12-1-2015

Modeling Heterogeneity in Growth Mixture Models: A Case Study of Model Selection using Direct Behavior Rating

Janice Kooken
University of Connecticut - Storrs, janice.kooken@uconn.edu

Modeling Heterogeneity in Growth Mixture Models:

A Case Study of Model Selection using Direct Behavior Rating

Janice Kooken, Ph.D.

University of Connecticut, 2015

Abstract

This study investigates student classroom behavior changes over one year using multilevel growth mixture modeling to demonstrate how modifying assumptions of invariance affects parameter estimates, number of classes, and proportion of students assigned to each class. Current best practices for growth mixture modeling emphasize the importance of the proper specification, but the impact of these assumptions on the parameters and latent class composition has not been thoroughly addressed in applied research in multilevel growth mixture models. Using the Direct Behavior Rating Single Item Scale measures from 1975 students in lower elementary, upper elementary and middle school, a series of models were compared from full invariance to partial noninvariance. This research provides a description of steps, decisions, and results from testing for noninvariance, and how these affect the resulting subgroups and model parameters. Results indicated a dramatic shift in the students from higher classes to lower classes as the model was relaxed to allow for class level difference in variance parameters. The best fitting models for each grade group contained three latent classes characterized by students with consistently good classroom behavior, students with less consistent moderate behavior, and students with highly variable behavior. Criterion measures provided validation of these results. Research on classroom behavior heterogeneity using GM modeling represents an important

addition to the knowledge base by providing a descriptive typology of student behavior, lending information necessary to fine tune behavior interventions. This research suggests that using variability as a criteria for the typology results is a more sensitive screening instrument identifying more students for follow up. This research also suggests that because student classroom behavior is highly variable for the students most at risk due to behavior problems, variability should be measured and tracked in single case studies to identify interventions to reduce this variability.

Modeling Heterogeneity in Growth Mixture Models:

A Case Study of Model Selection using Direct Behavior Rating

Janice Kooken

B.A., Queens College of the City University of New York, 1982

M.S., University of Illinois at Chicago, 1998

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at

The University of Connecticut

2015

APPROVAL PAGE

Doctor of Philosophy Dissertation

Modeling Heterogeneity in Growth Mixture Models:

A Case Study of Model Selection using Direct Behavior Rating

Presented by

Janice Kooken, B.A., M.S.

Major Advisor

D. Betsy McCoach

Associate Advisor

Noel Card

Associate Advisor

Sandra Chafouleas

Associate Advisor

Faith G. Miller

Associate Advisor

H. Jane Rogers

University of Connecticut
2015

Acknowledgements

The journey to this point has been one characterized by many unexpected detours and delightful sojourns. I never lacked aspiration and desire to pursue a PhD, but I never had any expectation that I would be able to reach this goal. I was raised on a farm in the Midwest, the granddaughter of Polish immigrants. My own parents spoke Polish, but they would not teach it to me. Polish people were considered ignorant, and my parents did not want me to be burdened with evidence of that heritage. My parents taught me the value of hard work and an education. Without their love, encouragement, and support, I would never have even finished my Bachelor's. So, first of all, I would like to acknowledge my Mom and Dad. Nature or nurture, I am what I am in no small part because of them.

I grew up before Title IX, so girls' sports were not accessible to me. I was more studious than most of my peers, and I loved mathematics. Katherine Pedersen was my first advisor in college mathematics and has been a support and encouragement to me for my entire adult life. I am very grateful for her continued support.

After graduating with a degree in mathematics for secondary education, my life took some unexpected twists and I began a career in insurance. During this phase of my life, I met my husband Mike. There is no one who has had a greater or more positive influence on my life. He believed in me, challenged me and supported me through so many different journeys. He also taught me to have fun. Without him I would never have even considered pursuing my PhD.

At several points along my path, the sojourns occurred. I have three strong, intelligent, and delightful sons, Kyle Williams, Luke Kooken and Calvin Kooken. They have had to put up with me as a mother, and yet through it all, we have stayed very close. I am thankful that they let me talk about mathematics, statistics, and Calculus since they were toddlers at the same time we

were giggling, reading, baking cookies and finger painting. All three of them are brilliant and the apple of my eye. I would like to acknowledge them for their support of me through this program.

Time passed, and one day I walked into UConn and started a PhD program. When someone asks me why, I tell them there are hundreds of reasons but really only one: I finally believed that I could.

Betsy McCoach has been without a doubt, the single most influential person at UConn. Betsy provided me with the light and compass I needed to direct my studies. Betsy listened, challenged, and validated my interests, my questions, and my frustrations. Betsy answered my questions and provided the perfect level of direction at each point along the way. I have learned so much from her and will always aspire to be a scholar and an advisor like her. I am deeply grateful to Betsy.

I am also deeply grateful to Sandy Chafouleas for her mentoring and for providing me the opportunity to work with the DBR data. The data and substantive issues provided just the perfect mix of measurement, mathematics, and a focus on helping children succeed in school to stimulate my interests.

I am also deeply grateful to my other advisors: Faith Miller, Noel Card, and Jane Rogers. Their invaluable suggestions and technical advice greatly improved my work. I also want to acknowledge the other professors in the UConn measurement program. I have learned so much through the classes and conversations we have shared.

I would also like to acknowledge my friends with whom I have laughed and cried, drank and prayed, but with whom I have not had enough time to enjoy their company because I have been too busy!

Note. Preparation of this article was supported by funding provided by the Institute for Education Sciences, U.S. Department of Education (R324A110017). Opinions expressed herein do not necessarily reflect the position of the U.S. Department of Education, and such endorsements should not be inferred. Correspondence regarding this article should be addressed to the Project Director, Sandra M. Chafouleas, at the University of Connecticut, Department of Educational Psychology (U-3064), Storrs, CT 06269. Email: sandra.chafouleas@uconn.edu

Table of Contents

Chapter 1: Introduction

Students in our nation's schools represent an increasingly diverse population in terms of racial and ethnic mix, academic abilities, language, behavior, motivation and attitudes. A recurring theme of current educational research is to promote research and practice that supports and encourages students of all backgrounds, with an emphasis on justice to right inequalities (e.g., American Educational Research Association Annual Meeting, 2015). The statistical procedures used in much of educational research represent a framework whereby students are compared based upon their membership in defined groups: males versus females, regular education versus special education, or one race versus another. This type of research has supported tremendous growth in our knowledge of educational effectiveness, policy, and practice. Yet in doing so, the use of defined groupings does not suggest that being in the group causes or fully defines that particular outcome, leaving questions unanswered that may be very helpful in providing targeted assistance to these students. More advanced analytics provide opportunities to increase our ability to thoroughly explore how students vary across a variety of educational outcomes while demographic groupings do not provide mutually exclusive and exhaustive classifications in terms of these outcomes.

The term "big data" (Gandomi & Haider, 2015) is used in business when referring to the increased availability of large, complex data which when combined with advanced analytic techniques can uncover new insights and innovative solutions. Large complex databases are increasingly more available in educational and psychological research as well, including large national databases and data from smaller research projects. The availability of data and advanced computing capabilities results in a widening range of opportunities to ask richer more focused

research questions and apply advanced analytics to learn about how students learn (Podesta, Pritzker, Moniz, Holdren, & Zients, May, 2014).

An emerging area of research in education and psychology involves the use of a family of analytic techniques referred to as latent class analysis. Rather than using defined groups such as race, gender, or disability status to study differences in student outcomes, these techniques rely upon patterns in the student outcomes themselves to define distinct classes based upon these outcomes. These classes are called latent classes because the groupings are not defined based upon observable variables.

Similar to latent class analysis, growth mixture (GM) modeling and latent class growth (LCG) modeling represent person-centered modeling techniques useful in studying variations in change over time using longitudinal data (Bauer & Curran, 2003; Muthén, 2001; Muthén & Shedden, 1999; Nagin, 1999). GM and LCG modeling are similar in that they use response patterns to identify subgroups within the population that are qualitatively and quantitatively distinct in their composition and change patterns but not completely determined by known characteristics such as gender, race, or value of a related covariate. Unlike latent growth curve modeling (Bollen & Curran, 2006) where within and between subject variability is modeled using random effects and explanatory covariates, LCG and GM models represent heterogeneity by identifying patterns that characterize unobserved subpopulations and sort individuals into these groups.

In educational research, model specifications need to be flexible enough to capture characteristics of student diversity, which in statistical terms is referred to as heterogeneity, but rigorous enough to maintain strong psychometric properties. When using GM models, the researcher can impose assumptions of invariance which may compartmentalize the location of

the heterogeneity. In this context, invariance refers to the assumption that parameters are the same, e.g. are invariant, across subgroups within the population. This is a different context than measurement invariance. A measure is invariant whenever its use in differing situations, such as with people from different groups, at different times, or under different conditions, results in the same score for individuals with the same trait level (Brown, 2006; Kline, 2011; Vandenberg & Lance, 2000). Statistical models should capture the heterogeneity in our population. Current research has provided limited examples and guidance on how varying parameterization choices can affect results or what techniques may be particularly effective in identifying optimal choices.

