing research questions can be applied to factor analysis of change scores and factor analysis of cross-sectional scores. The two methods can be viewed as answering two different research questions, just as the ANCOVA and ANOVA in Lord's paradox. Cross-sectional analyses, such as an analysis of post-test or pre-test scores only, model the latent dimensions underlying observed responses at a single occasion, whereas change score analyses model latent dimensions underlying change, growth, or development. Given this distinction, the structures may be at least partially independent of one another because they answer disparate research questions. The decision to use which method depends on the interest of the researcher. Researchers interested in the processes underlying changes in observed scores should model the dimensions of change scores, while those interested in processes or constructs influencing responses at a single occasion should model dimensions of cross-sectional scores.

Identifying dimensions of change can reveal a breadth of new information. Novel factors may be extracted to represent more dynamic constructs than when analyzing at any single occasion of data. For example, a construct may be best represented with one factor at any single occasion, but change in the construct may have two or three factors as responses on groups of items change together. Knowing the structure of change also allows the researcher to identify covariates of change that would otherwise be masked or diluted if we assume the structure of change is identical to that of cross-sectional scores. Additionally, factor analyzing change scores allows researchers to examine patterns of change across time (e.g., determining the items on which respondents tend to change similarly across time and in what ways those items relate to one another), a critical component to understanding underlying mechanisms of growth.

Given factor analysis of change scores may aid in understanding change processes, it can be a valuable method for disciplines that focus on development and growth, such as higher education research. In the following sections, I will guide the reader through how to apply this method in greater detail. Specifically, I will discuss how to (a) apply and conduct exploratory and confirmatory factor analysis on change scores, (b) interpret change score factors, and (c) collect validity evidence for change score factors.

## Exploratory Factor Analysis with Change Scores

**General Overview.** In situations where there is limited research regarding the factor structure of a construct, researchers rely on exploratory factor analysis (EFA) to identify and explore underlying dimensions of a set of observed variables. The logic behind EFA is simple. If variables are highly correlated with one another, they are likely tapping into the same underlying dimension. The purpose of using EFA is to identify a smaller set of latent dimensions that explain the correlations among the variables. Researchers using factor analysis assume correlations among observed variables are a result of one or more underlying latent constructs. Identifying factors underlying the observed variables allows the researcher to examine how the observed variables cover the breadth of the construct. For instance, if a latent construct is assumed to be multidimensional, it is critical to include several items on the scale that represent each dimension of that construct (Bandalos & Finney, 2010).

The focus of EFA is not on model fit, but rather on the estimated parameter values (e.g., structure coefficients, pattern coefficients, factor correlations, and communalities) obtained from the model. The goal of EFA is to reproduce the correlations or covariances among the observed variables from loadings and factor correlations on a smaller set of latent variables (Bryant & Yarnold, 1995). Latent factors are extracted from the correlation matrix. However, because EFA extracts factors that best explain shared variance among the variables, communalities must first be estimated for each variable. Communalities are measures of shared variance among variables. These values cannot be directly calculated, but are estimated through an iterative process. Typically, researchers begin with starting values for communality estimates. The starting values are used to estimate the initial communalities. Through a series of iterations, the communalities are updated until the best values are obtained. The correlation matrix is then “reduced” by replacing the ones on the diagonal with the best estimated communalities; factors are then extracted from the “reduced” correlation matrix.

In EFA, the part of the variance in the scores that is accounted for by common factors is known as common variance. However, latent common factors may not account for all of the variance in the scores. Each variable is associated with a latent unique component as well, typically called a “uniqueness.” Some traditions refer to uniquenesses as residuals or residual variances, or the part of the observed scores not explained by the common factors. In terms of common variance, there are two important estimated parameters that should be mentioned: structure coefficients and pattern coefficients; both known as factor loadings. In general, loadings summarize the relationship between latent factors and observed variables. Structure coefficients represent the zero-order correlations between observed scores on each variable and factor scores, and pattern coefficients represent this same correlation, but with variance explained by all the other factors partialed out. In cases in which factors are uncorrelated or there is only one factor, structure and pattern coefficients will be identical. With multiple correlated factors, researchers should take into account both of these coefficients during interpretation (Bandalos & Finney, 2010).

Just as observed score variance can be partitioned into common variance shared with all other variables and unique variance, observed scores can be decomposed into a unique component and a common component. The unique component of an observed score is the part unrelated to the factors (residuals) and the common component of an observed score is the part assumed to be influenced by the factors. A person's score on an observed variable can be represented as a linear combination of the common component(s) and unique component of the score. Thus, an observed score (X) for a person can be expressed in equation form as (Bandalos, in prep):

$$X_{iv} = w_{v1}F_{1i} + w_{v2}F_{2i} + \ldots + w_{vf}F_{fi} + U_{iv} \qquad 7$$

Where $w_{vf}F_{fi}$ is the influence of the factors on each variable score. The term $w_{vf}$ represents the weight or loading associated with variable *v* on factor *f*. The term $F_{fi}$ represents the factor score on factor *f* for individual *i*. Thus, each variable (subscripted *v*) can have a different loading on each factor (subscripted *f*), and each individual (subscripted *i*) can have a different score on each factor (subscripted *f*). Lastly, the term $U_{iv}$ represents the unique component of each variable *v* score for individual *i* (Bandalos, in prep).

In EFA, variables can take on a variety of forms such as subscale scores, scale scores, or item scores. The only difference between conducting an EFA with change scores and cross-sectional scores is the scores themselves. In cross-sectional EFA, the scores are observed variables measured at a single occasion. Factors are extracted from the observed reduced correlation matrix. Alternatively, in change score EFA, the calculated change scores are used in the analysis. Change is usually calculated by subtracting the first occasion from the second occasion (i.e., change = post - pre).

Correlations among every pair of change score variables are calculated, and factors are extracted from the change score correlation matrix. Thus, the same equation (7) can be generalized to factor analysis of change scores, although the term $X_{iv}$ becomes the change score for individual *i* on variable *v* and $w_{vf}F_{fi}$ becomes the influence of the factors on each change score. An individual's change score on a variable can be represented as a linear combination of factor loadings (or weights) and factor scores plus some uniqueness.

In EFA, a number of decisions must be made throughout the course of the analysis such as method of factor extraction, number of factors to extract, and rotation of factors. Many of these decisions are subjective and are ultimately guided by theory. Given the only major difference between factor analysis of change scores and factor analysis of cross-sectional scores is the scores themselves, common practices associated with factor extraction such as: identifying the number of factors and factor rotations for factor analysis of cross-sectional scores, are also applicable to factor analysis of change scores. Among these decisions, the number of factors to extract is often viewed as one of the most important decisions in factor analysis (Comrey & Lee, 1992; Gorsuch, 1983; Hakstian, Rogers, & Cattell, 1982).

**Methods of Extraction.** As noted in the previous section, the end goal of factor analysis is to extract a smaller set of latent factors that explain the most common variance in the observed scores. Thus, extraction refers to the process by which factor parameters are estimated from the correlation matrix to form a number of factors that best describe the data. Although there are a number of extraction methods, two widely used methods are principal axis factoring (PAF; Benson & Nasser, 1998) and Maximum Likelihood (ML; Bentler & Bonett, 1980).

***Principal Axis Factoring.*** Principal axis refers to a least squares-type method where the residuals (differences) between the observed correlation matrix and the model-implied correlation matrix are minimized (Benson & Nasser, 1998). As in most methods of extraction, communalities must be estimated through an iterative process. Once the best estimates of communalities are obtained, they are inserted on the diagonal of the correlation matrix to represent the amount of variance shared by the observed variables. Placing the communalities on the diagonal implies observed variables contain some degree of error (e.g., the uniqueness component) because the communalities are often below one. Finally, the factor parameters are estimated (or extracted) from the reduced correlation matrix such that each factor explains the maximum amount of observed variance, independent of all previously extracted factors.

For example, imagine we are conducting an EFA. The first factor extracted explains as much common variance as possible in the observed variables. Once the variance associated with the first factor is removed from the correlation matrix, a one-factor EFA model is used to generate a model-implied correlation matrix. The model-implied correlation matrix is compared to the initial correlation matrix; the difference between the two matrices is referred to as the residualized matrix. The second factor is extracted from this residualized matrix and explains as much of the common variance left over in this matrix as possible (what is not explained by the first factor). This procedure continues until there are zeros on the diagonal of the residualized matrix (i.e., no common variance left) or until some extraction criterion is met. However, the only time there will be no variance left in the residualized matrix is when the number of factors extracted is equal to the number of observed variables. For this reason, the decision of how many factors to retain is critical. The most popular methods for identifying the number of factors to retain are discussed in the following section.

***Maximum Likelihood.*** Another widely used factor extraction method is maximum likelihood (ML; Bentler & Bonett, 1980). Assuming there are a specified number of factors underlying the data and the data are from a sample where the distribution is multivariate normal, ML estimates factor parameters (e.g., loadings, intercepts, and error variances) based on the sample size and number of observed variables. Simply put, ML estimates factor parameters that are most likely to reproduce the observed correlation matrix through an iterative process. A set of starting values is initially used to estimate the factor parameters. The produced correlation matrix from the estimated factor parameters is then compared to the observed correlation matrix to examine how well the estimated factor parameters reproduce the observed data. The initial starting estimates are replaced with new estimates, each associated with a log-likelihood, representing the likelihood of the data given the set of estimated parameters. This iterative process is continued until convergence is met (i.e., the estimates associated with the highest log-likelihood is generated; Bentler & Bonett, 1980). Convergence of a solution, however, may not always be possible. For instance, ML cannot be used when the observed correlation matrix is singular (i.e., matrix does not have an inverse).

**Number of Factors.** A variety of statistical and visual methods have been established to determine the number of factors to extract. However, these methods can yield contradictory results. There is not a correct answer to this question. Instead, the most important things to consider are theory and interpretability. Researchers should have a justification or reason for extracting a certain number of factors. This may be from theory about why and how the variables should covary with one another. For areas where there is limited research, agreement among different statistical and visual methods as well as interpretability is ideal. Regardless, a factor is only useful to researchers if its interpretation is meaningful. Common methods for determining the number of factors include the Kaiser criterion (K1 method), parallel analysis, and the scree plot. It should be noted that each method has flaws and results should be compared across different methods for better accuracy.

***K1 Method.*** One of the most commonly used methods is the eigenvalue-greater-than-one rule, or Kaiser criterion (K1; Kaiser, 1960). Under this method, eigenvalue decomposition is conducted on the correlation matrix. The number of eigenvalues greater than one is the number of factors that should be extracted from the data. This method originated from Guttman (1954), in which it was first proposed as a method for estimating the lower bound for the rank (or dimensionality) of a population correlation matrix. According to Guttman (1954), the minimum number of factors extracted from a correlation matrix with unities on the diagonal should be equal to or greater than the number of eigenvalues greater than one. However, there are some reservations with the K1 method. It is intended that researchers be able to identify the lower bound for the rank of a correlation matrix. Despite this, they often use this criterion to determine the exact number of factors. Guttman’s findings were also derived from population data, which does not take into account sampling error. In finite samples, sampling error tends to increase the rank of a correlation matrix. Thus, using the K1 method may overestimate the number of factors (Horn, 1965). Moreover, the first few eigenvalues in a sample correlation matrix tend to be larger than those in the population correlation matrix (Nunnally & Bernstein, 1994). In spite of these limitations, the K1 method is the default method of determining the number of factors in SPSS and is widely used and adopted by many researchers.

***Parallel Analysis.*** Research indicates parallel analysis (PA; Horn, 1965) is one of the most accurate methods in determining the number of factors to extract (e.g., Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1982). PA is an extension of the K1 rule with an adjustment for sampling error (Carraher & Buckley, 1991; Zwick & Velicer, 1986). In PA, correlation matrices of random variables, which are based on the same sample size and number of variables in the observed dataset, are simulated. Given variables are randomly generated, the correlations among the variables should be zero. However, the correlations will never be exactly zero due to sampling error, and thus eigenvalues will never be exactly one for each variable. PA adjusts for the sampling error. Eigenvalues are first calculated for each simulated correlation matrix. The average eigenvalues across the simulated correlation matrices are then compared to the observed eigenvalues, such that the first eigenvalue of the real data is compared to the average of all first random eigenvalues, and so on with successive eigenvalues. The number of factors that should be retained is equal to the number of observed eigenvalues that are greater than their corresponding average eigenvalues (Hayton, Allen, & Scarpello, 2004; Horn, 1965).

***Scree Plot.*** Another commonly used method for determining the number of factors to retain is Cattell's (1966) scree plot test. This method involves plotting the eigenvalues, with eigenvalues on the Y-axis and number of factors on the X-axis.

Researchers look for a steep curve on the plot followed by a bend and then a flat horizontal interval on the resulting line. The bend in the curve is known as an elbow, a point at which the eigenvalues level off and begin to form a horizontal line. Factors before the elbow should be retained (Cattell & Jaspers, 1967). There are a few problems associated with this method. In situations where the number of variables and sample size are both small, there is not always a clear break (Cliff & Hamburger, 1967). Additionally, the method is subjective to the researcher, especially when there is not a clear break or there are multiple breaks. Some researchers include the elbow as a factor that should be retained, some do not include the elbow, and some include one factor after the elbow. Whether or not a researcher should keep the elbow is debatable. Despite these problems, studies have reported high interrater reliability for the scree plot test (e.g., Cliff, 1970; Zwick & Velicer, 1982).

**Methods of Rotation.** Once factors are extracted from the observed reduced correlation matrix, rotation of the factors often eases interpretation. Recall factors are extracted such that each maximizes the amount of observed variance explained. Once the number of extracted factors is determined, this becomes the unrotated factor solution and accounts for a proportion of the total shared variance. The unrotated factor solution is then rotated to allow the variance accounted for by the extracted factors to redistribute among the factors. This enables the variance to spread out among all factors rather than remain concentrated in the first few factors (Benson & Nassar, 1998). This redistribution often makes the factors more meaningful and easier to interpret. However, the total variance accounted for in the rotated and unrotated solutions will always be the same. There are two primary types of rotation: oblique or orthogonal. Specifying an oblique rotation allows the factors to be correlated with one another. With orthogonal rotation, the factors are restricted to be uncorrelated. The choice between oblique and orthogonal often depends on the construct being examined. For example, factors that underlie constructs in higher education and psychology tend to be correlated and thus, oblique rotation is often more appropriate.

There are a variety of oblique and orthogonal rotation methods. In essence, the goal of each method is to simplify the patterns of loadings to approximate a simple structure. This can be achieved by doing one of two things: minimizing the number of factors on which each variable loads or minimizing the number of variables that load on each factor. Mathematically, factors are rotated by multiplying the unrotated matrix by a transformation matrix. Transformation matrices are solved using algorithms that take into account any restrictions imposed on the factors (correlated or uncorrelated). The most commonly used oblique rotation methods are Direct Oblimin and Promax while the most commonly used orthogonal methods are Varimax and Quartimax.

## Confirmatory Factor Analysis with Change Scores

**General Overview.** Unlike EFA, where the factor structure is not specified in advance, but extracted from the reduced correlation matrix, researchers typically use confirmatory factor analysis (CFA) to test and investigate known or hypothesized factor structures. When conducting a CFA, the researcher is able to specify a reduced set of relationships between variables and latent factors (i.e., each variable is often typically only related to one factor). CFA is thus viewed as more restrictive than EFA because an a priori factor structure must be specified. The specified model is then estimated and evaluated on a variety of fit indices (e.g., log likelihood (*LL*), BIC, AIC, incremental fit, RMSEA). These fit indices provide the researcher with criteria for how well the observed data fit the specified CFA model (DeVellis, 1991). Because CFA is more stringent than EFA, it is often recommended that researchers generate models based upon prior literature and theory of a construct (e.g., Bandalos & Finney, 2010). When there is minimal research on the construct, an EFA on an independent data set should first be conducted to explore the dimensionality of the construct (Worthington & Whittaker, 2006). The latter is often the case when factor analyzing change scores because of the lack of research on change score factors. Similar to EFA, the only difference between conducting a CFA on change scores versus cross-sectional scores is the scores themselves.