Statement of the Problem

This research addresses both a methodological question and a substantive question. First, applied researchers desiring to model heterogeneity do not currently have access to research that provides a thorough discussion of how using different model specification can have a dramatic impact on class composition and trajectory shape. In methodological literature (e.g. Muthén, 2004; Petras & Masyn, 2010), proper investigation and decision making regarding testing the mean and covariance structure is emphasized, focusing on proper parameterization including the level of invariance. Enders and Tofighi (2008) found that misspecification of residual variance parameterization negatively affects the accuracy of the parameter estimates, number of classes, and class proportions. In particular, the level of invariance in an applied example of multilevel GM modeling has not been explored and reported in the literature. It is unknown how the systematic relaxation of invariance assumptions affects class size, composition, initial values and changes over time. These variations may have a sizable impact on substantive conclusions, resulting in a great deal of uncertainty on how to interpret results.

Second, the study and classification of people and organizations in education and psychology often involves seeking answers to questions involving the level of an attribute, to what extent that attribute changes over time, and what characteristics affect its level and change. Prior research on identifying and assessing student classroom behavior has provided a strong theoretical framework for a group of behavior targets theoretically linked to classroom success and measured using the Direct Behavior Rating Single Item Scale (DBR-SIS) (Chafouleas, 2011; Chafouleas, Volpe, Gresham, & Cook, 2010;). Students need to consistently listen, pay attention, follow directions, stay on task, control their emotions, control externalized behaviors, and be respectful of others. The evaluation, assessment and classification of students based upon behavior in the classroom begins before the first day of school, as the teacher collects and considers information from records and prior teacher evaluations, and continues as the teacher monitors and manages student behavior throughout the school year. Except for an individual teacher's idiographic understanding of her own classroom, characteristics of a normative classroom distribution of students across the spectrum based upon level and variability on these target behaviors are still largely unknown. If unobserved subgroups are identified and validated, the characteristics of these subgroups could provide insights to improve interventions for students whose behavior interferes with their ability to succeed in the classroom. No typologies currently exist of student classroom behavior across the lower elementary, upper elementary and middle school years.

Research Questions

The purpose of this study was to gain an understanding of how invariance assumptions in multilevel GM modeling can result in differences in latent class composition and growth curve characteristics. Using the context of student classroom behavior, the study disentangles how

mean levels and variability in the intercepts, slopes, random effects, and residual variances within and between students and teachers affect latent class profiles using data collected over one school year. Using GM modeling to study student classroom behavior as measured longitudinally using the DBR-SIS (Chafouleas, 2011) provides baseline descriptive growth models and helps in defining a typology of student classroom behavior over time.

This study addresses three questions. The first question addresses the methodological concerns related to how residual noninvariance may affect decision making when using GM modeling. The second question analyzes the selected models for evidence of the validity of the classes. The last question seeks to define the typology of student classroom behavior identified in the modeling steps using the distribution of student demographics, the shape and scale of the trajectories, and the level of variability.

1. When selecting the multilevel GM model that best represents the characteristics of a real data sample of student classroom behavior using the DBR-SIS, in what ways and to what extent does noninvariance affect decision making, including application of enumeration indices, and resulting latent class characteristics?
2. To what extent does the trajectory class membership selected from research question 1 predict distal outcomes of student behavior measured using alternatives measures such the Behavior Assessment System for Children-2 Behavioral and Emotional Screen System (BESS) (Kamphaus & Reynolds, 2007) risk classification, office disciplinary referrals (ODRs), and suspensions, providing validity evidence for the latent trajectory classes?

3. What are the characteristics of the students comprising latent classes of classroom behavior trajectories that emerge from the use of GM models, and in what ways do the latent classes vary in intercepts, slopes, and variances?

Background

GM modeling. Although GM modeling and LCG modeling are similar, they differ in the level of heterogeneity reflected in the model. For LCG models, growth trajectories are assumed to be the same for all subjects within a latent class. That is, although intercepts and slopes vary by class, all subjects within that class are assumed to have the same intercept and slope as a result of constraining the intercept variance and slope variance to zero. From a measurement perspective, this means the latent classes vary by mean levels but are otherwise invariant. From a substantive perspective, this means that the latent classes of growth trajectories are adequate to describe all the heterogeneity in the population. The assumption of invariance is relaxed in GM modeling where heterogeneity can be represented in the model through estimating a variety of additional parameters including differences in intercepts, slopes, the variability of the intercepts and slopes and residual variances. These additional parameters can help capture the characteristics of the heterogeneity. In doing so, we must be careful that in our quest to model and understand population heterogeneity, we give proper methodological and substantive consideration to choices that explore psychometric properties in the measure and in the model and not confuse measurement noninvariance and heterogeneity. Measurement noninvariance is an undesirable measurement property. Heterogeneity refers to detailing the differences in a particular population. To explore this more, we need to consider some of the basic properties of invariance, discussed briefly here and more fully developed in Chapter 2.

Measurement invariance represents a multifaceted property with implications to nearly all aspects of the measurement of latent traits in psychometrics and structural equation modeling. When a measure is noninvariant, two people with the same trait level may receive different scores on the measure. Therefore, interpretations and inferences drawn from comparing scores of people in different groups lack validity. The application of invariance to GM models has a related but slightly different meaning. In GM modeling, heterogeneity in the population is modeled through identification of unobserved subpopulations. Contrary to the requirement for measurement invariance, the parameterization in GM modeling in some cases reflects noninvariance, e.g. that the measurement is not invariant, between the latent subpopulations. The key is to identify the source of this noninvariance. Differences may be due to level, amount of change over time, random effects, covariances between the latent variables, or residuals.

The level of complexity of the models increases as parameters are allowed to randomly vary. Morin et al (2011) provided results which described how varying choices to fix or estimate these parameters resulted in changes to the number of classes, trajectories, and class distributions. Models are created using a simple parameterization with only fixed effects, run with stepwise increases in the number of classes, and compared based upon selected fit indices. Parameterization is then changed by estimating random effects, allowing these effects to vary by class, and residuals to vary across time or by class. Again, each unique model parameterization needs to be run with stepwise increases in the number of classes and compared. This situation is exacerbated when data are clustered, such that heterogeneity may exist at many different levels of the model such as through random effects at the student level, within latent class within cluster, between latent class within cluster level, or within student residual. After completing these steps, a best model needs to be selected, but which model is best may not be clear.

One of the most difficult aspects of GM modeling is class enumeration. The research technique seeks to use the response patterns to define a classification system. This classification is not directly observable in the population. Often teachers are skilled at identification and documentation of differences in behavior on a student by student basis. They may even know which students are consistently well-behaved, which are less consistent, and which students have severe behavioral difficulties. Unfortunately, this knowledge may be largely subjective based upon the teacher's level of experience and comfort in classroom management. GM modeling seeks to use scientific methods based upon statistical analytics to define these groups.

An active body of research currently seeks to identify and confirm the accuracy of techniques to identify the number of unobserved subgroups (Peugh & Fan, 2015), referred to by Nylund, Asparouhov and Muthén (2007) as class enumeration. Model misspecification can have negative effects on parameter estimates, class enumeration, and class assignment (Enders & Tofighi, 2008). Assuming invariance when the some aspects of model specification are noninvariant has been shown to have a negative impact on the identification of the number of subgroups and the substantive interpretation of those subgroups in applied research (Enders & Tofighi, 2008; Morin et al., 2011). Both Enders and Tofighi and Morin et al. recommend that the modeling process include testing invariance of within class residual variance. This is accomplished by comparing the model fit of competing models where one uses an invariant residual structure and another allows residual variances to vary between latent subgroups.

Class enumeration in GM modeling may present additional challenges for studies within educational settings where students are nested within classrooms necessitating the use of multilevel modeling. The research literature lacks examples of applied longitudinal studies using GM modeling that provide insight into the challenges of making modeling choices in an

environment where model specifications are in flux. Of 571 GM modeling research studies identified from Psychinfo from 2000 to October, 2015, nine were multilevel defined as students nested within classrooms or schools (Brown, et al., 2008; Chen, Kwok, Luo, & Willson, 2010; Finch & French, 2014; Kellam et al., 2014; Kuntsche Otten, & Labhart, 2015; Morin, Maiono, Marsh, Nagengast, & Janosz; 2013; Owen et al., 2015; Palardy & Vermunt, 2010; Simmons et al., 2015). Of these nine studies, the studies by Kellam et al., Kuntsche et al., Morin et al., Owen et al., Palardy and Vermunt, and Simmons et al. represent applied research. These studies provide a variety of levels of success at reflecting clustering. For example, Morin et al. did not have sufficient level 2 units to use multilevel modeling. Owen et al. had convergence problems with the multilevel models. Kuntsche et al. modeled at the event level, not the individual level. Simmons et al. used separate steps with GM modeling at the student level, propensity scores and then multilevel regression. Palardy and Vermunt modeled heterogeneity between clusters, assigning all individuals in a cluster to the same latent class. None of these studies represent an applied study using GM modeling to model observations within individuals, assigning individuals to latent classes, and reflect random effects for individuals within clusters. No comparable research found in the literature addressed the interaction of modeling choices, parameters, and number of classes in multilevel GM modeling on either real or simulated data.

Student classroom behavior typologies and the DBR-SIS. The motivation for the current study stems from the desire to understand the level, change and variability in the classroom behavior of students in elementary and middle school over a period of one school year. Student academic success is known to be a function not just of efforts in the cognitive and attitudinal domains, but also is strongly affected by the behavior of students in the classroom (Hinshaw, 1992; Wentzel, 1993; Georges, Brooks-Gunn, & Malone, 2012). For teachers to

maximize their impact in understanding and managing student classroom behavior, they, and the researchers who support them, need to understand and utilize scientifically based classification systems, also known as typologies, for student classroom behavior. An active line of research has investigated the typology of student behavior groupings (DiStefano, Kamphaus, Horne, & Winsor, 2003; DiStefano, Kamphaus, & Mîndrila, 2010; Huberty, DiStefano, & Kamphaus, 1997; Kamphaus, Huberty, DiStefano, & Petoskey, 1997; Kim, Orpinas, Martin, Horne, & Sullivan, 2010; Nagin & Tremblay, 1999) using behavior rating scales such as the Behavior Assessment System for Children Teacher Rating Scales for Children (BASC TRS-C; Reynolds & Kamphaus, 2006) and the Child Behavior Checklist (CBCL; Achenbach, 1991).

The DBR-SIS (Chafouleas, 2011) represent a hybrid assessment technique that contains characteristics of both the systematic direct observation and behavior rating scales. The DBR-SIS have been used to record student behavior in a longitudinal study involving students in elementary and middle school (Johnson, Miller, Chafouleas, & Kooken, 2011). With the DBR-SIS, the rater, who is the teacher, rates a student on a single behavior, measuring the proportion of time a student exhibits the behavioral construct in question. For example, a teacher would use a scale of 0-10 to rate the percentage of time a student is exhibiting behavior that indicates academic engagement, with 0 meaning the student did not exhibit that behavior to 10 meaning the student exhibited that behavior 100% of the time. The resulting measures are not normally distributed.

To support the use of GM modeling, the following two conditions must hold: prior research needs to suggest that the population is comprised of heterogeneous sub-populations and the data must lack normality (Bauer & Curran, 2003). A longitudinal study of classroom behavior using the DBR-SIS provides a unique opportunity to use GM modeling to study

heterogeneity, including residual noninvariance, the impact of modeling choices on class identification and parameter estimation, and the impact of clustered data. The DBR-SIS have been used in research and applied settings for progress monitoring to track change, but little is known about the characteristics of normative profiles of classroom behavior trajectories. Several studies have reported strong evidence for the validity of the DBR-SIS as measures of behavior change in single case design as reported by Chafouleas, Sanetti, Kilgus, and Maggin (2012) and Christ, Nelson, Van Norman, Chafouleas, and Riley-Tillman (2014), yet for some students, there is a high degree of variability in day to day measures. Little is known regarding the characteristics of the variability, the characteristics of the students with higher variability, and if there are unobserved subgroups whose characteristics would provide insights to improve interventions for students with high variability.

The purpose of this study was twofold. First, the study investigated the level of noninvariance in GM modeling using real data to identify how differing invariance assumptions can affect class composition and characteristics. Through research questions 1 and 2, this research investigated how to best describe the heterogeneity in the underlying population and use methodological and substantive approaches to identify the source of these differences in the subpopulations. Given that this investigation uses classroom behavior measures, the class composition and characteristics can be explored and validated using predictive and concurrent measures related to classroom behavior. Second, through research question 3, the findings from these investigations contributed to developing typologies of student classroom behavior.

Methodology

The study utilizes data from a longitudinal validation study of the DBR-SIS in which 1975 students in grades 1,2,4,5,7 and 8 were assessed on their classroom behavior by teachers

twice daily for five days, during each of three data collection periods, fall, winter, and spring, totaling 30 data points. Grades 1-2 are considered lower elementary (LE), grades 4-5 upper elementary (UE), and grades 7-8 middle school (MS). During each rating period, ratings were structured such that five students were concurrently rated, using the DBR-SIS, administered twice daily for five consecutive days. In this study, teachers rated student behavior using two teacher-rated measures of student behavior: the DBR-SIS and the BESS.

To address research question 1, statistical modeling captured heterogeneity in the behavior trajectory of groups of students using multilevel GM modeling of patterns of observed variables at the student level, adjusting for correlated outcomes due to the teacher level. Following the procedures outlined in Jung and Wickrama (2008), Nylund, Asparouhov, and Muthén (2007), and using Mplus 7.3 (Muthén & Muthén, 1998-2014), I adopted an iterative process to test models with successively greater complexity from an unconditional multilevel model to a multilevel GM model. Starting with a fixed effect model, intercepts, slopes and residuals were freely estimated in a stepwise progressive fashion starting at the lowest level within students, between teachers, and then between latent classes. A combination of techniques including model quality based upon convergence, fit criteria, likelihood ratio tests, size of classes, and substantive interpretability informed the selection of final models. Data availability for three grade groups allowed the replication of the study using separate analyses for LE, UE, and MS.

Research question 2 provided a method to investigate the validity of the latent classes in predicting other behavioral outcome measures and in the composition and shape of their trajectories. Validation of the latent classes involved analyzing how well the classes predicted distal outcomes which included ODRs, incidence of suspension or expulsion, disability status,

and risk level as determined by the BESS. Theory supported the hypotheses that students in classes with high and consistent ratings have a lower incidence of ODR’s and other disciplinary problems, whereas students in latent classes with lower and inconsistent ratings have a higher incidence. Second, the final selected models reflect a predicted trajectory of the DBR-SIS for each of the latent classes. Parallel results between LE, UE, and MS provide evidence for the validity of the class groupings by representing replication of results. In addition, differences in gender, race, ethnicity, and disability status were compared across the latent classes. Given that other research has identified that boys, students of color, Hispanic students, and students with disabilities as having the greatest behavioral challenges, this study also examined these factors. If latent classes characterized by lower engagement, higher disruptiveness, and lower respectfulness have a lower proportion of students in these demographic groups, this would be of interest, but the opposite would not represent evidence against the validity of the latent class. If latent classes with more problematic behavior have a lower proportion of students identified with criterion measures as at risk at school (e.g., BESS and ODRs) this would be provide evidence against the validity of the latent class structure. A secondary goal of this study is to add to theory needed to identify students with problem behavior based upon their patterns of behavior rather than by demographic characteristics.

The results from the analyses conducted to answer research questions 1 and 2 is fully examined for research question 3. Descriptive statistics provide the composition of each latent class by gender, race, ethnicity, disability status, and level of behavioral support. In addition, information on parameter estimates and a plot of the average trajectory is provided and discussed with emphasis on the substantive interpretation. The analysis of variation in intercepts, slopes, and residual variances includes connections with substantive theory and practice. Specifically,

recommendations and insights aim to connect these results with research and practice, especially using the DBR-SIS.

Using the descriptive information and shape of the trajectory, the classes are described in terms of their typology of student behavior. These trajectory classes are compared to those found in the behavior typology literature. Latent trajectory classes are described by their level of risk for behavioral problems in school. These risk levels are compared to findings from other literature (Dowdy et al., 2014; Dever, Dowdy, Raines, & Carnazzo, 2015).

Scholarly Significance

The approach used in this study reflects what Marsh and Hau (2007) referred to as a "methodological-substantive synergy" (p. 552) in which methodological insights are gained through the application of modeling techniques using real data to gain answers to questions of substantive interest. Marsh and Hau argue that these synergies are of great importance in the field of educational psychology because of the substantial amount of measurement error in most indicators, making them especially suitable for models using latent variable techniques. Applied studies which use sophisticated latent variable techniques contribute greatly to our understanding when they also include comparisons to more traditional techniques and identification of potential pitfalls. The use of real data from classroom behavior studies provides an opportunity to study the effect of systematically relaxing invariance assumptions on class composition and growth parameter estimates, while using real contextual distal variables such as the number of ODRs to validate latent class profiles.

From a methodological perspective, this study adds to the current literature by providing technical information on how changes in parameterization, and in particular residual variance assumptions can impact class assignment and characteristics in GM modeling. It provides

techniques which were useful in completing a thorough analysis of the class composition, how class assignment changed as different parameterizations were chosen, and how trajectory characteristics changed. No prior research has provided this much detail on the effect of fine tuning invariance assumptions in GM modeling. Applied researchers can use both the techniques and the results to assist them in identifying the proper parameterization for their own GM models.

From a substantive perspective, this study adds to the current literature by representing the first time growth mixture modeling has been used to analyze student classroom behavior change. This study provides a comprehensive set of descriptors providing information on the number, initial status, and rate of change of the classes, information on whether they vary by grade group, and whether class assignment is predictive of distal outcomes such as ODRs and risk status on the BESS. The results of this analysis provide a large body of knowledge to assist both researchers and practitioners in understanding, diagnosing and intervening to improve student classroom behavior. For example, finding that differences in behavior relate to the within student variability on a day to day basis, teachers and practitioners may wish to target interventions that reduce variability rather than those that are targeted to improve a mean score. In addition, where class groupings are predictive of ODRs, suspensions, and expulsions, this analysis provide important validation to the procedure and the measure. Research on classroom behavior heterogeneity using GM modeling represents an important addition to the knowledge base by providing a descriptive typology of student behavior, lending information necessary to fine tuning behavior interventions.