In addition to examining the model fit of a specified factor structure, CFA can be used to evaluate and compare competing nested models (Bryant & Yarnold, 1995). For example, CFA could be used to compare two different factor models that underlie a ten-item scale measuring diversity: (a) a two-factor model consisting of five items on each latent factor, representing multicultural competency and racial diversity and (b) a one-factor model consisting of ten items on a single latent factor, representing a general diversity construct. Because the two proposed models are nested (one contains a subset of the parameters estimated in the other) the difference in fit between the two models can be tested using a chi-square difference test (also known as likelihood ratio test). The chi-square value and degrees of freedom for the general model is subtracted from the chi-square value and degrees of freedom for the more restrictive model. The difference in chi-square is evaluated as if it were an ordinary chi-square test, using the difference in degrees of freedom as degrees of freedom for the significance test. A significant difference in chi-square values means the more complex model (the model with fewer degrees of freedom) provides a significantly better fit to the data compared to the more parsimonious model. Thus, a significant chi-square test indicates the more complex model should be championed as the best fitting model.

In CFA, observed scores (or change scores) are decomposed into two or more latent factors, at least one representing the construct of interest and one representing error associated with the scores. In specifying a factor model, the researcher would model scores on each observed variable as being influenced by one or more latent constructs and by unexplained latent error. Additionally, researchers could also specify whether factors are oblique or orthogonal to one another. In contrast to EFA, where the uniqueness of each variable is assumed to be independent of one another, CFA allows these errors to be either independent or correlated.

**Evaluating Model Fit.** In CFA, the factor structure specified by the researcher is used to produce a model-implied covariance (or correlation) matrix. The model-implied covariance matrix is compared to the observed covariance matrix (the difference between the two matrices is known as the fitted residual matrix), which allows researchers to evaluate how well the specified factor model reproduces the observed covariance matrix. If the discrepancy between the observed and model-implied covariance matrix is large, the specified model should not be used to represent the relationships in the data (Hu & Bentler, 1995). Researchers often rely on multiple model fit indices to determine how well the data fit the specified model. Each CFA model yields a chi-square ($\chi^2$) value, which is a measure of the exact data fit, and a probability value ($p$-value) associated with it. The $\chi^2$ $p$-value indicates the probability that the fitted residuals generated by the model are different from zero assuming the model actually fits the data well. Thus, a non-significant $\chi^2$ value with $p > .05$ is desirable, though the $\chi^2$ significance test can be overly sensitive to sample size (Bryant & Yarnold, 1995; Hu & Bentler, 1998). Therefore, it is common practice to examine approximate model fit statistics in addition to exact data fit statistics (Schreiber, Nora, Stage, Barlow, & King, 2014).

Approximate model fit can be further categorized into two different types: absolute and incremental. Absolute fit indices measure how well the model reproduces the observed correlation matrix with no comparison to a reference model (e.g., standardized root mean square residual and root mean square error of approximation). Incremental fit indices measure how much better the specified factor model fits the data relative to a more restricted model (e.g., comparative fit index).

***Standardized Root Mean Square Residual.*** The standardized root mean square residual (SRMR) is a measure of the average fitted residual generated by the specified model. It is the square root of the average of the squared fitted residuals. Recall fitted residuals are the difference between the observed covariance matrix and the model-implied covariance matrix. The SRMR values range from 0 to 1.0, with lower values being indicative of a better fit.

***Root Mean Square Error of Approximation.*** The root mean square error of approximation (RMSEA; Steiger, 1990) is also a standardized measure of the lack of fit of the specified model. The RMSEA takes into account model complexity and adjusts for parsimony. Simpler models (models with more degrees of freedom) will have lower RMSEA values than equally well-fitting but more complex models. RMSEA values range from 0 to 1.0, with values closer to zero indicating a better fit. RMSEA is sensitive to model misspecification and the number of variables in the model. It tends to increase as the number of variables in the model increases (Kenny & McCoach, 2003).

***Bentler's Comparative Fit Index.*** The comparative fit index (CFI; Bentler, 1990) compares the fit of the specified model to that of the most restrictive model (null model), in which there are no underlying latent factors and correlations among the variables are purely a result of sampling error (Tanaka, 1993). CFI values range from 0 to 1.0; higher values indicate better fit to the data. CFI penalizes the model for every parameter estimated. Thus, complex models are penalized more when using CFI.

There are a number of other fit indices that one could use to evaluate approximate model fit, including: the Tucker-Lewis coefficient (TLC, Tucker & Lewis, 1973); Bollen's (1989) incremental fit index (IFI); and goodness-of-fit (GFI) and adjusted-goodness-of-fit (AGFI) indices (Jöreskog & Sörbom, 1986). However, the three most common are SRMR, RMSEA, and CFI (Jackson, Gillaspy, & Purc-Stehenson, 2009). Extant literature has recommended the following guidelines: CFI $\geq$ .95, SRMR $\leq$ .08, RMSEA $\leq$ .06 (Bentler, 1990; Browne & Cudek, 1993; Hu & Bentler, 1999).

***Local Fit.*** To evaluate local fit of the model, correlation residuals for each variable should be examined. Correlation residuals are calculated by taking the difference between elements of the observed correlation matrix and the model-implied correlation matrix. Ideally, if the CFA model is correctly specified, the correlation residuals (i.e., the difference between the two matrixes) for each variable should be small. Correlation residuals greater than an absolute value of 0.10 are an indicator of modelmisspecification (Kline, 2013). In some instances, it is possible for a model to have good global fit overall, but poor local fit.

**Interpretation of Change Score Factors**

The other major difference between factor analysis of change scores and cross-sectional scores, other than the scores themselves, is the interpretation of the factors. As with any factor analysis, the factor solution from change scores should be interpreted using all of the estimated parameters (e.g., structure coefficients, pattern coefficients, and factor correlations if applicable). The initial step of the interpretation process is to identify which variables load on to which factors. To do this, the researcher must determine a cut-off value for the coefficient to be considered salient. Most factor analysts are comfortable with structure coefficient values greater than .30 or .40 (e.g., Yong & Pearce, 2013) to indicate salient loadings. However, the choice of which value to use is often arbitrary and could vary across researchers and disciplines. Additionally, researchers could square the pattern coefficient terms to obtain the proportion of unique variance accounted for in each variable by each of the latent common factors. If factors are orthogonal (i.e., uncorrelated), squaring the terms will result in the proportion of variance accounted for in the variable by the factor. If factors are oblique, squaring the terms will represent the unique proportion of variance accounted for in the variable by the factor, controlling for the other factors in the model.

Change score factors represent latent change or growth processes that influence the change in responses from time one to time two. Thus, change scores that load onto the same factor change in a similar way across the two occasions. Researchers should examine the content of the items (or variables) carefully to aid interpretation of factors. Moreover, theory surrounding the construct should be taken into account when naming and interpreting change factors. This is especially critical for constructs theorized to be fluid, in which development may be multifaceted. There should be a clear alignment between the interpretation of the factors and the theory behind the construct. In situations where this may not be possible, researchers should consult and collaborate with their colleagues in interpreting change score factors.

**Validity Evidence for Change Score Factors**

Because EFA and CFA are used to model latent change processes, it is important for researchers to collect external validity evidence for the change score factors. This process for change score factors is identical to that of cross-sectional factors. Reliability estimates for each change score factor should be computed and reported. If the construct is best represented using one change factor, then only one reliability estimate (typically Cronbach's alpha) is necessary. If the construct is multidimensional, reliability estimates should be calculated for each of the dimensions or subscales (if they are to be used independently). In addition, researchers should examine the relationship of the change score factors to other theoretically related variables. This may be difficult with change score factors, because research on the relationships between different latent change processes may be limited. In such cases, researchers should use and rely on previous theory of the construct involving cross-sectional factors as a starting point. For example, if sense of identity is positively related to students' GPA, it is logical that change in sense of identity (represented by the change score factors) might also be related to students' GPA. This process can provide researchers with additional research questions to explore. For instance: if change in sense of identity is best represented with two factors, do these factors differentially predict GPA? This information may yield some of the first insights on how different latent change processes relate to one another and how their predictive utilities may not be consistent with those of cross-sectional factors. Additional valuable information is gained through the validating process of change score factors.

**Factor Mixture Modeling With Change Scores**

Thus far, I have described factor analytic methods (EFA and CFA) to model latent change processes using change scores, assuming the change score factors are the same for everyone in the population of interest. For example, the change score CFA model represents the relations among observed change scores computed from a sample of individuals drawn from a homogenous population. However, this may be a faulty assumption for populations investigated in the social sciences and other related fields of research. The change score factor model parameters (i.e., factor means, loadings, and intercepts) may differ across subpopulations if the population studied is actually heterogeneous. If this is the case, the change score CFA model can be extended to allow parameters to differ across multiple groups (Jöreskog, 1971). Typically, groups are well defined or known in the population (e.g., gender and race) and the process of obtaining and comparing group parameters is fairly straightforward. However, it is possible for change score factor model parameters to differ across not only defined groups, but also unobserved (or unknown) groups in the population. To account for this statistical possibility, we can employ a technique called factor mixture modeling (FMM).

In factor analysis of change scores, the extracted change score factors are interpreted as latent change processes and are assumed to be the same for everyone in the population. In some empirical fields, these change processes may differ by person. For example, suppose a researcher surveyed a group of students on aspects of alcohol use at two occasions. The researcher analyzed the factor structure of change scores and results supported a two-factor solution. One change factor represented change in negative experiences with alcohol and the other factor represented change in alcohol consumption. In this context, it is unlikely every student will exhibit the same pattern of change. One group of students may report substantial growth in both negative experiences and consumption, with high averages on both the change in negative experiences factor and the change in consumption factor. Another group of students might report a high positive average score on the change in negative experiences factor, but this increase in negative experiences has deterred them from consuming more alcohol. Thus, their consumption decreases across the two occasions and their consumption change factor score is, on average, low. Not only do the mean factor scores differ between these two groups, but the factor covariance differs as well. For one group, the two change factors are positively related, and for the other, this relationship is negative.

These differences between unknown groups can be accounted for by adding a single categorical latent variable to the two-factor change score model. The categorical latent variable models the unknown population heterogeneity as unknown groups with different patterns of change processes. Combining a single categorical latent variable and a factor analysis model is known as factor mixture modeling (FMM). The purpose of factor mixture modeling is to assess whether the data consist of unknown groups that differ in their factor model parameters (i.e., loadings, covariances, and intercepts; Pastor & Gagné, 2013). Note in the example above that the groups differed in their factor means and covariance. However, it is possible for groups to differ on other parameters as well, such as loadings, variances, and intercepts, which I will discuss in greater detail.

## Mixture Modeling

**General Overview**. Before further describing change score factor mixture modeling, it is important to understand the basics of mixture modeling. Mixture analyses are often called upon to explore unknown groups in a population. Specifically, mixture modeling assumes the population distribution is a mixture of multiple distributions, with each distribution and corresponding probability density function belonging to an unknown “class” in the data (Pastor & Gagné, 2013). Thus, the population distribution is a weighted sum of all of the underlying distributions (Lubke, 2010). Each underlying distribution is considered a class. If only one variable is modeled, each class is assumed to have its own univariate distribution. If two or more variables are modeled, each class has its own multivariate distribution and covariance matrix. The goal of mixture modeling is to estimate parameters describing these distributions (e.g., means, variances, and covariances) for each of the classes.

Consistent with factor analysis, the observed covariance (or correlation) matrix is also a focus in mixture modeling. Whereas factor analysis assumes correlations among variables are a result of a set of underlying continuous factors, mixture modeling assumes the correlations among variables reflect the presence of unknown discrete classes in the population. In other words, the two methods differ in the type of latent variables they model. Factor analysis is used to model continuous latent variables (factors), and mixture modeling is used to model categorical latent variables (unknown classes in the population; Pastor & Gagné, 2013). For example, suppose a researcher found a positive correlation between alcohol use and drug use. Through a factor analytic framework, the correlation reflects an underlying sensation-seeking dimension. Alternatively, the same observed correlation between alcohol use and drug use could be a by-product of two classes in the population: one group characterized by high levels of alcohol use and drug use and the other characterized by low levels of alcohol use and drug use. When the two groups are mixed together, the overall population would produce the observed positive correlation between alcohol use and drug use. Thus, the observed correlation is simply a result of mixing classes with different means on two continuous variables. In mixture modeling, all covariation in the variables is assumed to be due only to differences between classes (Pastor & Gagné, 2013). Once classes are accounted for (using a categorical latent variable), the observed variables do not covary and any correlations left between the variables are due to sampling or measurement error. This is an assumption known as local independence.

Two important decisions must be made prior to using a mixture model. First, the researcher must choose how many classes to model. Researchers typically model several numbers of classes and compare model fit across solutions. Second, the researcher must specify a distributional form for the classes. The distributional form can be either univariate or multivariate, depending on the number of variables being modeled. Although many applications of mixture modeling assume normality within classes, other distributions can be specified (e.g., Peel & McLachlan, 2000). For example, if theory dictates the distribution for each class is positively skewed, specifying a positively skewed distribution would allow the model to better fit the data (Pastor & Gagné, 2013).

In addition to specifying the number of classes and distributional form, the researcher may choose to constrain parameters in the model. As noted, each class has its own multivariate distribution with its own mean, variances, and covariances. These parameters can be constrained across classes, within classes, or both and they can either be freely estimated or fixed to a particular value (usually 0 or 1). Any combinations of specifications are possible for each parameter in the model. For example, means can be constrained across classes and freely estimated (i.e., one mean value is estimated and constrained across classes). Meanwhile, variances can be free across classes, but fixed to particular values (i.e., not freely estimated) within the same model. It is possible for parameters to be freely estimated both across classes and within classes. Constrained models are considered simpler because fewer parameters are estimated by the model. Thus, the estimation process is easier for constrained models than unconstrained models (Bauer & Curran, 2004; Pastor & Gagné, 2013).

Similar to CFA models, mixture models can be compared to one another. However, this is only possible under certain situations. Models that differ in the number of specified classes but have the same model parameterization are considered nested models and can be compared directly to one another. Models with the same number of classes but nested parameterizations may also be directly compared to one another. Researchers often estimate a variety of models with the same model parameterization, starting with a one-class model and increasing the number of classes in subsequent models. This allows researchers to examine the fit of each model individually and to compare fit across models to find the best fitting model (Pastor & Gagné, 2013).

**Evaluating Model Fit.** In order to evaluate how well any given mixture model fits the data, a log-likelihood (*LL*) or -2*LL*, obtained by multiplying the *LL* by -2, is calculated for the model. These values convey the likelihood of the data given the estimated model parameters. Lower values of -2*LL* (corresponding to higher values of *LL* and likelihood) indicate superior model-data fit. Although -2*LL* values provide the researcher with a good indicator of model-data fit, researchers may find it most useful to compare -2*LL* values from multiple models. However, because -2*LL* values will always be lower or more desirable for more complex models (i.e., models with more classes), a variety of information criteria measures should always be used in tandem with -2*LL* to evaluate model fit. Information criteria measures account for model complexity (Pastor & Gagné, 2013).