Chapter 2: Review of Literature

Techniques for Statistical Modeling of Change across Time

The study of people and organizations in education and psychology often involves seeking answers to questions involving the level of an attribute, to what extent that attribute changes across time, and what characteristics affect variability in its level and change. When studying change, statistical models are available in a wide range of sophistication from multiple regression and multilevel modeling to structural equation modeling of latent variables. To study change, we need data to be collected longitudinally, meaning measures repeated over time for the same individual. With repeated measures, the observations are not independent within a given person, and techniques need to be used to adjust for the lack of independence. Researchers seek a defensible statistical model that answers the research question, accurately models the measurement attributes of the data, and reflects the heterogeneity in the population. Multilevel GM modeling techniques used in this study are related to multilevel models, latent growth curve models (LGCMs), and latent class growth (LCG) models. Characteristics of each of these models presented separately can help describe the nuances of GM modeling.

Multilevel Models. When conducting research in educational settings, it is often the case that questions of interest involve observations or students who are in some way related to each other and not independent (McCoach, 2010; Peugh, 2010; Raudenbush & Bryk, 2002). Cross sectional research may include multiple students from a single classroom, referred to as students nested or clustered within classrooms. Students in the same classroom are more likely to be correlated because they have been exposed to the same environmental factors such as the same teacher, peer group, physical environment, rules, and educational materials. In longitudinal studies, repeated measures over time are nested within individuals violating the assumption of

independence. Correlations must be reflected in the modeling technique to reduce bias in the standard errors and reduce the likelihood of Type I errors. When longitudinal studies involve students nested within teachers or schools, it is sometimes helpful to consider these as three level models with observations nested within students, and students nested within contexts (i.e., classrooms or schools).

Multilevel models, also referred to as hierarchical models, represent a technique for modeling clustered data. For longitudinal studies, data are represented in long format, where each row represents the value for the observation at a single point in time. These models are structured to capture the initial status, represented by the intercept, and change per unit in time, represented by the slope. Multilevel models provide a statistical technique to measure the amount of variability in the outcome measure that is due to characteristics of the observation, known as Level 1, or between individuals, referred to as Level 2 (McCoach, 2010; Raudenbush & Bryk, 2002). The equations for multilevel models and the definitions of the terms follow. In all cases, model equations are presented without covariates or predictors.

Level 1 Model: (1)

$Y_{ti} = \pi_{0i} + \pi_{1i} \text{Time}_{ti} + e_{ti}$

Level 2 Model:

$\pi_{0i} = \beta_{00} + r_{0i}$

$\pi_{1i} = \beta_{10} + r_{1i}$

i = 1, 2, …n students

t = variable representing observations in time.

Y_{ti} = outcome measure at time t for student i

$Time_{ti}$ = variable representing time for the observation t for the student i, which in this example and in this study represents linear change.

π_{0i} = predicted mean of the outcome measure of student i.

π_{1i} = predicted change for each unit of time of the outcome measure of student i.

β_{00} = grand mean of the outcome measure for all students and times.

β_{01} = average change in the outcome measure for each unit of time for all students and times.

e_{ti} = residual (error) representing the difference between the predicted outcome measure for student i at time t and the observed outcome measure, e.g., the within student random effect.

r_{0i} = residual (error) representing the difference between the grand mean outcome measure and the observed outcome measure for student i, e.g. the between student intercept random effect.

r_{1i} = residual (error) representing the difference between the predicted change in the outcome measure per unit in time and the observed change for student i, e.g. the between student slope random effect. (Raudenbush & Bryk, 2002). .

Latent Growth Curve Models. Latent growth curve models (LGCM, Bollen & Curran, 2006) use structural equation modeling methods to measure growth through estimation of initial status and rate of change as represented by latent variables. The path diagram for a linear change model is shown for an example with five observations in Figure 1.

Figure 1. Path diagram for a latent growth curve model

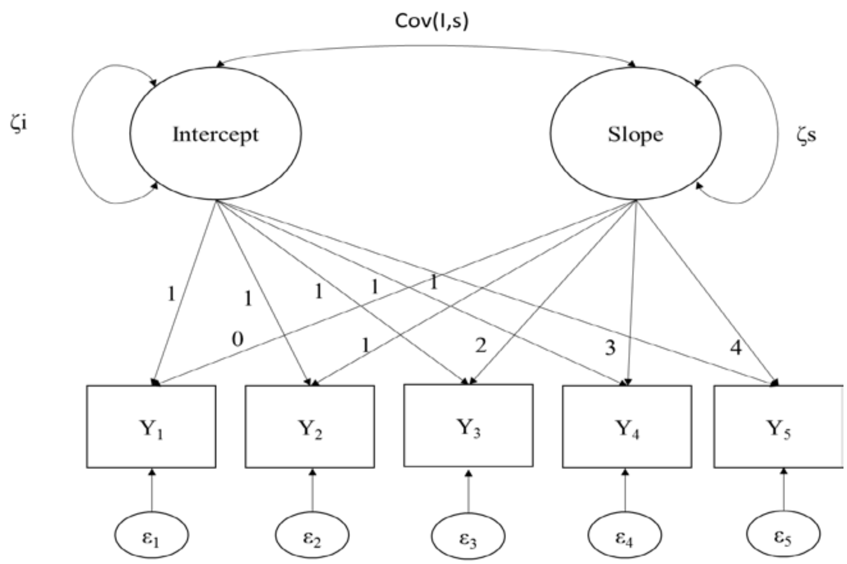


Figure 1. Path diagram for a latent growth curve model with five observed variables, random intercepts, and random slopes.

The equations representing a linear latent growth curve follow. First using matrix notation,

$$\mathbf{Y} = \boldsymbol{\Lambda}\,\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

$$\begin{pmatrix} y1 \\ y2 \\ y3 \\ y4 \\ y5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} I \\ S \end{pmatrix} + \begin{pmatrix} \varepsilon 1 \\ \varepsilon 2 \\ \varepsilon 3 \\ \varepsilon 4 \\ \varepsilon 5 \end{pmatrix} \qquad (2)$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \zeta.$$

The latent variable matrix $\boldsymbol{\eta} = \begin{pmatrix} I \\ S \end{pmatrix}$ is estimated with means $\boldsymbol{\alpha} = \begin{pmatrix} \mu i \\ \mu s \end{pmatrix}$ and variance $= \begin{pmatrix} \zeta\text{i} \\ \zeta\text{s} \end{pmatrix}$. The means of the intercept and slope capture the expected values for the initial status for person i and rate of change. The parameter estimates for the intercept and slope variances, ζi and ζs, capture

the individual variation around the average trajectory. In each case, the residuals, also called random effects, are assumed to be normally distributed with mean = 0 and variances ζi and ζs. The intercept and slope parameters may be allowed to covary: the covariance captures the relationships between the growth parameters, whether the intercept is higher when the slope is higher, lower when the slope is lower, or if they change in opposite directions. Individual growth modeling, using either multilevel modeling or structural equation modeling is referred to as a variable centered approach because variation in individual change patterns are reflected through the use random effects and covariates (Bollen & Curran, 2006).

GM modeling. One limitation of using either multilevel growth models or LGCMs is that the sample is assumed to come from a homogeneous population, which is often not the case. Therefore all heterogeneity in the population is represented in terms of intercept and slope random effects. One option is to introduce covariates as a technique for explaining variability in the intercepts and slopes. As described in Bauer and Curran (2003) and Bollen and Curran (2006), when the sample comes from a multiple group population where the groups are known, the categorical variable representing the group can be introduced into the model through the use of a covariate as a predictor of the intercept and slope. When used in this way, parameters other than the intercept and slope do not differ by group. In contrast, means, variances and covariances can be uniquely estimated for each group when using multiple group growth modeling (Bauer & Curran).

GM and LCG models (Bauer & Curran, 2003; Muthén, 2001; Muthén & Shedden, 1999; Nagin, 1999) are two related techniques used to study variations in how measures change over time which do not require the assumption that all individuals come from the same population. In addition, these techniques do not require known group membership. GM and LCG models reflect

the heterogeneity in the population through identification and classification of subjects into unobserved subgroups. GM and LCG models use response patterns within the repeated measures to identify the mapping of each individual to the unobserved subgroup, identifying the subgroup and unique change pattern with highest probability of its membership (Bollen & Curran, 2006). This technique is classified as a person centered modeling approach because it focuses on identifying the characteristics of the distribution of a group and the individuals belonging to that group based upon their responses patterns on indicator variables of interest (Muthén & Muthén, 2000).

GM and LCG modeling combine growth modeling as described in the preceding sections, and finite mixture modeling as developed in McLachlan and Peel (2000). Finite mixture modeling uses a probability density function created as a weighted sum of a finite number of density functions. This can be represented as f (y_j) = the probability density function of an outcome variable y_j, by a mixture of i = 1,2…g groups of normal distributions, weighted using $\pi(i)$ as

$$f\,(y_j) = \sum_{i=1}^{g} \pi(i)\; fi(y). \qquad (3)$$

This technique offers flexibility in modeling heterogeneity in the underlying population by representing subgroups using normal distributions with unique means and standard deviations. McLachlan and Peel (2004) provide a thorough discussion of theory and applications of finite mixture models, including presentation of applications, estimation procedures, and mixture of distributions besides the normal distribution. Mixture distributions can be traced back to Pearson, but not until techniques were developed to estimate the parameters did researchers fully realize their usefulness (McLachlan & Peel). The development of the expectation maximization (EM)

algorithm published in Dempster, Laird, and Rubin (1977) accelerated the use of mixture modeling.

According to Muthén and Shedden (1999), when there is evidence that a population is made up of multiple unobserved heterogeneous subpopulations, GM modeling can answer the question of how to relate the shape of a growth trajectory to the probability of a subject having that growth trajectory. GM models jointly estimate the latent growth curve parameters and the logistic regression parameters representing the probability of being in a particular growth curve class. Unlike LGCMs, LCG and GM models introduce a categorical latent variable which indicates a class of trajectories. This categorical variable is represented by a latent exogenous variable on the path diagram as shown in Figure 2.

Figure 2. Path Diagram for a Growth Mixture Model

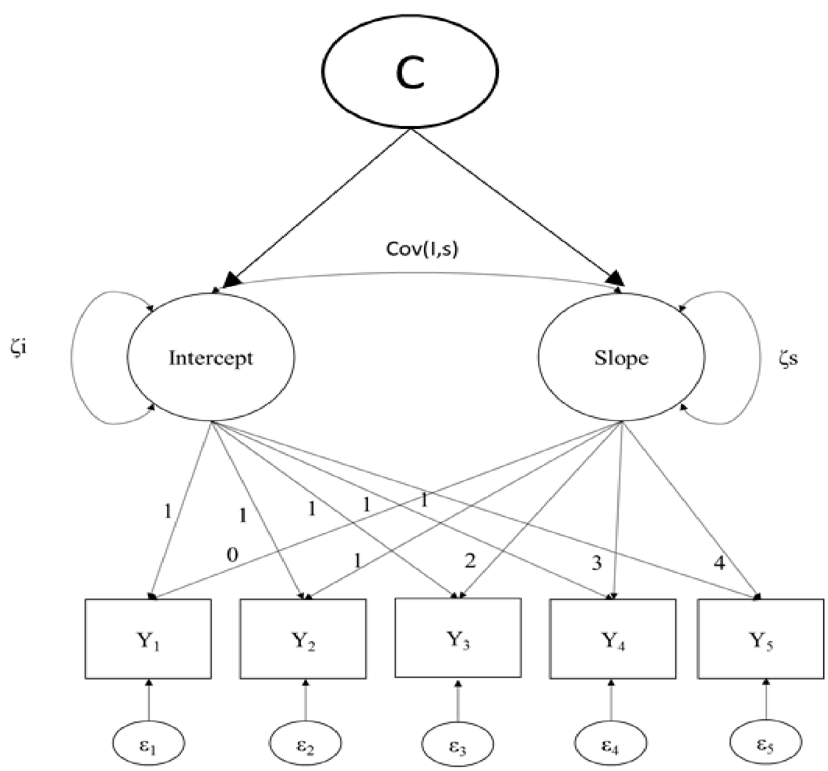


Figure 2. Path diagram for a growth mixture model with five observed variables, random intercepts, and random slopes where the variable C represents the latent class trajectory.

GM and LCG models assume that subjects are so different that observations are more accurately modeled using multiple groups of normally distributed variables each with a unique mean and possibly a unique variance (Muthén & Muthén, 2000). The model matrix equations shown below where c represents the latent class, indicate that each latent class can have unique model parameters, including the means, variances, and covariances of the latent variable and residual variances from the model.

$$\mathbf{Y} = \mathbf{\Lambda}\, \boldsymbol{\eta}_c + \boldsymbol{\varepsilon}_c \qquad (4)$$

$$\boldsymbol{\eta}_c = \boldsymbol{\alpha}_{ic} + \zeta_{ic}$$

The means, variances, covariances, and mixture weights must all be estimated in applied research. Further information is provided in the discussion on multilevel GM modeling.

EM algorithm. Although latent class is unknown, the model uses the patterns of the observed variables to assign individuals to classes within the E-step of the EM algorithm (Dempster, Laird, and Rubin, 1977; Muthén & Shedden, 1999; Muthén & Muthén, 2000; and others). Given the number of classes as an input, starting values assign each subject partially to each class. The probability of class membership is modeled using a multinomial logistic regression model, in this case class 1 as compared to class K, in the equation that follows.

$$\log \frac{P(cij=1)}{P(cij=K)} = \lambda \qquad (5)$$

Class assignment generates a modeled log odds and model implied posterior probability (Nagin, 1999). The EM algorithm involves estimating individual growth models and also estimating each individual's most likely class membership in an iterative process, seeking the minimization of the log-likelihood and model convergence at optimization. Once optimization has been reached, each individual has a posterior probability for each class representing the likelihood of

assignment to a particular class. When a single class assignment must be selected, the subject is assigned to the class with the highest posterior probability.

GM Modeling Context from Current Study. To provide some context, consider the current substantive research question of how student behavior changes over time as measured using the DBR-SIS. The composite measure ranges from 0 to 30, with 30 representing the optimal situation where the student is scored as engaged and respectful 100% of the time during the assessment period and disruptive 0% of the time. I hypothesized that initial status and degree and direction of change over time are different for different groups of students. For example, students who are consistently well behaved receive high scores with little variability (Class 1). In addition, these scores do not change much over time because it is impossible for students to be scored better than 100%. Another group of students may not be as well behaved. For example, if these students are observed with the optimal behavior about 80% of the time for the first rating, they receive an average initial score of 24. These students may start the period at 24 and show improvement, as represented by a positive slope (Class 2). Class 1 may be considered ideal, but Classes 1 and 2 may be considered normative in that students in both groups still function and learn well in the classroom environment. Now, consider a third group of students who are only on task 70% of the time with an average score of 21, and their behavior deteriorates over the rating period (Class 3). The students in Class 3 are likely in need of support services to help them improve their behavior. If LGCMs or multilevel models are used, the results only provide a single average trajectory, representing the average intercept and slope for all students. A graph of these three groups and the average is shown in Figure 3.

Figure 3. Example Graph of Growth Trajectories

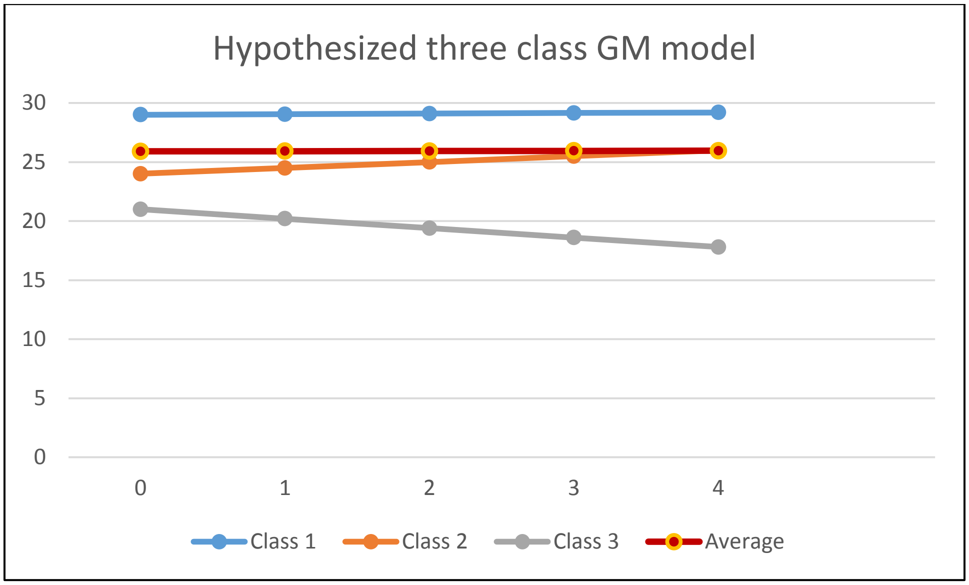


Figure 3. Example of the graph of the growth trajectories for students as represented by three separate classes, Class 1, 2 and 3, and the average which includes all students.

The average trajectory does provide us with some information on how students change across time, but it is not useful in all contexts. The model representing the average trajectory can be relaxed to include random effects, representing the distribution of the differences between students within the same latent class, between latent classes, and between teachers. The variance of the random effects reflects the variability in the intercepts and slopes across individuals. Further, residual variability in the intercepts, slopes or over time can be explained by the use of known covariates such as gender, race, and disability. However, it is possible that by using averages and known covariates, we are not fully explaining behavior variability because averages can tend to mask unique patterns in the outcomes. For example, the students who are in Class 3 start low and deteriorate across time. These students are likely in need of special services to provide assistance in improving their behavior. If we instead used LGCMs and covariates,

these differences would be reflected in the changes in model parameters from that covariate. If these students are not represented by one of the known covariates, or are distributed across the covariates, these unique patterns may be averaged across the known groups and remain hidden. The three groups of students described above are not directly observable, in that there is no covariate that explicitly defines the groups. Unfortunately, we do not know with certainty how students are distributed across these three groups, nor do we know in what proportions. GM modeling is a statistical technique that combines features of mixture models and LGCMs.