***Information Criteria.*** Measures of information criteria (IC) are often used to assess model fit because they penalize *-2LL* values based on the number of parameters being estimated (model complexity). Some also penalize for sample size (Henson, Reise, & Kim, 2007). Because of this, ICs are more appropriate for model comparison than just the -2*LL* value. Akaike Information Criterion (AIC; Akaike, 1973), consistent AIC (CAIC; Bozdogan, 1987), Bayesian Information Criterion (BIC; Schwarz, 1978), and sample-size adjusted BIC (SSABIC; Sclove, 1987) are commonly-used ICs for determining model fit. The model associated with the smallest ICs is considered the best fitting model. Simulation studies found BIC and SSABIC to be most suitable for use with mixture modeling techniques over other ICs measures (Henson et al, 2007; Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2007; Yang, 2006). The equations for these two ICs are as follows:

$$\mathrm{BIC} = -2LL + \ln(N)\, q \qquad 8$$

$$\mathrm{SSABIC} = -2LL + \ln\left(\frac{N+2}{24}\right) q \qquad 9$$

Where *q* equals the number of parameters in the model and *N* equals the sample size (Schwarz, 1978, Sclove, 1987).

***Lo-Mendell-Rubin Likelihood Ratio Test.*** Typically, the fit of nested factor models are compared statistically using the likelihood ratio test (LRT). Recall in CFA that a one-factor and two-factor model can be compared using a LRT to determine if the addition of the second factor is necessary. Similarly, mixture models can also be nested within one another and thus, may also be compared directly. For mixture models with the same number of classes but nested parameterizations, the LRT is appropriate for model comparison, just like in factor analysis. However, for mixture models that differ in the number of specified classes but have the same model parameterization, the LRT is inappropriate to use for model comparison. This is because the LRT statistic obtained from two mixture models with the same parameterizations ($k$-1 and $k$) does not follow a chi-square distribution. To obtain a $k$-1 class model in mixture modeling, parameters for one of the classes in the $k$ model must be fixed to zero (i.e., the probability of being in the $k$th class is set to zero). As a result, a parameter in the $k$ model is set to a value at the boundary of the parameter space (zero), prohibiting the difference in *LLs* from being chi-square distributed. Therefore, the LRT cannot be used as a measure of comparative fit with mixture models that differ in the number of classes specified yet have the same model parameterization (Lo, Mendell, & Rubin, 2001; Tofighi & Enders, 2007).

The fit between two mixture models differing in the number of classes with the same parameterization can instead be compared using the Lo-Mendell-Rubin (LMR) test (Lo et al., 2001). The LMR test uses an approximation for the distribution of the LRT statistic and allows for models to be compared to an adjusted chi-square distribution. A significant LMR value indicates the more complex ($k$-class) model fits the data better than the more parsimonious ($k$-1 class) model. Simulation studies have shown the LMR test performs well at identifying best fitting models over other ICs measures (e.g., AIC, BIC; Henson et al., 2007; Nylund et al., 2007; Tofighi & Enders, 2007). However, the LMR test can only be used to compare fit between neighboring nested models (i.e., comparing *k*-class and *k*-1 class models of the same parameterizations). Researchers are advised to use both the LMR test and ICs in determining the best fitting mixture model.

**Change Score Factor Mixture Modeling**

**General Overview**. Factor mixture models (FMMs) are a special case of mixture modeling that incorporate a single categorical and one or more continuous latent variables into the same model. In other words, FMMs can be viewed as a hybrid between a factor model and a mixture model. In FMMs, the single categorical latent variable (mixture model) accounts for sources of heterogeneity in the population, and the specified continuous factors (factor model) are used to model dimensionality (i.e., covariance matrix) in each class. Thus, the general FMM can be thought of as an extension of the factor analysis model. Recall that an individual's change score on a variable can be represented as (Bandalos, in prep):

$$X_{iv} = w_{v1}F_{1i} + w_{v2}F_{2i} + \ldots + w_{vf}F_{fi} + U_{iv} \qquad 10$$

Where $X_{iv}$ is individual *i's* change score on variable *v*, $w_{v1}$ is the loading of variable *v* on factor 1, $F_{1i}$ is individual *i's* score on change factor 1, and $U_{iv}$ is the unique component of each score on variable *v* for individual *i*. We can extend the previous equation to represent the relationships among change scores across all respondents by rewriting the equation in matrix form (Bandalos, in prep):

$$\Sigma_X = \Lambda\Phi\Lambda' + \Theta_\delta \qquad 11$$

Where $\Sigma_X$ is the observed change score correlation matrix, the $\Lambda$ matrix contains the loadings of each change variable on each factor, $\Phi$ is the correlation matrix for the change score factors, and the $\Theta$ matrix contains variances (and covariances, if permitted) of the residuals or uniquenesses.

This equation can be transformed into that of a FMM by allowing the factor loadings, correlations, and residuals to vary as a function of class. For class $k$, a FMM can be expressed as:

$$\Sigma_k = \Lambda_k \Phi_k \Lambda'_k + \Theta_{\delta_k} \qquad 12$$

Where the only difference between equations 11 and 12 is the *k* subscript. The *k* subscript allows the factor parameters to vary across classes.

With the development of FMMs, researchers are no longer restricted to modeling latent constructs under either a factor analytic framework or mixture framework. Researchers are able to use both frameworks at once to model latent constructs with FMMs (Kuo, Aggen, Prescott, Kendlet, & Neale, 2008). In fact, we can think of the factor analysis model and mixture model as special cases of the general FMM. The factor analysis model is analogous to a FMM with only one latent class, in which every individual's probability of being in that class is unity. The mixture model is equivalent to a FMM where the factor covariance matrix is fixed to zero (i.e., no underlying factors). Because certain constraints can be made to the general FMM to obtain the factor analysis model and mixture model, the two models are considered nested within the FMM.

Given FMMs include both a single categorical latent variable and one or more continuous latent variables, researchers typically use FMMs when they suspect and anticipate latent classes in the population that differ in their factor parameters (i.e., there is not measurement invariance across latent classes, but latent classes have different factor means, variances, and covariances). It is possible to use FMMs to test for measurement invariance among unknown groups and for differences in factor parameters between groups. These multiple purposes of FMMs are depicted in the path diagram for a general FMM (see Figure 1), where multiple arrows are drawn from the latent categorical variable to both the structural components (factor means and variances) and measurement components (loadings and intercepts) of the model. In other words, the latent categorical variable (or latent class membership) influences both the structural and measurement parameters in the model. For this reason, the FMM does not assume measurement invariance across all latent classes (Masyn, Henderson, & Greenbaum, 2010). However, there is a special case of the FMM in which measurement invariance can be assumed via parameter constraints.

**Mixture Factor Models Versus Factor Mixture Models.** In situations where researchers believe latent classes do not differ in their measurement parameters, they should use a special case of the FMM known as the mixture factor model (MFM). Simply put, MFMs are a constrained version of the FMM where measurement invariance holds across classes (see Figure 2). The only difference between Figure 1 and Figure 2 is the dashed arrows from the latent categorical variables to the measurement components of the model have been removed. Here, the latent categorical variable influences only the structural parameters in the model. Thus, a clear distinction between the FMM and MFM is the assumption of measurement invariance. One commonly used MFM is known as the semi-parametric factor model (Masyn, Henderson, & Greenbaum, 2010).

***Semi-Parametric Factor Model (SP-FM).*** In the factor analysis model, the distribution underlying the latent factors is assumed to be either univariate or multivariate normal. The conventional normality assumption for the latent factors is relaxed in the SP-FM. In the SP-FM, latent classes with a normal distribution are assumed to underlie the population. When multiple class distributions are "mixed" together, they form some type of non-normal overall population distribution. Thus, unlike in factor analysis, the overall population distribution of the latent factors can be non-normal, consisting of a number of underlying normal distributions that each represent a latent class. For instance, two normally distributed latent classes may form an overall positively skewed or bimodal population distribution. The SP-FM allows the researcher to model factor distributions that are not normally distributed (Masyn, Henderson, & Greenbaum, 2010; Pastor & Gagné, 2013)

The only parameters allowed to vary across classes in the SP-FM are related to the distribution of the latent factors (factor means and factor variances). There is strict measurement invariance in the SP-FM allowing factor means and variances to be comparable across individuals regardless of latent class membership (Masyn, Henderson, & Greenbaum, 2010). As with any mixture model, the researcher may also choose to constrain any parameters when estimating the SP-FM. This thesis project applies various SP-FMs to education change score data. Specifics of these SP-FMs are provided in the data analysis section. For the remainder of this literature review, best practices for mixture modeling are discussed in terms of the more general FMM.

**Constructing a Change Score Factor Mixture Model.** The construction of a FMM appears to be unclear – that is, different researchers tend to construct FMMs differently. In many instances (e.g., Lubke et al., 2007; Muthén, Asparouhov, & Rebollo, 2006), researchers analyzed their data using mixture modeling, factor analysis, and the combination of the two; then, researchers compared model fit among all models. While this is helpful in determining the best fitting model, it does not help the researcher determine the number of classes or factors to test. Thus, in a recent article, Clark and colleagues (2013) proposed a strategy for constructing a FMM through a four-step process, which I briefly discuss below (see Figure 3).

Prior to estimating a FMM, researchers should fit a variety of mixture models and factor models that increase in the number of classes and factors, respectively. The goal is to determine the best fitting mixture model and factor model. The best fitting mixture and factor models are suggested to serve as comparisons to the final FMM in the last steps of analysis. Specifically, the comparisons will aid in determining if it is necessary to include both factors and classes in the model. If the best fitting mixture model or factor model fits the data as well as the final FMM, the inclusion of both factors and classes may not be necessary (Clark et al., 2013).

After determining the best fitting factor and mixture models, the researcher should start by fitting a FMM with two classes and one factor, increasing the number of classes in subsequent models (step 1). Next, the researcher should fit a FMM with two classes and two factors and again increase the number of classes in subsequent models (step 2). This pattern of model fit is repeated until the combination of classes and factors is equal to the number of classes from the best fitting mixture model and the number of factors from the best fitting factor model (step 3). This iterative process should be conducted and applied to all types of FMMs (e.g., SP-FM). In the final step, the best FMM for the data is selected using a variety of model fit indices. This selected model is also compared to the best fitting mixture and factor models (step 4; Clark et al., 2013).

Although this four-step process is ideal to use when constructing a change score FMM, researchers should keep in mind the substantive research and theory behind the construct. Final model selection should be based on both statistical evidence and substantive research surrounding the construct(s). For instance, it may be that a model with two classes and two factors statistically fits the data best. However, previous research indicates the change construct is unidimensional. In this case, choosing a model with two classes and one factor might be more appropriate as it is a more reasonable representation of the change construct, even though it might not be the best fitting model statistically (Clark et al., 2013).

**Evaluating Model Fit.** Similar to mixture models of observed scores, the -2*LL* value and ICs (e.g., BIC, SSABIC, and AIC) are also used to evaluate the fit of a FMM. Lower values of -2*LL* and ICs indicate a good fit to the data. In addition to examining the -2*LL* value, ICs are reported because they penalize the -2*LL* for model complexity and sometimes sample size. Given FMMs can differ in both the number of classes and in parameterization, statistical indices used to compare across models is dependent on the kinds of models being compared. For models that have the same parameterizations, but differ in the number of classes, researchers should use BIC, SABIC, and LMR to compare models. For models that differ in the number of classes and parameterizations, researchers should rely on BIC and SABIC to compare models. The LMR test is not appropriate to use in this case because it should only be used to compare models that differ in the number of classes but have the same parameterization. Finally, for models with the same number of classes and nested parameterization, measures of information criteria (IC) can be used for model comparison along with the traditional likelihood ratio test (Pastor & Gagné, 2013; Tofighi & Enders, 2007)

**Validating latent change classes.** Because classes modeled in FMMs are latent or unobservable, identified classes should be validated. Latent change classes should be related to other variables in expected ways supported by previous research and theory. One common technique used to acquire validity evidence is the classify-analyze method (Clogg, 1995). With this technique, individuals are first "classified" into latent classes based on their highest posterior probability (modal assignment). The latent classes are then treated as known groups and used in other statistical models (e.g., ANOVA and regression) to examine relationships between class membership and validity variables (e.g., correlations and semi-partial correlations). However, the classify-analyze method often leads to attenuation of these estimates (Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010). Additionally, this method does not take into account classification accuracy of the FMM. If the model does not have near perfect classification accuracy, there is a high chance for individuals to be assigned the incorrect class, which could ultimately result in invalid conclusions about the classes (Clark, 2010; Pastor & Gagné, 2013).

To account for classification accuracy, other methods have been proposed to determine the validity of a classification solution. One approach is the pseudo-class draws method (Lanza, Tan, & Bray, 2013). Individuals are assigned to a class based on random draws from their posterior probability distributions. After each draw (or iteration), subsequent analysis is performed on group memberships and the validity variables (i.e., correlations and differences in means); results are combined across draws.

The subsequent analysis is performed 20 times rather than once and thus, results are more trustworthy (Asparouhov & Muthén, 2013). Another approach involves the inclusion of correlates in the FMM. This is called a single-step approach. In this process, a researcher could choose to include the correlates as either predictors or outcomes of the latent class variable. Researchers should provide a clear justification for the inclusion of correlates, particularly their specified role in the model. The FMMs solution can change depending on which correlates are included because they are involved in the estimation of model parameters. Thus, researchers are advised to consider extant research when specifying correlates and their roles in FMMs (Pastor & Gagné, 2013).

Given the disadvantages of the classify-analyze method, researchers interested in gathering validity evidence should use the pseudo-class draws approach (Lanza et al., 2013; Asparouhov & Muthén, 2013). This approach eliminates the need to incorporate correlates in the mixture model and, in turn, the need to accurately identify correlates of the latent class variable. Note that there are many other approaches one can take in gathering validity evidence for a mixture model or FMM solution. This is still an active area of research (see Clark, 2010; Petras & Masyn, 2010).

**Applied Example**

Sense of identity has become an increasingly important construct in higher education, particularly in relation to student academic performance. For example, sense of identity has been linked to greater academic performance and is a significant predictor of GPA over and above personality traits such as the Big Five (Lounsbury, Huffstetler, Leong, & Gibson, 2005). There are two mechanisms through which sense of identity may influence academic performance: confidence and motivation. In a small sample of college students, Aston, Baran, Brownfield, and Smith (2013) found a positive association between several aspects of identity development and academic confidence, which in turn had a positive relationship with GPA. In another sample of college students, Faye and Sharp (2008) found significant positive correlations between identity development, competence, and academic motivation. Identity development is positively related to competence and as a result, increases academic motivation.

These empirical results suggest that sense of identity plays a critical and important role for students in higher education and thus, should be measured by institutions and colleges. One scale that is often used to measure students' sense of identity is the Sense of Identity Scale (Lounsbury & Gibson, 2004). There has been limited research on change scores from the Sense of Identity Scale. For the applied example, change score factor mixture modeling was conducted on change scores from the Sense of Identity Scale to demonstrate the utility of the technique (Lounsbury & Gibson, 2004). In the next few sections, I will provide a brief theoretical background of sense of identity along with a literature review of past research on change scores from the Sense of Identity Scale.

**Theoretical Background.** In his book, Childhood and Society, Erikson (1978) conceptualized development as a lifespan model containing eight stages. He proposed that the individual would experience the eight stages as he or she progresses in life. Associated with each stage is a crisis the individual must overcome in order to progress onto the next stage. Based on Erikson's identity theory, individuals who successfully complete each stage are rewarded with positive life consequences. Meanwhile, those who fail to complete a stage and thus are unable to progress to the next stage are left with a poor sense of self and other negative consequences.

In the fifth stage of life development, which Erikson called “identity vs. role confusion,” the construct of identity is introduced. During this stage, the individual is considered an adult and is expected to make adult decisions. In addition to this, individuals also gain a strong sense of awareness of who they are and how others perceive them. An individual must balance between developing a stronger sense of personal identity, while at the same time, becoming aware of what others might think of them. This is known as the identity crisis (1978). The individual could experience the identity crisis both internally and externally. Internally, the individual must come to a sense of who they are. Externally, the individual must come to terms with how others perceive them. Factors contributing to the external and internal components of the identity crisis are prevalent in higher education.

For instance, college students are often challenged with making choices that may impact their sense of identity in both the short- and long-term. These may include choosing a field of study and/or what extracurricular organizations to join. In additional to these internal decisions that could shape identity, college students often are placed in diverse social environments where they are exposed to different types of peer groups and may meet individuals with values different from their own. These interactions could also shape identity.

**Previous Research.** Recent research on the Sense of Identity Scale, which is assumed to be unidimensional cross-sectionally, found change scores from the scale best represented with two correlated change score factors. Specifically, Ong and Erbacher (2016) conducted EFA and CFA on change scores from the Sense of Identity scale using two independent samples of students from a mid-sized public university. These researchers identified two change score factors, one representing change in sense of self and purpose and the other factor representing change in morals or beliefs (see Table 1). The change factors differentially predicted academic success such that change in sense of self and purpose significantly predicted students' GPA while change in morals or beliefs did not. These findings highlight two advantages to factor analyzing change scores. First, change score factors are somewhat independent of cross-sectional factors. For example, change in sense of identity is best represented with two correlated factors, whereas extant literature assumes sense of identity is one factor at any single occasion. Second, change score factors may have different relationships with other constructs than do cross-sectional factors. For instance, Ong and Erbacher (2016) found that change in sense of self and purpose is more closely related to academic success than is change in morals or beliefs. Thus, future studies should aim to further examine change in sense of self and purpose in academic contexts.

**Present Study.** The goal of the present study was to extend the findings of Ong and Erbacher (2016). In the current study, I further examine sense of identity change score factors to explore potential latent classes underlying factor scores. Results have implications for researchers, particularly those in the field of higher education. Identifying latent classes underlying dimensions of change in sense of identity provides greater insight into the relationship between GPA and sense of identity (Lounsbury et al., 2005). For example, one group of students may have a higher average score on the change in morals and beliefs factor, but a lower or even a negative average score on the change in sense of self and purpose factor. Meanwhile, another group of students may have high average scores on both change score factors. If students do exhibit different patterns of change across time, identifying these groups would allow researchers to answer more complex questions regarding student development of a sense of identity. For instance, how do different patterns of change relate to academic success? Is one group's pattern of change more conducive to academic success than others? The application of FMMs to change scores is imperative to answering these types of questions.

Conceptually, there are a number of reasons why different groups of students may demonstrate different patterns of change in sense of identity. According to Waterman (1982), college is a period where identity formation is most salient. College students are asked to make a variety of challenging choices that may impact their sense of identity. For example, students must declare a major field of study at the beginning of their college career. Students declaring a major who are sure of their decision may have a stronger sense of self and purpose than those who did not initially. Therefore, these students may change less or remain more stable across time in terms of their sense of self and purpose. Meanwhile, students who did not declare a major may exhibit more growth in their sense of self and purpose as they progress through college and solidify their career choice and major. Thus, we would expect patterns of change on sense of self and purpose to differ between these two groups. In addition, sense of identity has been significantly correlated with a variety of social and college behaviors (e.g., talking with other people, engagement in religious activities, and talking with professors outside of class; Lounsbury, Richardson, Saudargas, & Levy, 2008). Although this work was cross-sectional, the relationship between change in sense of identity and social behaviors could be inferred. For instance, students who change more positively in their sense of self and purpose may be more likely to engage in a variety of social behaviors such as meeting new groups of students or being heavily involved with their church. As a function of strong development in sense of self and purpose, they are more likely to be in positions where their morals and beliefs are challenged. Consequently, their beliefs and morals may change more across time.

**Research Questions.** Given these theoretical reasons why different groups of students may exhibit different patterns of change in the two change factors found by Ong and Erbacher (2016), the current study addressed the following research questions:

1. Are there latent classes that underlie change score factors from the Sense of Identity Scale?
2. Do latent classes underlying change score factors differ in latent means across classes?
3. Do latent classes underlying change score factors differ in latent variances and covariances across classes and/or within classes?
4. Does the final FMM fit the data significantly better than the best fitting mixture model and factor model?

# CHAPTER THREE

## Method

### Participants and Procedure

Study participants consisted of undergraduate students at a mid-sized public university on the east coast of the United States. Data were collected during an institution-wide, mandatory testing session known as Assessment Day. Students are required to participant in Assessment Day twice during their academic career: once as incoming freshmen and again during the spring semester of their sophomore year, once they have completed 45 to 70 credit hours. On Assessment Day, students are exempt from their classes and randomly assigned to rooms to complete a number of non-cognitive and cognitive measures, lasting two to three hours. In each assessment room, two trained proctors are present to ensure the quality of testing conditions. All room proctors follow a strict protocol (i.e., read the same instructions and follow the same timeline). Test administration is standardized across all rooms and testing sessions.

Study participants included two cohorts of students who completed the 8-item Sense of Identity Scale (Lounsbury & Gibson, 2004) as incoming freshmen (pre-test) and again as second semester sophomores (post-test). Students were separated into two samples by cohort: an exploratory sample and a validation sample. The exploratory sample (first cohort) consisted of 2,187 students with complete data at both occasions on the Sense of Identity Scale. Participants were predominantly female (65%) and Caucasian (88%), which is representative of the university's demographics. The average age of students at pre-test was 18.4 (*SD* = .37) years and the average age of students at post-test was 19.8 (*SD* = .37) years. The validation sample (second cohort) consisted of 706 students with complete data at both occasions. Demographic characteristics again indicate a predominantly female (67.3%) and Caucasian (80.3%) sample, similar to the exploratory sample. The average age of students for the validation sample at pre-test was 18.4 (*SD* = .38) years and the average age of students at post-test was 19.9 (*SD* = .38) years.

**Measures**

**Sense of Identity Scale.** The Sense of Identity Scale was administered to measure students' sense of identity. The Sense of Identity Scale, a subscale of the Adolescent Personal Style Inventory (APSI: Lounsbury & Gibson, 2004), contains eight non-cognitive items about various aspects of sense of identity (see Table 1). According to the authors, Lounsbury and Gibson (2004), sense of identity is conceptualized as "knowing one's self and where one is headed in life, having a core set of beliefs and values that guide decisions and actions; and having a sense of purpose" (p. 3). On the Sense of Identity Scale, students are asked to respond to each of the eight items on a 5-point Likert-type scale (1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Neutral*, 4 = *Agree*, 5 = *Strongly Agree*). One item on the scale is reversed scored. After reverse scoring, higher scores imply a stronger sense of identity. Recent studies support the reliability of the scores obtained from the scale (alpha > .70; Lounsbury et al., 2005; Lounsbury, Levy, Leong, & Gibson, 2007). However, the psychometric properties of the scores from the original study were not reported.

**Computing Change Scores.** Change scores were calculated by subtracting pre-test scores from post-test scores on each item of the Sense of Identity Scale. Positive change scores indicated a growth in sense of identity, whereas negative change scores indicated a decline in sense of identity across the two occasions. I calculated reliability for the change scores using Williams and Zimmerman's (1996) formula. The reliability of the change scores was .68 for the exploratory sample and .70 for the validation sample.

**Preliminary Data Analytic Plan**

Prior to the primary data analysis, the best fitting factor model and mixture model were determined (see Figure 3). All models were estimated using Full Information Maximum Likelihood Estimation (FIML) in Mplus Version 7.11 (Muthén & Muthén, 1998-2012).

**Factor Model.** Previous research has been conducted on the factor structure of change scores from the Sense of Identity Scale (Lounsbury & Gibson, 2004). Ong and Erbacher (2016) conducted EFA and CFA of change scores from the scale and found the scores were best represented by a two-factor model. Items 1, 2, 6, 7, and 8 of the Sense of Identity Scale loaded onto one change score factor and items 3 and 5 loaded onto another change score factor (see Table 1). The two change score factors were interpreted as representing two different aspects of development in sense of identity. Change scores from items 1, 2, 6, 7, and 8 represented change in sense of self and purpose and change scores on items 3 and 5 represented change in morals and beliefs. These two change score factors were validated on another independent sample (Ong & Erbacher, 2016). This work includes the same data sets (i.e., the exploratory and validation samples) as in the present study. Thus, the two-factor model was identified as the best fitting factor model.

**Mixture Model.** Mixture modeling was conducted on the eight change score variables (one variable for each item) to examine the number of underlying latent classes. Although there are a variety of mixture model specifications, three common mixture models (Model A, B, and C) were fit to the data, each differing in their parameterizations. For all three models, means were freely estimated across and within classes. In model A, variances were freely estimated within classes, but constrained to be equal across classes; covariances were fixed to zero across and within classes. In Model B, variances were freely estimated, but constrained to be equal across classes and within classes (i.e., all item variances were nonzero but were fixed to be equal within and across classes); covariances were freely estimated within classes, but constrained to be equal across classes. In Model C, variances were freely estimated within and across classes; covariances were fixed to zero across classes and within classes. For each parameterization, a one-class model was initially fit to the data and subsequent models (increasing by one in the number of classes) were analyzed. All model parameterizations were tested with up to five classes (or mixtures) or until estimation issues were encountered. Given the sample size, classes that emerge after the fifth class likely include only a small number of individuals. Thus, the number of classes was capped at five. I estimated all mixture models using a random start value of 500 and final stage optimization value of 150.

***Model Fit.*** A number of fit indices aided determination of the best fitting model. The -2 log-likelihood (-2*LL*) value was examined for each mixture model to evaluate the likelihood of the data given the estimated model parameters. Lower values of -2*LL* indicated superior model fit. Additionally, because values of -2*LL* will always be lower for models with more classes, two information criteria were also used to determine the best fitting mixture model. The Bayesian Information Criterion (BIC; Schwarz, 1978) and the Sample Size Adjusted BIC (SSABIC; Sclove, 1987) were used in the current study. Extant literature indicates these indices are more reliable with mixture modeling techniques than other information criteria (IC) measures such as Akaike Information Criteria (AIC; Henson et al., 2007; Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2007; Yang, 2006). IC values were also compared across models, each differing in the number of classes. Smaller IC values indicated better model fit and thus, the model associated with the smallest values was claimed as superior (or best fitting). Lastly, the Lo-Mendell-Rubin (LMR) test was used to compare nested *k* and *k*-1 class models with the same parameterization.

## Primary Data Analytic Plan

**Factor Mixture Modeling.** Given the focus of my study was to examine potential differences in factor means and factor variances across latent classes, I assumed measurement invariance across latent classes. I had to make this assumption because factor means and factor variances are only comparable when the assumption of measurement invariance is met (Millsap, 2011). For this reason, I chose to fit a series of FMMs, specifically SP-FMs, to change scores. The SP-FM constrains all measurement model parameters to be the same across latent classes and thus, assumes the measurement model fits data from all people equally well, regardless of latent class membership.

I estimated four different parameterizations of the SP-FM using the sense of identity (SOI) change score data (see Figure 4). In all four model parameterizations, the measurement model parameters (e.g., item intercepts, loadings, and error variances) were freely estimated, but constrained to be equal across classes, excepting model identification constraints. This was done intentionally to meet the assumption of measurement invariance. However, the structural model parameters (e.g., factor means, factor variances, and factor covariances) were freely estimated and allowed to vary across classes, within classes, or both, depending on the model. For all model parameterizations, I fit a two-class, one-factor model to the data and increased the number of classes in subsequent models. Then, I estimated a two-class, two-factor model to the data and in subsequent models increased the number of classes. I continued this pattern of model fit until the number of classes and factors was equivalent to the number of classes from the best fitting mixture model and the number of factors from the best fitting mixture model (Clark et al., 2013). I estimated all FMMs using a random start value of 500 and final stage optimization value of 150. For models that failed to converge, I increased these values in order to find convergence.

***Model A.*** In Model A, factor means were freely estimated across and within classes. Factor variances were freely estimated, but constrained across and within classes (i.e., factor variances were constrained to be equal across factors). Factor covariances were freely estimated, but constrained across classes. This model was used as a baseline model for comparison to other models in the study.

***Model B.*** In Model B, factor means were freely estimated across and within classes. Factor variances were freely estimated within classes, but constrained across classes. Factor covariances were freely estimated within classes, but constrained across classes. Change score factors in this model could differ in their variances within classes. For example, one change score factor may be more stable compared to the other (i.e., have less variance) or vice versa within classes. Thus, this model helped identify differences in factor variances within classes.

***Model C.*** In Model C, factor means were freely estimated across and within classes. Factor variances were freely estimated across classes, but constrained to be equal within classes. Factor covariances were freely estimated across classes, but constrained within classes. Change score factors may also differ in their variances across classes as well. For instance, one class may have more variance on all change factors whereas another class may have less variance on all change factors. Model C allowed me to explore this possibility.

***Model D.*** In Model D, factor means, variances, and covariances were freely estimated across and within classes. It is possible for change score factors to differ in their variances within classes and subsequently across classes as well. For example, a class may have more variance on one change score factor compared to the other factor. However, this pattern may not be consistent across classes. Model D allowed a test of this possibility.

**Model Fit.** Similar to evaluating fit for mixture models, values of -2*LL* and IC indices (e.g., BIC and SSABIC) were examined to determine model fit to the data. Additionally, the Lo-Mendell-Rubin (LMR) test was also used to compare certain models (i.e., models with the same parameterizations and *k* vs. *k*-1 classes). Both theory and model fit were taken into account when determining and selecting the best fitting FMM.

**Validity.** An important step in conducting FMMs is to validate the classes obtained from the solutions. To increase confidence in claiming that the classes could be distinguished based on their patterns of development in sense of identity, I examined the relationship between class membership and GPA.

***GPA.*** Previous research indicates a relationship between sense of identity and academic success (Lounsbury et al., 2005). Individuals reporting a strong sense of identity tend to achieve a higher GPA than those lower in sense of identity. Thus, individuals in a latent change class with higher means on the two change score factors (i.e., exhibiting more positive growth across time in sense of identity) were expected to have higher GPA than those in other classes.

# CHAPTER FOUR

## Results

The first three research questions in my study pertained to the FMM results. Thus, I will focus on those results first. To determine the best fitting FMM, I followed the four-step process recommended by Clark et al. (2013) for constructing a FMM. Prior to fitting any FMM, I determined the best fitting mixture model and factor model to the data. The best fitting mixture model and factor analysis results informed the final number of classes and factors to include in the FMM analyses. First, I report on the mixture model results. Second, I comment on the best fitting factor model. Then, I describe the best fitting FMM. In the end, I interpret model results across these three methods to answer each of my four research questions.

### Best Fitting Mixture Model

To identify the best fitting mixture model, I estimated one-, two-, three-, four-, and five-class solutions for each of the three mixture model parameterizations (A, B, and C). Fit indices for the mixture models are included in Table 2. The two-, three-, four-, and five-class solutions for Model C (variances freely estimated, but constrained to be equal across classes and within classes and covariances freely estimated within classes, but constrained to be equal across classes) did not converge. Thus, I did not interpret these solutions because the results were deemed untrustworthy.

Out of the converged models, the four and five-class Model B solutions had the lowest IC values (BIC and SSA-BIC; see Table 2). These two models also had the highest entropy values. The LMR test was statistically significant ($p = .043$) for the four-class model, which indicated the four-class Model B provided significantly better fit to the data than the three-class Model B (Tofighi & Enders, 2007). The LMR test for the five-class Model B was not statistically significant ($p$ = .703). Given this strong support for the four-class Model B, I championed the four-class Model B as the best fitting mixture model and examined FMMs with up to four classes (Clark et al., 2013).