As described in Chapter 1, GM models are used with increasing frequency in educational and psychological research. However, as described in Bauer (2008), there are limitations which must be considered whenever adopting this technique. For example, GM modeling is designed for situations where the underlying outcome data lack normality, but it is assumed that once the latent classes are identified, outcomes are normally distributed within classes. Bauer and Curran (2003) found that multiple classes may be extracted from a homogenous population when the outcome measures are not normally distributed. This concern can be reiterated as follows: does lack of normality imply a heterogeneous population or just that the outcome measure is not normally distributed? Additionally, the latent subgroups identified from modeling may or may not follow substantive theory. Muthén (2004) recommends that all latent class modeling, including GM modeling, should be done with strong interdependence on substantive theory to drive modeling decisions. In addition, estimation procedures are often wrought with problems including Heywood cases with negative variances, lack of convergence, and non-positive definite variance-covariance matrices. Although these limitations are areas for concern, GM modeling can be used as an exploratory method to identify patterns that simpler techniques such as LGCM and multilevel models would miss.

LCG Models. LCG models represent a type of GM model in which the heterogeneity is reflected in the model through fixed effects only. When using LCG models, all error variances except residual variances are constrained to zero in the models. Differences in the trajectories are depicted by intercepts and slopes that vary by latent class. As reported in Kreuter and Muthén (2008) and Ram and Grimm (2009) the LCG model represents an efficient starting place when conducting a GM modeling study. LCG models estimate fewer parameters than GM modeling, and typically they do not experience the convergence problems introduced when attempting to estimate parameters, particularly variances, that are close to zero.

Multilevel GM modeling. Multilevel structural equation modeling (Kline, 2011; Stapleton, 2006; Stapleton, 2013) is foundational to examining multilevel GM modeling, considering intercepts and slopes as latent variables at the within cluster and between clusters level. Multilevel structural equation modeling partitions the variance into the amount within cluster and the amount between clusters. Different covariates can be used at the within classroom level as compared to the between classroom level to further explain the variance at each level. For example, we may hypothesize that students in classrooms with more experienced teachers have exhibit fewer problem behaviors than less experienced teachers. In this type of model, a variable representing teacher experience would be entered as a covariate at the classroom level.

When modeling heterogeneity where the individuals are clustered such as in the current study, with students nested within classrooms, multilevel GM models provide a method for reflecting the lack of independence and heterogeneity (Muthén, 2004; Palardy & Vermunt, 2010; Vermunt, 2010). There are a variety of approaches for describing the heterogeneity in the growth trajectory. Multilevel GM models can be used to model heterogeneity based upon individuals, and then cluster them or they can model heterogeneity between clusters, assigning all individuals

in a cluster to the same latent class, or they can use both techniques (Palardy & Vermunt, 2010). Palardy & Vermunt (2010) and Vermunt (2010) describe modeling and interpretation of three level GM models, representing a three by three matrix of possible configurations, with continuous, discrete or both latent variables at the within or between levels.

In the case of the research described herein, consider a three level model, with observations nested within students and students nested within classrooms. The equations for a GM model where the assignment to unobserved classes is completed at the individual level for k classes using multilevel notation follow.

Level 1: (7)

$Y_{tijk} = \pi_{0ijk} + \pi_{1ijk}\, Time_{tij} + e_{tijk}$

Level 2:

$\pi_{0ijk} = \beta_{0jk} + r_{0ijk}$

$\pi_{1ijk} = \beta_{1jk} + r_{1ijk}$

Level 3:

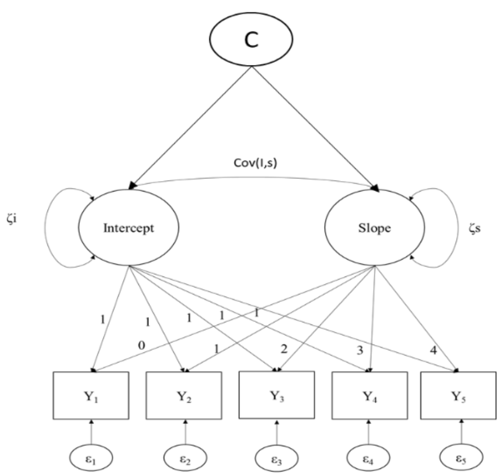
$\beta_{0jk} = \gamma_{00k} + \mu_{0jk}$

$\beta_{1jk} = \gamma_{10k} + \mu_{1jk}$

The definitions are aligned with those provided for multilevel models, but with i = individual, j = cluster, and k = class. Starting at Level 1, every observation is estimated by an intercept and slope for person i, cluster j, class k, and a residual $e_{tijk.}$ The Level 1 intercept and slope are estimated at Level 2 for each individual with random effects r_{0ijk} and $r_{1ijk.}$ Similarly, the Level 2 intercept and slope are estimated at Level 3 for each teacher with random effects μ_{0jk} and $\mu_{1jk.}$ As a result, for this highly parameterized model, we have class varying intercepts, class

varying slopes, class varying Level 2 residual variances for both the intercept and slope, class varying Level 3 residual variances for both the intercept and slope, and class varying residuals.

Building upon the path model picture from Figures 1 and 2, a three-level GM model can be depicted as a combination of the two. Figure 4 (a) represents a within teacher level GM model where the latent class is modeled at the individual level. Figure 4 (b) represents the between teacher level model, in which outcome variables, intercepts and slopes are represented as latent variables. Just like the LGCM, the model can include modeled covariances, random intercepts and random slopes. Model choice depends upon the measures, populations, and research questions. For the current study, prior research has indicated that although the variability in the DBR-SIS that can be explained by the influence of the teacher ranged from 26% to 41%, the majority of the variability is still at the student and observation level, supporting the assignment of latent class at the student level, referred to as the within teacher level (Kooken et al., under review).



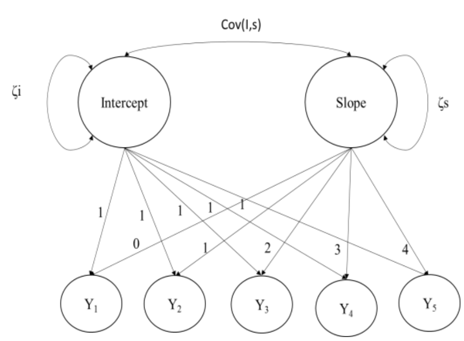


Figure 4. Multilevel GM Path Model.

Figure 4 (a) Figure 4 (b)

Figure 4. Path model for a multilevel GM model where Figure 4 (a) represents the within cluster level path diagram with C representing the latent class variable. Figure 4 (b) represents the between cluster path diagram where individuals are clustered at the between teacher level.

As mentioned in Chapter 1, the amount of research on multilevel GM models is limited. One area of current research involves investigation of effects of ignoring either the clustering or the mixture in GM model specification and interpretation. In a methodological study incorporating repeated observations nested within students, Muthén and Asparouhov (2009) found that heterogeneity at level 1 (within student) was incorrectly assigned in the models to level 2 random effects when mixture modeling was not used. When mixture modeling is used, the within subject heterogeneity can be properly reflected in the model by the varying intercepts and slopes of the latent trajectory classes. Chen, Kwok, Luo, and Willson (2010) conducted a simulation study to investigate the effect of ignoring a level of nesting on GM models. Using data generated for a 3-level model with 2 subpopulations and modeled correctly or modeled ignoring the third level, they found that ignoring the nesting resulted in no impact on fixed effects, but reduced accuracy in classification, biased variance estimates, and biased standard errors. Specifically, the correct number of classes was identified in 97% of the cases with the true model and 93% of the cases with a misspecified model. Hit rates representing the proportion of individuals properly classified indicate slight improvement for the correct model of the misspecified model, 91% versus 88%. Misspecification also resulted in understated intercept standard errors and overstated slope standard errors. Ignoring nesting also resulted in reduced accuracy of the student level (Level 2) random effects but not the Level 1 residuals. Estimation errors were exacerbated for higher interclass correlation coefficients (ICC). The results from these studies support the need to incorporate a multilevel design to reflect the nested structure of the data.

From an applied perspective, there have been a few studies which have reflected the nesting of Level 2 units within Level 3 units, e.g., students within classrooms. Owen et al. (2015) examined change in client functioning resulting from psychotherapy over short term treatment. Individuals were nested within therapists, but the researchers were not able to examine the Level 3 effect due to non-convergence. Morin, Maiano, Marsh, Nagengast and Janosz (2013) studied self-esteem trajectories of high school students where students were nested within schools using GM modeling at the student level. Multilevel GM modeling was not used of an insufficient number of Level 3 units, only 5 schools. As an alternative, they used a design based correction to standard errors and group mean centering to improve estimation accuracy. Kuntsche Otten, and Labhart (2015) investigated young adult alcoholic drinking patterns where incidence of drinking was nested within event. Rather than use Multilevel GM modeling, they used a three step approach in which GM modeling was used to identify evenings with similar drinking patterns. Simmons et al. used GM modeling in a study examining the effect of formative assessment on student reading achievement for students receiving a reading intervention. The analysis first used GM modeling to identify different progressions, then used propensity scores to match students who received the intervention with student who did not receive it. In a final step, they used multilevel modeling to test the intervention effect. In all of these cases, data were clustered, but the researchers did not use Multilevel GM modeling. Palardy and Vermunt (2010) modeled heterogeneity between schools in student cognitive growth using Multilevel GM modeling, but they used the between Level 3 unit model, assigning all individuals in a cluster to the same latent class. No applied Multilevel GM modeling study which assigned individuals to classes and modeled the Level 3 effects was identified in the research literature.

Measurement invariance. An important distinction between LGCMs, GM and LCG models is the level of heterogeneity that is reflected in the models. Measurement invariance generally refers to an important measurement property sought for variables used in educational and psychological assessment and research to measure latent traits—traits that are not directly observable—such as intelligence, content knowledge, attitudes and beliefs. A measure that is noninvariant may be considered to be biased, which invalidates comparisons between people or subgroups (Brown, 2006; Kline, 2011; Vandenberg & Lance, 2000). For example, boys may consistently receive a lower score than girls on a measure, even when trait levels are the same. In psychometrics, more specifically as related to item response theory (IRT), invariance, referred to as a "cornerstone of IRT" (Hambleton, Swaminathan, & Rogers, 1991, p. 18), refers to the property whereby model parameters do not vary across groups within the population.

For the purposes of this research, we are concerned with the application of invariance to structural equation modeling. Van de Schoot, Lugtig, and Hox (2012) provide a thorough discussion of terminology and procedures for testing invariance in structural equation modeling. Figure 5 provides a graphic organizer representing the nested nature of the levels of invariance. When factor loadings are equal across groups, the measure has metric (weak) invariance. When both factor loadings and intercepts are invariant, the measure has scalar invariance. When factor loadings, intercepts, and error (residual) variances are equal across groups, the measure is considered to have strong measurement invariance, also known as full uniqueness invariance. An ongoing debate continues as to what type of invariance must exist in order to justify the use of a particular measure or test to compare groups (Byrne, Shavelson, Muthén, 1989; Kline, 2011; van de Schoot, Lugtig, & Hox, 2012; Vandenberg & Lance, 2000). For a measurement model with

latent variables and manifest indicators to be fully invariant, the items, the factor loadings, intercepts, and residual variances should be invariant across groups (Kline, 2011).

Wu, Li, and Zumbo (2007) provide a set of mathematical conditions that create a line in the sand to distinguish between invariance properties that exist for optimal performance of a measure in group comparisons and those that may not exist but do not invalidate the use of the measure as follows:

1. the model specification (number of factors and loading patterns)
2. the regression coefficient,
3. the regression intercept term,
4. the regression residual variance,
5. the means of the common factors,
6. the variances of the common factors, and
7. the covariances among the common factors (Wu, Li & Zumbo, 2007, p. 3).

Figure 5. Concept map of invariance

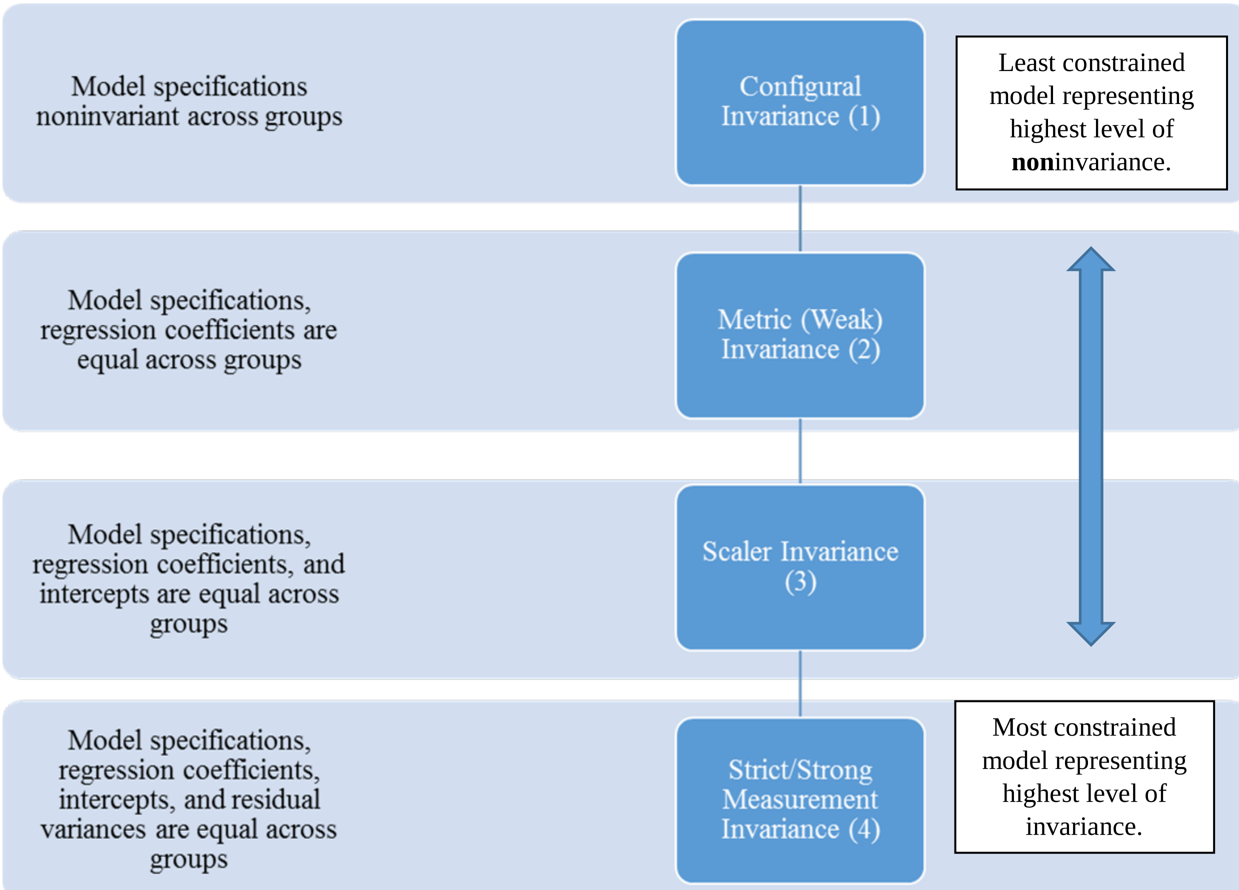


Figure 5. Concept map of invariance.

The parameterization of the GM model contains a measurement model portion that reflects the first four items in the above list and a structural model which reflects the last three. The measurement model part contains the latent growth curve model parameterization, with a fixed number of factors represented by intercepts and slopes, fixed regression coefficients and intercepts, and an estimated residual variance. The structural part is represented by the factor means and variances, with factors represented by the intercept and slope parameters, and covariances between the intercepts and slopes. The parameterization of the growth model provides a guarantee of invariance with respect to items 1, 2, and 3 above. The invariance

question as it applies to GM modeling applies to items 4-7. By choosing a statistical model that assumes heterogeneity in the underlying population, we hope to capture the essence of the heterogeneity in the model parameterization. We expect there to be differences in the means, variances, covariances, and residuals, but we need to clarify to what extent, if any, the differences in means, variances, covariances and residuals reflect true measurement invariance and to what extent do they reflect population heterogeneity? In other words, we need to test whether the measure is simultaneously sensitive enough to pick up heterogeneity and also structured enough to measure the same construct the same way across the latent subpopulations.

In GM models, not only can the subgroups vary in their intercepts and slopes, but the models can also reflect variability in the residual variances. The residual variance is a modeled parameter which measures the variability in the outcome that is not captured by the model. The residual is the difference between the observed and predicted value, and variability in the residual over all subjects and observations provide the distribution with a mean and variance. For a growth model, the residual variances can be either fixed across time points or allowed to vary across time points. For example, when residual variances are fixed, this assumes the modeled residual has the same mean and variance for all observations. When residual variances are allowed to change across time points, this means that the distribution mean and variance of the parameter is allowed to freely vary across time. Residual variances also can be fixed or allowed to vary across latent class. This means that the residuals which are time invariant or time noninvariant can be modeled to reflect different distributions for each class.

The within class variance in slopes, intercepts, and residual variances can be estimated, but whether or not they vary across groups depends upon the study. With real data, correct model specification is not known, and so uncertainty exists around how the level of invariance and

model specification affects the number of latent classes and latent class assignment. As reported in Enders and Tofighi (2008), published research lacks consistency in how variance components are treated, whether they are fixed or allowed to vary across Level 2 units or across classes. Enders and Tofighi conducted a simulation study in which they examined the impact of constraining residual variances when it varies by class. Enders and Tofighi found that misspecification of residual variance parameterization negatively affects the accuracy of the parameter estimates, number of classes, and class proportions.