**Best Fitting Factor Model**

Given Ong and Erbacher (2016) conducted EFA and CFA on the same data set used in my study, I used results from their study to inform the best fitting factor model. Ong and Erbacher (2016) estimated a one-factor and two-factor model with change score data and compared the fit via a variety of indices (e.g., CFA, SRMR, and RMSEA). Fit indices for the one- and two-factor models from their study are presented in Table 3. Their results indicated the one-factor model did not fit the data well, $\chi^2(20) = 739.064$, $p < .001$, CFI = .767, SRMR = 0.069, RMSEA = 0.129, 90% CI = .121 to .137. The two-factor model provided better fit to the data, $\chi^2(19) = 298.112$, $p < .001$, CFI = .910, SRMR = 0.044, RMSEA = 0.082, 90% CI = .074 to .091, and fit the data statistically significantly better than the one-factor model, $\Delta\chi^2(1) = 440.95$, $p < .001$. Based on these findings, I championed the two-factor model as the best fitting factor model and examined FMMs with up to two factors (Clark et al., 2013).

**Best Fitting Factor Mixture Model**

Because the two-factor model and the four-class mixture model were championed as the best fitting models, I estimated FMMs with one to two factors and one to four classes for each of the four modeling parameterizations (A, B, C and D; Clark et al, 2013). Fit indices for the FMMs are provided in Table 4. Out of the 18 models, six models did not converge[1]. Thus, only 12 models were interpretable.

When comparing models using only IC values, two-factor FMMs, regardless of the number of classes, fit the data better than one-factor FMMs. Additionally, aside from the 4-class, 2-factor Model B, all other models had higher entropy values with two factors compared to one. These findings indicated the second factor is essential. Out of all the two-factor FMMs, the two-class Model B (factor means freely estimated; factor variances freely estimated, but constrained across class; and factor covariances freely estimated) provided an overall good fit to the data compared to the other models. No other models had relatively low IC values, a significant LMR *p*-value, and a high entropy value. For example, the four-class, two-factor Model A has lower IC values and a LMR *p*-value less than .10, but the entropy value is much worse than the two-class, two-factor Model B. Although the three-class, two-factor Model B and four-class, two-factor Model B both have lower IC values and higher entropy values than the two-class, two-factor Model B, neither have a statistically significant LMR test result. The two-class, two-factor Model B seems to be the best balance between all the fit indices of interest.

In addition to comparing models via fit indices, I also considered the practicality of the results for each model. For instance, many of the models with three or four classes included classes with very small mixing proportions (i.e., containing a small number of students). In terms of information gain, these models did not provide us with much additional information about students in our sample. Classes identified under the three- and four-class models are likely to be a result of sampling error and may not actually

[1] Models did not converge despite using a random start value of 4000.

exist in the population. The two classes obtained using the two-class, two-factor Model B contained a larger number of students compared to the three-class, two-factor Model B and the four-class, two-factor Model B. Taking together the fit indices of the models and practicality of the results, I championed the two-class, two-factor Model B as the best fitting FMM in the exploratory sample.

Parameter estimates for the two-class, two-factor Model B are displayed in Table 5. The top part of the table shows the factor loadings and item intercepts for the two change score factors; the bottom part of the table shows the factor means and factor variances in each class. The factor loadings and item intercepts are identical across classes[2] and a SP-FM specification was used, which forced measurement invariance across classes. The eight change scores all loaded significantly onto their respective change factor (> .30). Change scores on items 3 and 5 loaded onto the morals and beliefs change factor and change scores on items 1, 2, 4, 6, 7, and 8 loaded onto the sense of self and purpose change factor.

I used results obtained from the two-class, two-factor Model B solution to answer the first three research questions in my study.

**Research Question One: Are there latent classes that underlie change score factors from the Sense of Identity Scale?**

Since the two-class, two-factor Model B fit the data the best, I concluded there are two latent classes underlying the two change score factors from the Sense of Identity Scale. Class One consisted of 50 students and Class Two consisted of 2,127 students.

---

[2] Loadings for change scores on items 1 (factor 1) and 3 (factor 2) were fixed to one for identification purposes and to set the scale of the change score factor. Similarly, intercepts for change scores on items 1 and 3 were fixed to zero in order to freely estimate factor means.

**Research Question Two: Do latent classes underlying change score factors differ in latent means across classes?**

Class One had an estimated factor mean of -0.591 on the morals and beliefs change factor and an estimated factor mean of -1.848 on the sense of self and purpose change factor. Class Two had an estimated factor mean of 0.030 on the morals and beliefs change factor and an estimated factor mean of -0.013 on the sense of self and purpose change factor. In comparison, students in Class One (50 students) had lower factor means on both change factors than students in Class Two (2,127 students). One possible interpretation of the classes is that students in Class One were more fluid on development of sense of identity than students in Class Two, because the factor means of Class One were farther from zero or more extreme than the factor means of Class Two.

Although factor means are often more desirable than observed means, given no measurement error is associated with latent scores, it may be hard to interpret factor means when using change scores. For instance, the scale of the change factor may not align with the raw change score metric and so a positive or negative factor mean value does not necessary indicate a growth or decline in the construct (or in this case, sense of identity). Therefore, I also examined the observed class means on the two change factors to further investigate class mean differences. The observed class means were computed using modal assignment, where I assigned students to one of the two classes based on their highest posterior probability. For each class, I computed the total score for the change factors and calculated the means. These class means are presented in Table 6.

Consistent with factor means, Class One had overall lower observed means on the two change factors than Class Two. More specifically, Class Two had observed means that were near zero on both change factors. This indicated, on average, students in this class changed minimally in their sense of self and purpose and morals and beliefs across time. Class One, however, had observed means that were below zero on both change factors. This finding suggests students in Class One declined more in their sense of self and purpose and morals and beliefs across time than students in Class Two. Although students in Class One and Class Two both tended to change minimally on the morals and beliefs change factor, students in Class One reported a greater decline on the sense of self and purpose change factor than students in Class Two. Students in Class Two had an observed mean of -0.34 and those in Class One had an observed mean of -11.46 on the sense of self and purpose change factor. Class One decreased dramatically on sense of self and purpose compared to Class Two. There was over a 10 point difference in observed means. These change patterns based on observed means are consisted with the interpretation of the factor means above; Class One is fluid in their sense of self and purpose and Class Two is more stable.

**Research Question Three: Do latent classes underlying change score factors differ in latent variances and covariances across and/or within classes?**

Recall, I used the four FMM parameterizations (A, B, C, and D) to test differences in factor variances across and within classes. Each parameterization constrained factor variances in a different way. The estimated factor variances based on the two-class, two-factor Model B can be seen in Table 5. Note the two-class, two-factor Model B allowed factor variances to be freely estimated within class, but constrained the factor variances to be the same across classes. This constraint is why the factor variances have the same values across both classes in Table 5. Because the two-class, two-factor

Model B provided better fit to the data than the two-class, two-factor Model A (see Table 4), the two change factors differed in their variances within classes. The morals and beliefs change factor had a slightly higher variance than the sense of self and purpose change factor. In other words, there was more variability in students' change in moral and beliefs factor scores than in their change in sense of self and purpose factor scores. Because of convergence issues with the two-class, two-factor Model C and Model D, I was unable to provide further interpretations about the factor variances.

**Research Question Four: Does the final FMM fit the data significantly better than the best fitting mixture model and factor model?**

To address my final research question, I fit the best fitting mixture, factor, and factor mixture model obtained from the exploratory sample to the validation sample ($N$ = 706). After comparing the IC values (BIC and SSA-BIC) of the best fitting mixture, factor, and FMM (Table 7), it was clear the factor model did not fit the data better than the FMM and the mixture model. When comparing BIC values alone, the FMM had the lowest BIC value (14972), which suggested the best fitting model is the FMM. Even though the FMM had the lowest BIC value, the mixture model had the lowest SSA-BIC value (14788) compared to the other two models. Given the discrepancy between the BIC and SSA-BIC values, it was inconclusive on whether the FMM or the mixture model fit the data better. Practically speaking, however, there were a few reasons why the FMM should be championed as the best fitting model.

The difference in BIC and SSA-BIC values (Table 7) between the FMM and the mixture model were minimal. Recall BIC penalizes -2*LL* for the number of estimated parameters (i.e., model complexity). In the mixture model, 71 parameters were estimated and in the FMM, 28 parameters were estimated. The difference in parameters estimated between the two models is 43. In terms of model fit and their number of estimated parameters, the mixture model yielded roughly similar model fit as the FMM even though 43 additional parameters were estimated. The FMM fit the data nearly as well as the mixture model with far fewer estimated parameters. More so, when examining the number of students in each class under the mixture model (four-class Model B), three out of the four classes consisted of fewer than 20 students: Class One is the smallest-sized class with only 3 students; Class Two is made up of 14 students; Class Three, the largest-sized class, consisted of 677 students; and Class Four is made of 19 students. The mixture model fit provided similar fit to the FMM, but the number of students assigned to each class is troublesome, especially with the sizes of Class One and Class Two. Class One and Class Two may not represent qualitatively different groups of students in the population, which is a concern when using mixture modeling.

For the reasons listed above, I championed the FMM (two-class, two-factor Model B) as the overall best fitting model and further interpreted the model below.

**Two-Class, Two-Factor Model B.** Parameter estimates for the two-class, two-factor Model B fit to the validation sample are presented in Table 8. Class One consisted of 664 students and Class Two consisted of 39 students.

***Factor Means.*** Class One had an estimated factor mean of -0.07 on the morals and beliefs change factor and an estimated factor mean of -0.05 on the sense of self and purpose change factor. Class Two had an estimated factor mean of 1.07 on the morals and beliefs change factor and an estimated factor mean of -0.08 on the sense of self and purpose change factor. On the sense of self and purpose change factor, students in Class

One had a mean factor score similar to students in Class Two. However, students in Class Two had a higher factor mean on the morals and beliefs change factor than students in Class One, suggesting more growth than Class One.

These results were also consistent when examining the observed class means (calculated using modal assignment) reported in Table 9. Class One had observed means near zero on both change factors, which indicated that students in this class changed minimally in their sense of self and purpose and morals and beliefs. Class Two, however, had observed means that were both greater than zero, suggesting students in this class grew in their sense of self and purpose and morals and beliefs. Although students in both classes tended to both change minimally on the sense of self and purpose change factor, students in Class Two reported greater growth on the two-item morals and beliefs change factor ($M$ = 3.10) than students in Class One ($M$ = -0.08).

***Factor Variances.*** Factor variances for both classes are located in Table 8. These variances were freely estimated within class, but constrained to be equal across classes. The morals and belief change factor had slightly lower variance than the sense of self and purpose change factor. In other words, there was more variability in change of sense of self and purpose factor scores than in change in morals and beliefs factor scores.

## Validity of the Classes

To examine the extent to which my findings aligned with my expectations regarding sense of identity classes and academic achievement, I conducted validity analysis on the two classes using GPA. I entered GPA in the factor mixture model analyses in Mplus (Muthén & Muthén, 1998-2012) as an auxiliary variable for the exploratory sample and validation sample. Rather than using modal assignment (as I did to calculate the observed class means on the two change factors), Mplus classified students into one of two classes using the Lanza method (i.e., pseudo-class draws method; Lanza et al., 2013; Asparouhov & Muthén, 2013). The validity results using the two-class, two-factor Model B solution are presented in Table 10. In the exploratory sample, Class One and Class Two statistically significantly differed from each other on GPA, $\chi^2(1)= 12.346$, $p < .001$, with Class One exhibiting lower mean GPA. This was expected because students in Class One had lower factor means than students in Class Two. Moreover, students in Class One decreased on sense of self and purpose by a large amount compared to students in Class Two. Thus, students who decreased in sense of self and purpose had lower GPAs than those who remained stable in their sense of self and purpose.

In the validation sample, Class One had higher GPA than Class Two (Table 10). However, the two classes were not statistically significantly different on GPA, $\chi^2(1)= 1.590$, $p = .207$. The relationship between class membership and GPA was not in the expected direction. Students in Class Two, who had a higher factor mean on the morals and beliefs (indicating more growth), were expected to have a higher GPA than students in Class Two. However, this was inconsistent with the empirical results.

In addition to exploring the relationship between GPA and class membership, I also examined potential outliers in the data. It is possible for the fluid classes (consisting of relatively smaller number of students than the stable classes) found across the two samples to emerge as a result of containing mostly of outliers. To check for multivariate outliers, I examined Mahalanobis distance values. In the exploratory sample, 11 out of 50 cases (i.e., students) in Class One and 59 out of 2068 cases in Class Two were deemed as multivariate outliers. In the validation sample, 12 out of 652 cases in Class One and seven out of 32 cases in Class Two were deemed as multivariate outliers. Given these results, the fluid classes (i.e., small-sized classes) did not appear to consist mostly of outliers.

# CHAPTER FIVE

## Discussion

Change scores represent a person's change on a construct across two time points. They are simple to calculate, easy to interpret, and provide insightful information about change. Although researchers have argued against reliability of change scores, many have based their argument on incorrect mathematical bases, as I discussed in Chapter 2. The purpose of this thesis project is to provide a detailed overview of two longitudinal methods that can be used with change scores. The first method is factor analysis of change scores, in which EFA and CFA are used to explore the dimensionality of change or growth in a construct. The second method involves applying a person-centered approach, mixture modeling, to change score factors to uncover latent classes underlying change score factors, known as change score factor mixture modeling. The utility of both methods are demonstrated in this thesis through an applied example using real change score data. The applied example consisted of conducting change score factor mixture modeling on change score data from the Sense of Identity Scale (Lounsbury & Gibson, 2004), obtained from two independent samples of college students at a mid-sized university.

In the following sections, I will summarize and discuss the findings in the applied example. Then, I will highlight the major advantages to both methods as shown with the applied example. Finally, I will discuss novel information that could be obtained through both of these methods.

## Applied Example

There were four research questions in the present study. The first research question regarded the number of latent classes underlying the two change factors from the Sense of Identity Scale (Lounsbury & Gibson, 2004). The second and third research questions prompted comparisons of latent classes on factor means and variances. Finally, the fourth research question prompted comparisons of change score factor mixture modeling results to results obtained from factor analysis of change scores and change score mixture modeling. This final question addressed whether there is any gain in using change score factor mixture modeling over the other two methods.

**Research Question One and Two.** To answer these first two research questions, I determined the best fitting FMM. I fit a variety of FMMs to the data, each differing in the number of classes, factors, and parameterizations, to sense of identity change scores from the exploratory sample. I found the best fitting FMM to be the two-class, two-factor model with parameterization B. This indicated there are two latent classes underlying the Sense of Identity Scale change score factors in the exploratory sample. Class Two consisted of the majority of students in the sample and Class One consisted of a small subset of students. The two classes differed in their factor means on both the sense of self and purpose change score factor and the morals and beliefs change score factor. Overall, Class One had lower factor means than Class Two. These results were corroborated by class means of observed sum scores (i.e., sum of change scores on all items loading on a factor). Class One had change sum scores near zero on morals and beliefs items (i.e., mean of change sum scores is near zero), but had extremely low, negative change sum scores on sense of self and purpose items. In contrast, Class Two had observed sum score means near zero on both sets of items. Thus, I deemed Class One as the fluid class and Class Two as the stable class.

The difference in factor means and observed means across and within classes provided evidence for two change patterns associated with development of sense of identity during the college years. For one group of students, development of sense of identity was stable across time (i.e., no growth or decline in sense of self and purpose and morals and beliefs). Meanwhile, for another group of students, development of sense of identity was fluid. These students on average decreased slightly in morals and beliefs, and decreased substantially in sense of self and purpose across time. This latter decrease was particularly large when examining the observed change sum score class means. Class One was characterized by students who are struggling in developing their sense of identity, particularly in establishing their sense of self and purpose. Students in class one seem to be losing their sense of who they are. Alternatively, Class Two was characterized by students who, regardless of whether they came in with a strong or weak sense of sense of self and morals or beliefs, maintained their sense of identity across time.

The two change patterns indicate development of sense of identity is not same for all students, or at least for students in higher education. Rather, student development of sense of identity could follow one of multiple trajectories, two of which were identified in this thesis.

**Research Question Three**. For both classes, the two change score factors differed in their factor variances. Overall, students tended to vary more in change of morals and beliefs than in change of sense of self and purpose. This was indicated by a higher factor variance on the morals and beliefs change factor. Unfortunately, I was unable to draw further conclusions about the factor variances due to convergence issues with the two-class, two-factor Models C and D. Recall that Model C allowed factor variances to be freely estimated across classes, but constrained to be equal within classes. Model D, the most complex model, allowed factor variances to be freely estimated across and within classes. It is possible that Model C, if it had converged, would fit significantly better than Model B. This would provide evidence that the two change factors have the same factor variances within classes, but classes have different factor variances. In other words, students could vary similarly in change of sense of self and purpose and morals and beliefs; a pattern that may be not consistent across classes (e.g., one class may vary more on both while another may vary less on both). It is also possible for Model D, if it had converged, to fit significantly better than Model B. In this case, variability in change score factor scores would differ both within classes (between factors, one factor has more variability than the other) and across classes.

Although I could not explore the two possibilities listed above statistically, as the models did not converge, the two-class, two-factor Model B provided insightful information on sense of identity development. Recall factor variances were set to the same scale as one of the change score variables. For the sense of self and purpose change score factor, the variance was set to the same scale as change scores on item 1. For the morals and beliefs change score factor, the variance was set to the same scale as change scores on item 3. Overall, there was more variability in the morals and beliefs change factor scores than in sense of self and purpose change factor scores. This could indicate change in morals and beliefs is more fluid than change in sense of self and purpose, as more students can have extreme scores on the morals and beliefs change factor.

**Research Question Four.** As noted in Chapter 2, FMMs are a hybrid between a factor analysis model and mixture model. It is important to evaluate whether the use of FMM is necessary. Is there an advantage or gain in using FMMs to model change scores over more parsimonious models such as factor and mixture models? If the factor model fits as well as or better than the FMM, this would suggest no underlying latent classes exist for the sense of identity change factors, or our data come from a homogenous population.  If the mixture model fits as well as or better than the FMM, this would suggest there are latent classes that underlie change in sense of identity; however, we do not need continuous latent factors to model change in sense of identity, only categorical latent classes. If the FMM fits best, then both continuous and categorical latent variables are needed to adequately capture the dimensionality and heterogeneity in the data. Conceptually speaking, these model comparisons allow for better understanding of sense of identity development. Statistically speaking, these comparisons allow us to determine whether two different types of latent variables (categorical and continuous) in the same model provide a better fit to the data compared to a mixture model with one latent categorical variable or a factor model with one or more latent continuous variables.

For the applied example, comparisons between the factor model, mixture model, and FMM were conducted to answer the fourth research question. Based on results from the exploratory sample, I fit the best fitting factor model, mixture model, and FMM to change scores from the validation sample and compared the fit of each model to one another. The factor model (2-factor) was found to fit worse than both the mixture model and the FMM. Although, the mixture model (four-class Model B) and FMM (two-class, two-factor Model B) were similar based on IC values (e.g., BIC and SSA-BIC), the FMM solution was more desirable from a practical standpoint (e.g., classes obtained using the FMM contained larger numbers of students). Additionally, given approximately equal fit, the FMM required far fewer parameters than the mixture model. Thus, I identified the best fitting model on the validation sample to be the FMM, specifically the two-class, two-factor Model B.

The two classes found in the validation sample differed in factor means. Class Two had a higher factor mean on the morals and beliefs change score factor compared to Class One. This indicated students in Class Two reported more growth in morals and beliefs than students in Class One. The two classes also differed in factor variances within classes as well. There was more variability in change in sense of self and purpose factor scores than in development of morals and beliefs factor scores. Similar to results obtained in the exploratory sample, these differences in factor means and variances support two distinct change patterns associated with development of a sense of identity.

The two classes obtained from the exploratory sample, however, did not share the same change patterns as the two classes obtained in the validation sample. This was apparent when comparing observed class means across the two samples. In the exploratory sample, students in Class Two reported minimal change on both change score factors, and students in Class One, although also reporting minimal change on the sense of self and purpose change score factor, reported a greater decline on the morals and beliefs change score factor. In the validation sample, students in Class One, who were similar to Class Two in the exploratory sample, reported minimal change on both change score factors. Students in Class Two, however, reported a greater increase on the morals and beliefs change score factor than Class One in the validation sample. The discrepancy between change patterns in both classes across the two samples is somewhat troublesome for the validity of these classes.

Although classes found in the exploratory and validation samples produced inconsistent change patterns, there were some promising similarities. Both samples consisted of one class that was stable on the two change factors (i.e., minimal change) and another class that was more fluid on at least one of the two change factors (i.e., changed more on one factor). Thus, the classes in both samples shared similar change characteristics. Moreover, the fluid class contained a small number of students in both samples, suggesting most students tended to be stable in development of sense of identity. The fact that both samples consisted of a fluid class containing only a small group of students provided evidence for two distinct patterns of change in sense of identity.

**Validity of the Classes.** In an attempt to gather validity evidence for the two classes found in the exploratory sample and validation sample, I used GPA as an auxiliary variable in my FMM analyses. I found Class One and Class Two to significantly differ on GPA in the exploratory sample, with Class One having lower mean GPA by 0.20 points. This was consistent with what I expected based on the literature surrounding academic achievement and sense of identity. Compared to Class Two, Class One decreased by a substantial amount in sense of self and purpose. Given previous researchers found sense of identity to be positively related to GPA (Lounsbury et al, 2005), it is not surprising that students who are severely declining on one aspect of sense of identity (Class One) have lower GPA than those who remained stable. This decrease in sense of self and purpose could have had a negative impact on students' academic performance in some way. Thus, change in sense of self and purpose may be an important covariate of students' success in higher education.

For the validation sample, students in Class One and Class Two differ on GPA, but the difference was small (raw difference = 0.08) and not statistically significant. Moreover, this difference was inconsistent with the literature. Students in Class Two had lower GPAs than those in Class One, even though students in Class Two increased more in morals and beliefs than those in Class One across time. Given sense of identity is positively related to GPA, I expected students who increased in morals and beliefs (e.g., Class Two) to have higher GPA than students who did not change as much. This was not the case for this sample.

One possible explanation for this is change in morals and beliefs may not be strongly related to students' academic performance. Thus, unlike in the exploratory sample where students who decreased in sense of self and purpose reported a significantly lower GPA than students who remained stable, the relationship between change in morals and beliefs and GPA may be different. That is, change in morals and beliefs may be less related to academic performance than change in sense of self and purpose. It is also possible for this relationship to be reversed. For example, academic performance may have an influence on change in sense of self and purpose and change in morals and beliefs and this influence may be stronger for sense of self and purpose than morals and beliefs.

**Implications and Future Directions.** The discovery of two latent classes with different change patterns on components of sense of identity has important implications, especially for educational researchers. The two classes found in this study may differ on a number of other outcomes related to academic success that were not included in this research (e.g., graduation rate). For instance, it is possible for one change pattern to be more conducive to academic success than another. Also, identifying classes that differ in their change patterns may help identify students who are struggling to develop a sense of identity. This is especially true for the two classes found in the exploratory sample, where Class One had a large decrease in sense of self and purpose. Students in this class may need additional resources and extra help with developing their sense of identity in college.

The next step is to uncover covariates of each pattern of change in sense of identity. For example, do students in the stable class tend to possess certain characteristics that students in the fluid class do not? If so, what are these characteristics? Additionally, because the two classes found in the exploratory sample were inconsistent with the ones found in the validation sample, further research should explore whether similar classes are supported in other independent samples. Exploring these types of questions will further research on sense of identity and may ultimately lead to a more comprehensive understanding of student sense of identity development in higher education.

**Limitations**. As with any person-centered approach, such as mixture modeling, the classes or groups identified should be interpreted with caution. The identified classes may not represent actual existing groups in the population of interest. In the applied example, a number of constraints were placed on the model (e.g., strict measurement invariance was imposed). Although these constraints allowed factor means and factor variances to be comparable across classes, the assumption of strict measurement invariance was not directly tested. It is possible for factor loadings, item intercepts, and error variances to differ across classes. This was not tested in the applied example but rather it was assumed. If there is not strict measurement invariance across classes, the specific FMM (SP-FM) would inaccurately represent (i.e., misspecify) the data and the results would be untrustworthy.

Additionally, factor mixture modeling will always produce the number of classes specified. The classes should be interpreted using a holistic approach (e.g., statistical fit of the model and practicality of the results) rather than just examining the statistical fit of the model to the data. Moreover, the applied example only used one variable, GPA, for validity evidence. Certainly, this is a limitation. Although the relationship between GPA and class membership did provide some validity evidence for the two classes, additional validity evidence is needed to demonstrate the classes differ both qualitatively and quantitatively. Finally, a number of FMMs failed to converge. This limited the conclusions I could draw about the identified classes.

**The Hidden Truth behind Change Scores**

Although the decision of whether one should factor analyze change scores or cross-sectional scores is dependent on the research question, factor analytic methods are often not applied to change scores due to reliability concerns. However, as I mentioned in Chapter 2, reliability of change scores is affected by a number of factors. Many researchers who have argued against change scores have based their arguments on a much simpler equation of reliability of change scores, where these factors are not taken into account. Thus, under certain circumstances mentioned in Chapter 2, change scores are just as reliable as cross-sectional scores. However, as a result of this misunderstanding, many researchers tend to overlook this method even though novel information could be gained by factor analyzing change scores.

Factor analyzing change scores provides one way of exploring change processes within a construct. Instead of assuming that change in a construct is unidimensional across time, researchers are able to test this assumption by directly factor analyzing the change scores themselves. The change score factors extracted represent different aspects of development in the construct. In the applied example, development of sense of identity was best represented using two change factors, sense of self and purpose and morals and beliefs. However, cross-sectionally, sense of identity is assumed to be unidimensional. The discrepancy between factor structures of change scores and cross-sectional scores highlights the most valuable reason for factor analyzing change scores. Researchers often assume the dimensions of change scores are equivalent to the dimensions of cross-sectional scores. However, this is, in fact, a faulty assumption, as shown by sense of identity in the applied example.

In addition to identifying different change processes, the change score factors could relate to and predict other outcomes differently. This has great implications for educational researchers, particularly those interested in growth processes related to success in higher education. Researchers are able to answer complicated research questions about the change processes of these constructs using this method. Additionally, factor analysis of change scores is easier to conduct compared to other sophisticated longitudinal models such as growth curve modeling. Even for those familiar with longitudinal data analysis, this technique provides another way of exploring change in a construct that could be added to their statistical toolbox. If the results obtained from factor analyzing change scores align with other longitudinal models that are theorized to answer similar research questions, the validity of the results would be further supported. Regardless, whether you are familiar with longitudinal data analysis or not, factor analysis of change scores is beneficial when answering research questions about the dimensionality of growth processes and validating results from other longitudinal models.

From a practical standpoint, change scores are not only easy to interpret, but require data from only two time points. In many longitudinal models, data from three or more time points are needed to conduct the analyses in order to explore change. Factor analysis of change scores requires only two time points. This is practical for most educational researchers working at institutions and colleges, as the odds of collecting pre- and post-test data are much higher than the odds of collecting data at three or more separate time points. Further, collecting data at three or more separate time points may require an abundant amount of resources from the researcher and their corresponding institution. Educational researchers interested in exploring change, but lacking the resources to collect data necessary for other sophisticated longitudinal models, may turn to factor analysis of change scores as an alternative. Factor analyzing change scores provides answers to similar research questions while requiring less intensive data. This, in and of itself, is a major advantage.

**Identifying Distinct Patterns of Growth**

The application of mixture modeling to change score factors adds a new layer of information about change in a construct. Factor analysis of change scores assumes the factor model fits all individuals in the population. In terms of change, the assumption is that change processes are the same for all individuals. However, change is often complicated; this is certainly the case for many constructs studied in higher education. For example, there may be unknown groups in the population that do not exhibit similar change patterns on the construct. Change score factor mixture modeling relaxes this assumption, allowing researchers to explore heterogeneity in patterns of growth or development. As demonstrated with the applied example, two latent classes that differ in sense of identity change patterns were identified in the present samples. The findings from the applied example highlight the many advantages to change score factor mixture modeling.

First, information about latent classes underlying change score factors can be obtained; particularly, the number of latent classes. This information is valuable to those interested in understanding a population of interest and its development on a construct. For instance, if development of a construct is significantly related to an outcome, it would likely be helpful to research whether development of that construct is the same for everyone in the population of interest. If not, what are different ways in which people could develop on the construct? These types of questions could be answered by change score factor mixture modeling.

Second, if latent classes are identified on change score factors, it indicates development of the construct is not uniform across individuals. In other words, development of the construct varies for groups of individuals in the population. Individuals with similar change patterns would be classified into one latent class. The characteristics of these change patterns are examined by comparing factor means and factor variances across latent classes. Differences between factor means and factor variances suggest differences in change patterns across latent classes.

Third, it is also possible for latent classes to differ in their item intercepts, loadings, and error variances. For example, a set of change scores may highly load onto the first factor for one class. For another class, a different set of change scores may have the strongest loadings on that factor. This is one scenario in which latent classes may differ in factor loadings. Change score factor mixture modeling could be used to test levels of measurement invariance between latent classes in the population of interest. For example, are the dimensions of change of the construct consistent across latent classes (configural invariance)? If so, are factor loadings the same across latent classes (metric invariance)? Finally, are item intercepts also the same across latent classes (scalar invariance)?

**Conclusion**

Change scores, when used in conjunction with latent variable modeling methods, can provide an abundant amount of information about growth processes. In this thesis, I explained and demonstrated the application and utility of two methods, factor analysis and factor mixture modeling, to change scores in uncovering growth processes of a construct and finding hidden patterns of growth. Both methods, when applied to change scores, provide unique information about change in a construct that is unobtainable with cross-sectional scores. Thus, this work is valuable to longitudinal researchers and practitioners interested in change. I argue for the use of change scores in a variety of analyses, particularly in education and psychology research where change scores are most likely to be available. It is my hope to inspire others to apply these methods to change scores.

Appendix A

Tables

Table 1.

*Change Score Factors of Sense of Identity Scale Items (Lounsbury & Gibson, 2004)*

| Item |
|---|
| <u>Change in Sense of Self Factor</u> |
| 1. I have a definite sense of purpose in life. |
| 2. I have a firm sense of who I am. |
| 4. I know what I want out of life. |
| 6. I don’t know where I fit in this world (r) |
| 7. I have specific personal goals for the future. |
| 8. I have a clear sense of who I want to be when I am an adult. |
| <u>Change in Morals and Beliefs Factor</u> |
| 3. I have a set of basic beliefs and values that guide my actions and decisions. |
| 5. I have a clear set of personal values or moral standards. |

*Note*. r = reversed scored.

Table 2.

*Fit Indices for the Three Mixture Model Parameterizations on Exploratory Sample*

| Model | LL | Par. | BIC | SAA-BIC | AIC | LMR *p*-value | Entropy |
|---|---|---|---|---|---|---|---|
| 1-Class | -23553 | 16 | 47228 | 47178 | 47138 | NA | NA |
| 2-Class | | | | | | | |
| Model A | -22794 | 25 | 45780 | 45701 | 45638 | < .001 | 0.761 |
| Model B | -21996 | 53 | 44227 | 44059 | 43926 | 0.437 | 0.975 |
| Model C* | -- | -- | -- | -- | -- | -- | -- |
| 3-Class | | | | | | | |
| Model A | -22389 | 34 | 45040 | 44932 | 44847 | < .001 | 0.828 |
| Model B | -21724 | 62 | 43925 | 43728 | 43572 | 0.253 | 0.987 |
| Model C* | -- | -- | -- | -- | -- | -- | -- |
| 4-Class | | | | | | | |
| Model A | -22250 | 43 | 44831 | 44694 | 44586 | 0.506 | 0.837 |
| **Model B** | **-21446** | **71** | **43438** | **43212** | **43034** | **0.043** | **0.996** |
| Model C* | -- | -- | -- | -- | -- | -- | -- |
| 5-Class | | | | | | | |
| Model A | -22110 | 52 | 44619 | 44454 | 44323 | 0.011 | 0.839 |
| Model B | -19788 | 80 | 40192 | 39938 | 39737 | 0.703 | 0.997 |
| Model C* | -- | -- | -- | -- | -- | -- | -- |

*Note*. * = did not converge despite 4000 random starts.

Table 3.

*Fit Indices for the One- and Two-Factor Model from Ong and Erbacher (2016) on Exploratory Sample*

| Models | *df* | $\chi^2$ | BIC | SSA-BIC | AIC | CFI | RMSEA | RMSEA 95% CI | SRMR | $\Delta\chi^2$ | $\Delta df$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| One-factor | 20 | 739.064* | 44914 | 44838 | 44778 | 0.767 | 0.129 | .121 - .137 | 0.069 | -- | -- |
| Two-factor | 19 | 298.112* | 44481 | 44402 | 44339 | 0.91 | 0.082 | .074 - .091 | 0.044 | 440.95* | 1 |

*Note.* N = 2128

Table 4.

*Fit Indices for the Four-Factor Mixture Model Parameterizations on Exploratory Sample*

| Model | LL | Par. | BIC | SSA-BIC | AIC | LMR *p*-value | Entropy |
|---|---|---|---|---|---|---|---|
| 2-Class, 1-Factor | | | | | | | |
| Model A/B | -22310 | 26 | 44820 | 44737 | 44819 | < .001 | 0.942 |
| Model C/D | -22283 | 27 | 44773 | 44687 | 44619 | 0.037 | 0.195 |
| 3-Class, 1-Factor | | | | | | | |
| Model A/B | -22289 | 28 | 44793 | 44704 | 44634 | 0.166 | 0.867 |
| Model C/D | -22270 | 30 | 44770 | 44675 | 44600 | 0.162 | 0.563 |
| 4-Class, 1-Factor | | | | | | | |
| Model A/B | -22283 | 30 | 44796 | 44701 | 44626 | 0.485 | 0.849 |
| Model C/D* | -- | -- | -- | -- | -- | -- | -- |
| 2-Class, 2-Factor | | | | | | | |
| Model A | -22095 | 27 | 44397 | 44311 | 44244 | 0.006 | 0.960 |
| **Model B** | **-22083** | **28** | **44381** | **44292** | **44222** | **0.001** | **0.954** |
| Model C* | -- | -- | -- | -- | -- | -- | -- |
| Model D* | -- | -- | -- | -- | -- | -- | -- |
| 3-Class, 2-Factor | | | | | | | |
| Model A | -22056 | 30 | 44343 | 44248 | 44173 | 0.067 | 0.931 |
| Model B | -21896 | 31 | 44030 | 43931 | 43853 | 0.315 | 0.987 |
| Model C* | -- | -- | -- | -- | -- | -- | -- |
| Model D | -22040 | 37 | 44364 | 44246 | 44153 | - | 0.957 |
| 4-Class, 2-Factor | | | | | | | |
| Model A | -22018 | 33 | 44290 | 44185 | 44102 | 0.058 | 0.825 |
| Model B | -21360 | 34 | 42981 | 42873 | 42787 | 0.149 | 0.996 |
| Model C* | -- | -- | -- | -- | -- | -- | -- |
| Model D* | -- | -- | -- | -- | -- | -- | -- |

*Note*. * = did not converge despite 4000 random starts.

Table 5.

*Parameter Estimates for the Two-Class, Two-Factor Model B on Exploratory Sample*

| | Factor Loading | | Item Intercept | |
|---|---|---|---|---|
| Change Score | Class 1 (*n* = 50) | Class 2 (*n* = 2127) | Class 1 | Class 2 |
| Sense of Self and Purpose Change Factor | | | | |
| Item 1 | 1.00* | 1.00* | 0.00* | 0.00* |
| Item 2 | 0.90 | 0.90 | 0.50 | 0.50 |
| Item 4 | 1.09 | 1.09 | 0.04 | 0.04 |
| Item 6 | 0.88 | 0.88 | 0.09 | 0.09 |
| Item 7 | 0.94 | 0.94 | 0.09 | 0.09 |
| Item 8 | 1.17 | 1.17 | 0.16 | 0.16 |
| Morals/Beliefs Change Factor | | | | |
| Item 3 | 1.00* | 1.00* | 0.00* | 0.00* |
| Item 5 | 0.86 | 0.86 | 0.04 | 0.04 |
| Factor Means | | | | |
| Sense of Self. | -1.84 | -0.01 | | |
| Morals/Beliefs | -0.59 | 0.03 | | |
| Factor Variances | | | | |
| Sense of Self. | 0.21 | 0.21 | | |
| Morals/Beliefs | 0.37 | 0.37 | | |

*Note.* * = fixed parameter; all unstandardized parameter estimates were statistically significant.

Table 6.

*Class Means by Change Factors for the Two-Class, Two-Factor Model B on Exploratory Sample*

| Class Means based on Posterior Probabilities | | |
|---|---|---|
| Change Factor | Class 1<br>( n = 50) | Class 2<br>(n = 2127) |
| Sense of Self. | -11.46 | 0.34 |
| Morals/Beliefs | -1.33 | 0.10 |

Table 7.

*Fit Indices of Best Fitting Mixture, Factor, and Factor Mixture Model on Validation Sample*

| | Par. | *LL* | BIC | SSA-BIC | AIC | LMR *p*-value | Entropy |
|---|---|---|---|---|---|---|---|
| Mixture Analysis | | | | | | | |
| 4-Class Model B | 71 | -7274 | 15013 | 14788 | 14690 | 0.4312 | 0.969 |
| Factor Analysis | | | | | | | |
| 2-Factor | 25 | -7413 | 14991 | 14911 | 14877 | -- | -- |
| FMM | | | | | | | |
| 2-Class, 2-Factor Model B | 28 | -7394 | 14972 | 14883 | 14845 | 0.1702 | 0.855 |

Table 8.

*Parameter Estimates for the Two-Class, Two-Factor Model B on Validation Sample*

| | Factor Loading | | Item Intercept | |
|---|---|---|---|---|
| Change Scores | Class 1 (*n* = 664) | Class 2 (*n* = 39) | Class 1 | Class 2 |
| Sense of Self and Purpose Factor | | | | |
| Item 1 | 1.00* | 1.00* | 0.00* | 0.00* |
| Item 2 | 0.92 | 0.92 | 0.07 | 0.07 |
| Item 4 | 0.98 | 0.98 | 0.05 | 0.05 |
| Item 6 | 0.76 | 0.76 | 0.17 | 0.17 |
| Item 7 | 0.85 | 0.85 | 0.11 | 0.11 |
| Item 8 | 1.21 | 1.21 | 0.16 | 0.16 |
| Morals/Beliefs Factor | | | | |
| Item 3 | 1.00* | 1.00* | 0.00* | 0.00* |
| Item 5 | 1.42 | 1.42 | 0.06 | 0.06 |
| Factor Mean | | | | |
| Sense of Self. | -0.05 | -0.08 | | |
| Morals/Beliefs | -0.7 | 1.07 | | |
| Factor Variances | | | | |
| Sense of Self. | 0.34 | 0.34 | | |
| Morals/Beliefs | 0.16 | 0.16 | | |

*Note*. * = fixed parameter; all unstandardized parameter estimates were statistically significant.

Table 9.

*Class Means by Change Factors for the Two-Class, Two-Factor Model B on Validation Sample*

| Class Means based on Posterior Probabilities | | |
|---|---|---|
| Change Factor | Class 1 (*n* = 664) | Class 2 (*n* = 39) |
| Sense of Self. | 0.26 | 0.51 |
| Morals/Beliefs | -0.08 | 3.10 |

Table 10.

*Validity Analyses for the 2-class, 2-factor Model B Solution*

| Auxiliary Variable | Statistic | Value |
|---|---|---|
| <u>Exploratory Sample</u> | | |
| GPA | Class 1 *M* (*SE*) | 2.90 (0.06) |
| | Class 2 *M* (*SE*) | 3.10 (0.01) |
| | $\chi^2$ | 12.35 |
| | *p*-value | < 0.001 |
| <u>Validation Sample</u> | | |
| GPA | Class 1 *M* (*SE*) | 3.160 (0.02) |
| | Class 2 *M* (*SE*) | 3.08 (0.06) |
| | $\chi^2$ | 1.59 |
| | *p*-value | 0.207 |

Appendix B

Figures

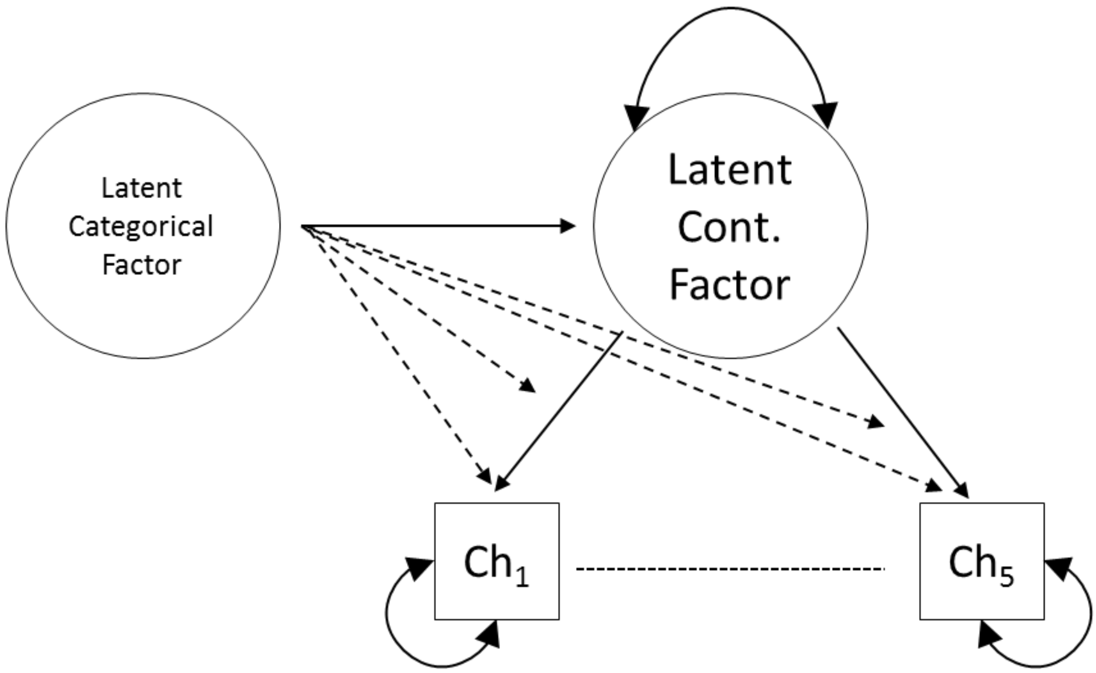


Figure 1. Path diagram for the general factor mixture model.

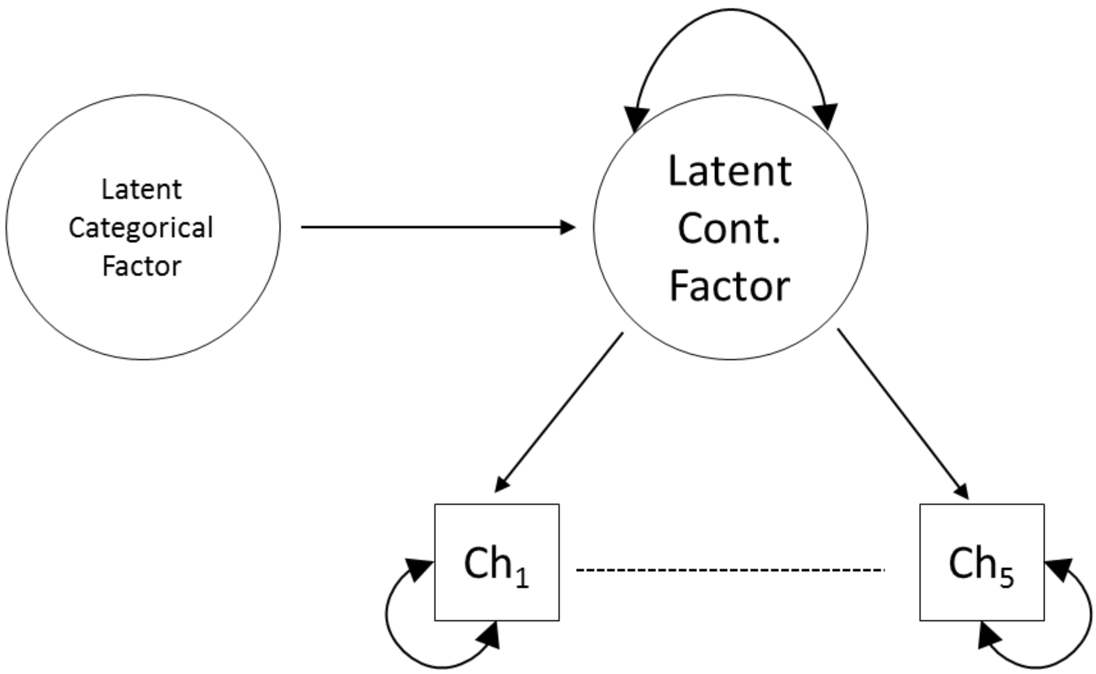


Figure 2. Path diagram for the general mixture factor model.

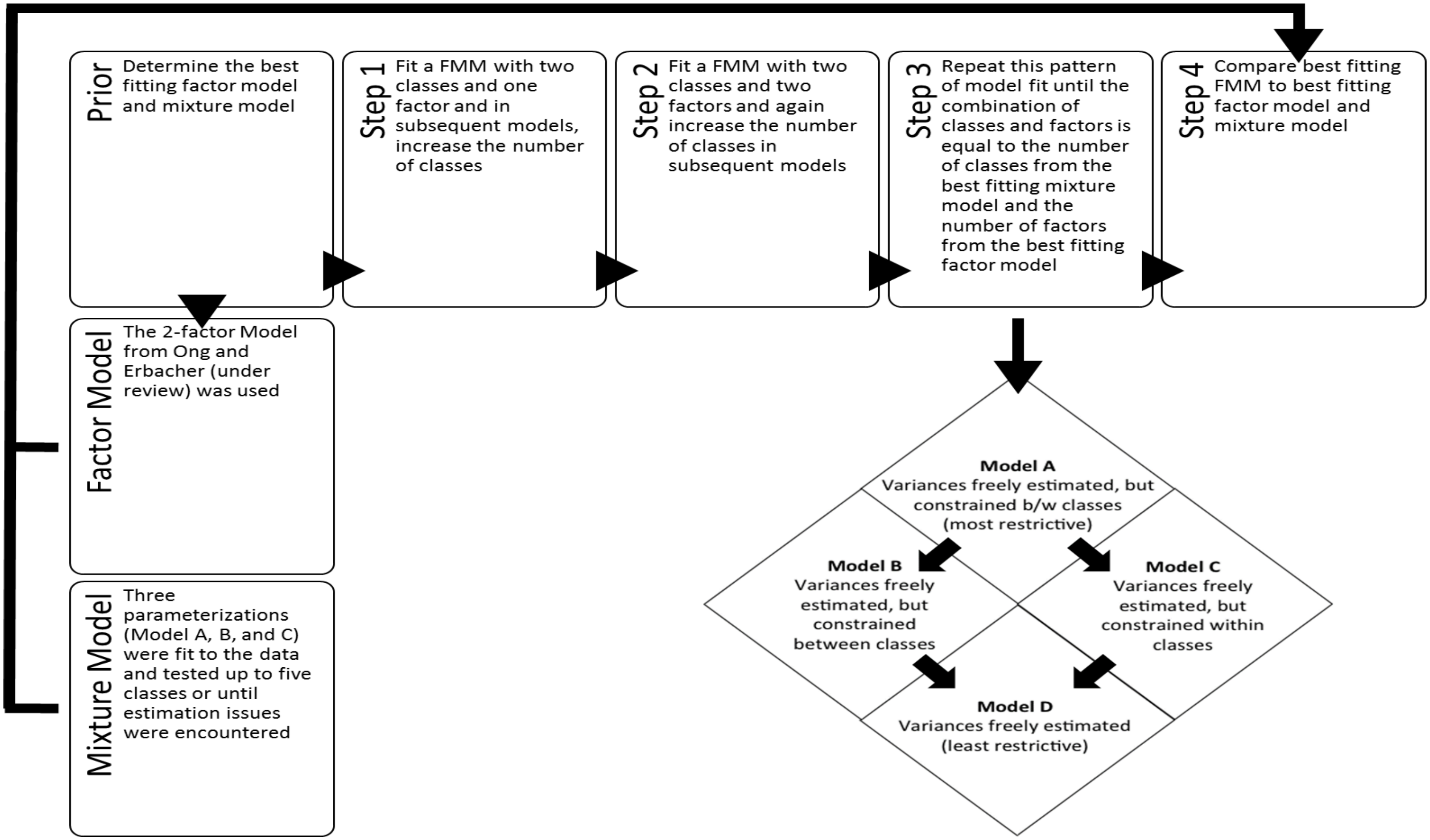


Figure 3. Outline for Constructing a FMM

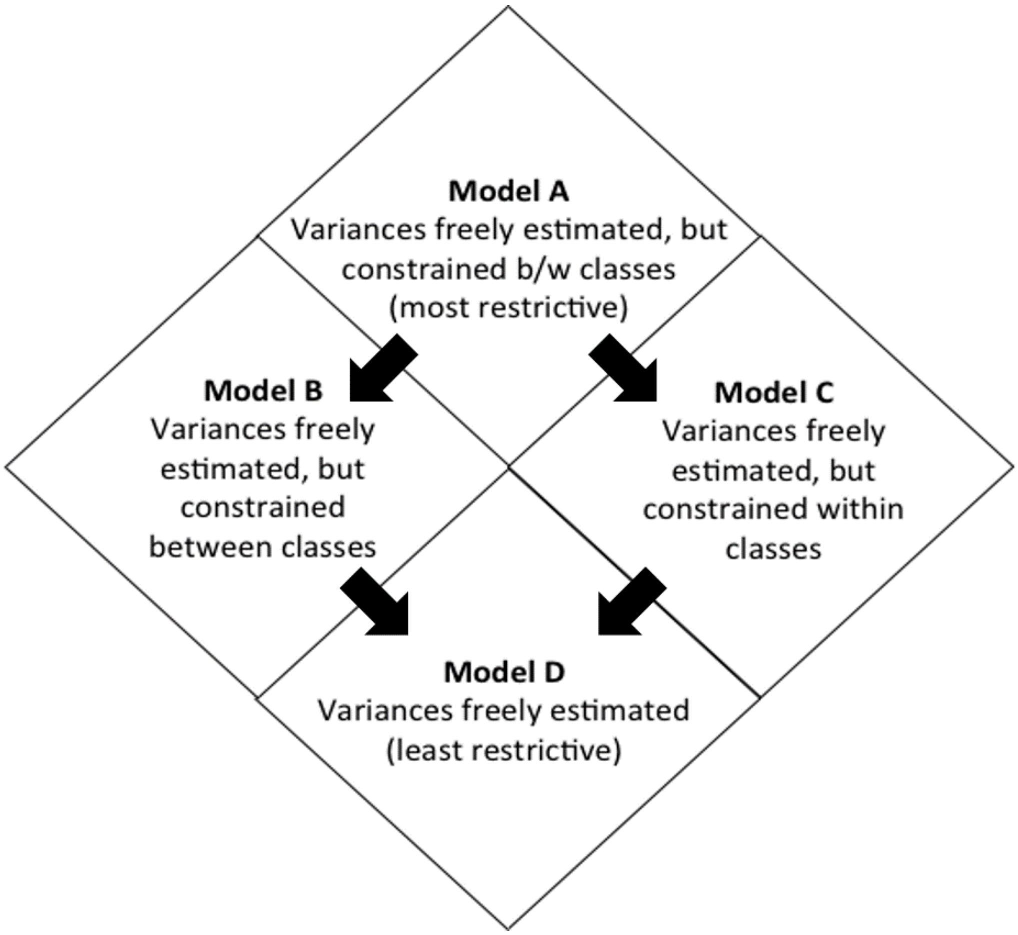


Figure 4. Summary of the Four SP-FM Parameterizations.

References

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

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, *20,* 93-114.

Asparouhov, T., & Muthén, B. (2013). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. *Mplus web notes*, *15*, 1-24.

Aston, K. L., Baran, R. D., Brownfield, K. D., & Smith, S. E. (2013). Who are you as a student? How does that effect your academic achievement and career aspirations? *Epistimi: Capital University's Undergraduate Research Journal,* 1-5.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G.R. Hancock & R.O. Mueller (Eds), *The reviewer's guide to quantitative methods in the social sciences* (93-114). New York, NY: Routledge.

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.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*(1), 3-29.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107,* 238-246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, *88*(3), 588-606.

Benson, J., & Nasser, F. (1998). On the use of factor analysis as a research tool. *Journal* *of Vocational Education Research*, *23*(1), 13-33.

Bergman, L. R., and Magnusson, D. (1997). A Person-Oriented Approach in Research on Developmental Psychopathology. *Development and Psychopathology*, *9*, 291–319.

Biesanz, J. C., West, S. G., & Kwok, O. (2003). Personality over time: Methodological approaches to the study of short-term and long-term development and change. *Journal of Personality*, *71*(6), 905-941.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Bohrnstedt, G. W. (1969). Observations on the measurement of change. In E. Borgatta (Ed.), *Sociological Methodology*. San Francisco; Jossey-Bass.

Bolck, A., Croon, M., Hagenaars, J. A. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*, 3-27.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods and Research*, *17*(3), 303-316.

Bozdogan, H. (1987). Model selection and Akaiki's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345-370.

Burr, J., & Nesselroade, J. R. (1990). Change measurement. In A. von Eye (Ed.), *Statistical methods in longitudinal research. Vol. 1: Principles and structuring change* (pp. 3-34). New York: Academic Press.

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading andunderstanding multivariate statistics* (pp. 99-136). Washington, DC: American Psychological Association.

Carraher, S. M., & Buckley, M. R. (1991). The effect of retention rule on the number of components retained: The case of the Pay Satisfaction Questionnaire. *Proceedings of the Southern Management Association*, 247-249.

Cattell, R. B. (1963). The structuring of change by P- and incremental-R technique. In C. W. Harris (Ed.) *Problems in Measuring Change* (pp. 167–198). Madison, WI: University of Wisconsin Press.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Cattell, R. B., & Jaspers, J. (1967). A general plasmode for factor analytic exercises and research. *Multivariate Behavioral Research Monographs*, *3*, 1-212.

Clark, S. L. (2010). M*ixture modeling with behavioral data* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3405665)

Clark S. L., Muthén, B., Kaprio, J., D'Onofio B. M., Viken R., & Rose R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling*, *20*(4), 681-703.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-359). New York: Plenum.

Cliff, N., & Hamburger, C. D. (1967). The study of sampling errors in factor analysis by means of artificial experiments. *Psychological Bulletin*, *68*, 430-445.

Comrey, L. A. & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates.

Crandal, R. (1973). The measurement of self-esteem and related constructs. In J.P. Robinson & P. R. Shaver (Eds), *Measures of Social Psychological Attitudes* (pp. 80-82). Ann Arbor, MI; ISR.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cronbach, L. J., & Furby, L. (1970). How we should measure change - or should we? *Applied Psychological Bulletin*, *74*, 68-80.

Culpepper, S. A. (2014). The reliability of linear gain scores as measures of student growth at the classroom level in the presence of measurement bias and student tracking. *Applied Psychological Measurement*, *38*(7), 503-517.

DeVellis, R. V. (1991). *Scale development: Theory and applications.* Thousand Oaks, CA: Sage.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*, 149–161.

Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, *17*, 193-219.

Hayton J. C., Allen D. G., & Scarpello V. (2004). Factor retention decisions in Exploratory Factor Analysis: A tutorial on Parallel Analysis. *Organizational Research Methods*, *7*(2), 191-205.

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

Holland, P. (2005). Lord's Paradox. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 2, pp. 1106–1108). Chichester, England: England John Wiley & Sons.

Horn, J. L. (1965). A rationale and a test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

Hu, L., & Bentler, P. M. (1995). *Evaluating model fit.* In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications.* Thousand Oaks, CA: Sage Publications.

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Jackson, D. L., Gillaspy, J. R., & Purc-Stehenson, R. P. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6-23.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-427.

Jöreskog, K. G., & Sörbom, D. (1986). LISREL VI: *Analysis of linear structural relationships by maximum likelihood and least squares methods*. Mooresville, IN: Scientific Software.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, *10*, 333-3511.

Kline, R. B. (1998). *Principle and practice of structural equation modeling.* New York, NY: The Guilford Press.

Kline, R. B. (2013) Exploratory and confirmatory factor analysis. In Y. Petshcer, C. Schatschneider, & D. L. Compton (Eds), *Applied Quantitative Analysis in Education and the Social Sciences* (pp. 170-201). New York, NY: Routledge.

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, *20*, 1-26.

Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, *52*(3), 377-389.

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

London, K. & Wright, D. B. (2012). Analyzing change between two or more groups: Analysis of variance versus analysis of covariance. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of Developmental Research Methods* (p. 279-290). New York, NY: Guilford Press.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *72*, 304–305.

Lorenzo-Hernandez, J., & Oullette, S. C. (1998). Ethnic identity, self-esteem, and values in Dominicans, Puerto Ricans, and African Americans. *Journal of Applied Social Psychology*, *28*, 2007-2024.

Lounsbury, J. W & Gibson, L. W. (2004). *Technical manual for the Resource Associates Personal Style Inventory and Adolescent Personal Style Inventory*. Knoxville, TN: Resource Associates.

Lounsbury, J. W., Huffstetler, B. C. Leong , F. T. Gibson , L. W. (2005). Sense of identity and collegiate academic achievement. *Journal of College Student Development*, *46*, 501-514.

Lounsbury, J. W., Levy, J.J., Leong, F.T.L., & Gibson, L.W. (2007). Identity and personality: The Big Five and narrow traits in relation to sense of identity. *Identity: An International Journal of Theory and Research*, *7*, 51-70.

Lounsbury, J. W., Richardson, J. D., Saudargas, R. A., & Levy, J. J. (2008). An investigation of extracurricular activities in relation to sense of identity of college freshmen. *Journal of College Orientation and Transition, 15*(2), 47-55.

Lubke, G. (2010). Latent variable mixture models. In G.R. Hancock & R.O. Mueller (Eds.). *The reviewer's guide to quantitative methods in the social sciences* (pp. 209-219). New York, NY: Routledge.

Lubke, G., Muthén, B. O, Moilanen, I. K, McGough, J. J., Loo, S. K., Swanson, J. M., Yang, M. H., Taanila, A., Hurtig, T., Jarvelin, M. R., & Smalley, S. L. (2007). Subtypes vs. severity differences in attention deficit hyperactivity disorder in the northern Finnish birth cohort (NFBC). *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*, 1584–1593.

Luyckx, K., Klimstra, T. A., Schwartz, S. J., and Duriez, B. (2013). Personality identity in college and the work context: *Development trajectories and psychosocial functioning. European Journal of Personality*, *27*, 222-237.

Magidson, J. & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, *31*(1), 223-264.

Magnusson, D. (2003). The person approach: Concepts, measurement models, and research strategy. In S. C. Peck & R. W. Roeser (Eds.), *New directions for Child and Adolescent development. Person-centered approaches to studying development in context* (pp. 3-23). San Francisco, CA: Jossey-Bass.

McDonald, R. P. (1967). *Nonlinear factor analysis*. Richmond, VA: Psychometric Corporation.

Millsap, R. E. (2011). Statistical Approaches to Measurement Invariance. New York, NY: Routledge, Taylor and Francis Group.

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. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nesselroade, J. R., & Cable, D. G. (1974). "Sometimes it's okay to factor difference scores"–the separation of trait and state anxiety. *Multivariate Behavioral Research*, *9*, 273-282.

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: A Multidisciplinary Journal*, *14*, 535–569.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw Hill.

Ong T. Q. & Erbacher M. K. (2016). *Items that hang together may not change together: Exploring the dimensionality of change in sense of identity*. Manuscript submitted for publication.

Pastor, D. A., & Gagné, P. (2013). Mean and covariance structure mixture models. In G.R. Hancock & R.O. Mueller (Eds.). *Structural equation modeling: A second course* (2nd ed, pp. 342-393). Charlotte, NC: Information Age Publishing Inc.

Peel D. & McLachlan G. J. (2000). Robust mixture modeling using t distribution. *Statistics and Computing, 10,* 339-348.

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 69–100). New York, NY: Springer

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

Schafer, W. D. (1992). Analysis of pretest-posttest designs. *Measurement & Evaluation in Counseling and Development*, *25*, 2-4.

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

Schreiber J. B., Nora A., Stage F. K., Barlow E. A., & King J. (2006). Reporting structural Equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338.

Sharma, K. K., & Gupta, J. K. (1986). Optimum reliability of gain scores. *Journal of Experimental Education*, *54*, 105-108.

Steiger, J. H. (1990). Structural model evaluation and modification: An internal estimation approach. *Multivariate Behavioral Research, 25*(2), 173-180.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K.A. Bollen, & J.S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.

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

Thurstone, L. L. (1935). *The vectors of the mind.* Chicago, IL: University of Chicago Press.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1-10.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41-47). Norwell, MA: Kluwer Academic.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*, 450-469.

Wainer, H. (1991). Adjusting for differential base rates: Lord's Paradox again. *Psychological Bulletin*, *109*, 147 – 151.

Waterman, A. (1982). Identity development from adolescence to adulthood: An extension of theory and a review of research. *Developmental Psychology*, *18*, 341-358.

Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, *20*, 59-69.

Williams, R. H., & Zimmerman, D. W. (1977). The reliability of difference scores when errors are correlated. *Educational and Psychological Measurement*, *77*, 679-689.

Worthington, R.L., & Whittaker, T.A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*, 806-838.

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

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94.

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research*, *17*, 253-269.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442.

Zimmerman, D. W., & Williams, R. H. (1982a). Gain scores in research can be highly reliable.

*Journal of Educational Measurement*, *19*, 149-154.

Zimmerman, D. W., & Williams, R. H. (1982b). On the high predictive potential of change and growth measures. *Educational and Psychological Measurement*, *42*, 961-968.