Applied GM model invariance research. Few researchers have considered the impact of variations in the residual variance on model results. Morin et al. (2011) analyzed the effect of untested invariance assumptions on model fit and accuracy in GM modeling. Using the example of studying adolescents' trajectories of anxiety, the authors modeled change using a LCG modeling with total invariance, a GM model where intercepts and slope variances were allowed to vary within latent class but were invariant between classes, and a GM model where intercepts and slope variances were noninvariant between classes. The authors did not address characteristics of the Level 1 (within person) residual variance. The authors interpreted their findings to mean that the fully noninvariant model provided the best fit to the data. Although information criteria identified the noninvariant models as superior, this was at expense of deterioration in entropy, a measure of class assignment accuracy, implying decreased accuracy (Akaike, 1977). Entropy deteriorated from .90 for the 5-class LCG model to .79 for the 5-class noninvariant model. This study represents an important contribution to the literature by investigating the impact of variance assumptions using real data. Nonetheless, the data were not nested, so the analysis was not multilevel. These challenges are exacerbated when modeling

clustered data where lack of independence requires the use of multilevel mixture modeling. In addition, the study did not consider the issue of invariance of Level 1 residual variance estimates.

Student Classroom Behavior and Assessment

As a student progresses through school, factors related to the environment and maturation act to produce changes in student behavior across time. Teachers expect students to learn and exhibit behaviors that are conducive to learning, respectful of others, and self-regulating (Welsh, Miller, Kooken, Chafouleas, & McCoach, in press). Often teachers are skilled at identification and documentation of behavioral change on a student-by-student basis. The extant literature provides a thorough and active investigation of student behavior profiles from a diagnostic perspective, with research on interventions to improve student behavior such as social skills and self-regulation, and the impact of behavior on other student outcomes such as academic achievement, some of which are reviewed below. Research is limited, however, defining a typology of characteristics of student behavior change over time in the classroom environment between students and across grade levels during an academic year. Theory and techniques for measurement of academic achievement at a specific point in time as well as change over time are well established and are the focus of the current national interest in cognitive testing (No Child Left Behind, 2003). Yet despite almost universal agreement in the importance of classroom behavior to student academic success, no published studies of the characteristics of student classroom behavior growth trajectories across time were found.

Current directions in school behavior screening. Student academic success is a function not just of efforts in the cognitive and attitudinal domains, but also is strongly affected by the behavior of students in the classroom (Georges, Brooks-Gunn, & Malone, 2012; Hinshaw, 1992; Wentzel, 1993). The classroom dynamic introduces unique demands on the student to

adjust to teacher expectations and to a social environment that includes many peers. Learning related skills including behavioral regulation and social competence can be predictive of academic success including growth in reading and math between kindergarten and second grade (McClelland, Acock, & Morrison, 2006).

While 17% of all K-12 school aged students require mental health services, approximately one third of all students have psychosocial problems manifested by student behavior that interferes with their ability to engage in learning (Epstein, Atkins, Cullinan, Kutash, & Weaver, 2008). Identification of these students for special intervention services and tracking their progress is an important step towards improving behavioral and academic student outcomes. Tools with solid psychometric properties are needed to aid researchers and practitioners in addressing student behavior interventions and research, for both screening and progress monitoring (Glover & Albers, 2007). Student behavioral assessment tools include systematic direct observation, rating scales, and direct behavior rating (Chafouleas, 2011). Systematic direct observations require intensive data collection techniques involving trained professional, which are costly in terms of training, personnel and time. Many psychometrically tested and commercially developed rating scales are available which measure student behavior, (e.g., BESS—Kamphaus & Reynolds, 2007; Social Skills Improvement System—Elliott & Gresham, 2007) but these measures are more time intensive, expensive, and their sensitivity to change is not well established (Chafouleas, 2011).

An assessment specifically designed to measure change in classroom behavior patterns is the DBR-SIS (Chafouleas, 2011). With the DBR-SIS, the rater, who is most likely the teacher, rates a student on a single behavior, measuring the percentage of time a student exhibits the behavior construct in question during a predefined observation period. For example, a teacher

would use a scale of 0-10 to rate the percentage of time a student is exhibiting behavior which indicates academic engagement during reading instruction. The DBR-SIS has desirable qualities for use in screening and progress monitoring at a single point in time in that it is contextually relevant, technically defensible, and usable (Chafouleas). Screening refers to identification of students at risk for behavioral difficulties. Progress monitoring refers to the assessment of change in this behavior over time, especially in the presence of an intervention, and the use of longitudinal data supports progress monitoring. Contextually relevant means the behavior is observed and the measure is recorded at the time and place the behavior occurs. Technically defensible refers to psychometric strength of the measure. Usability refers to the efficiency of the measure in terms of time, ease of use and cost (Chafouleas).

Research describing student behavior change. In describing changes in behaviors over time, other researchers have focused on either maladaptive behaviors or behavioral constructs in contexts outside the classroom. For example, using multilevel growth analysis, Bongers, Koot, van der Ende, and Verhulst (2003) found children ages 4-18 changed in levels of internalizing and externalizing behaviors over time. These trajectories varied significantly in both initial status and rate of change for males versus females, with boys showing higher initial status for externalizing behaviors and lower rates of change over time. There were no gender effects on initial status for internalizing behaviors, but girls had higher rates of change over time. Externalizing behaviors such as oppositionality, aggression, and conduct problems were shown to decrease whereas delinquent behavior increased. Similar results were found by Reynolds, Sander, and Irvin (2010), although they studied internalizing and interpersonal behaviors in a cohort of students from kindergarten through 5th grade and found significant differences by age, gender, and socioeconomic status (SES). Keiley, Bates, Dodge, and Pettit (2000) found

internalizing and externalizing behavior varied by initial status and rate of change based upon gender and SES between kindergarten and seventh grade. Using LGCMs, social skills of children as rated by parents were found to improve from kindergarten to third grade, while social skills in school settings decreased in the same time period (Chan, Ramey, Ramey & Schmitt, 2010). Studying the sons of low-income parents starting from around 18 months of age, Shaw, Gilliom, Ingoldsby, and Nagin (2003) found evidence of four distinct trajectories varying in level and rate of change of behavior problems, but unfortunately only risk factors associated with characteristics of the mother were tested for significance. Although all of these studies provide useful evidence for comparison to classroom behavior patterns, none of these studies focus specifically on classroom behavior needed for academic success.

Research defining student behavior typologies. The classification of individuals within social, emotional, and behavioral constructs is central to social science research, defining groups by their attributes on one or more dimensions, describing how these dimensions relate to desired outcomes, and identifying how to most effectively select interventions for improvement (Bailey, 1994). In addition, to fully understand human development, we need to understand how individuals change over time (Grimm, Ram, Estabrook, 2010). Typologies need to be based upon well researched statistical procedures and ideally provide a set of classes, description of the students within those classes, and a substantive framework grounding the classification structure to theory.

A review of the literature on characteristics of student behavior in school suggests that although many students exhibit behavior conducive to success in the classroom throughout the year, there is still much that needs to be learned with regard to the students who struggle. Using cluster analysis and the BASC TRS-C, Kamphaus et al. (1997) identified 7 clusters, representing

students with adaptive behavior in the largest cluster (53%) to students with disruptive behavior (8%) to students with severe behavior problems (4%). These results were later replicated by Kim et al., (2010), also using the BASC TRS-C, in which the 7 clusters with similar attributes were identified using 6th graders from ethnically diverse low SES communities. DiStefano and Kamphaus (2006) conducted both cluster analysis and latent profile analysis using the BASC TRS-C on children ages 6-11. The cluster analysis identified 7 groupings while the latent profile analysis identified 3 groupings. Class 1 (36%) represented students with optimal emotional and behavioral adjustment. Class 2 (42%) represented typical adjustment, and Class 3 (22%) represented students who were functionally impaired. DiStefano and Kamphaus (2008) used the results from this cluster analysis to analyze change over time using latent growth curves. Their research investigated the development of 162 students in grades 1-3, and they found all groups, including those at risk, exhibited a linear developmental pattern. Students in the average risk and those in lower risk groups had negative slopes for adaptive skills, internalizing problems, and externalizing problems, except for average risk students whose slope was not significant for externalizing problems. Even though these slopes were significant, the values themselves were not large, less than 2% change per year. The authors suggested the use of growth mixture modeling as an area of future research in situations where data are collected over a longer period of time. Using the BESS, Dowdy et al. (2014) found three latent groups using latent profile analysis. These groups consist of a normal group (47%), a slightly elevated group (45%), and an elevated risk group (8%).

Based upon these studies, student behavior patterns appear to include a large adaptive group which is over 50% and several mal-adaptive groups, including students with disruptive behavior and students who have severe behavioral problems. Differences were identified in class

assignment by gender, with males representing a disproportionately high number in the less adaptive groups and low representation in the more adaptive groups and race with African-American students over-represented in the less adaptive groups (Kamphaus et al., 1997).

This body of research provides some descriptive typologies of student behavior trajectories as measured by the behavior rating scales, but there are some limitations. Kamphaus et al. (1997) classified students based upon a cross sectional measurement, without providing information on patterns of change for the 7 clusters. DiStefano and Kamphaus (2008) only provided two classifications, lower risk and average risk; the trajectories of high risk students are not known. Also, neither of these studies used GM modeling to identify classes of growth trajectories, and this technique may identify different student groupings based upon the behavior change patterns. None of these studies considered variability as a feature contributing to identification of the class assignment. Specifically, the extent research has not fully investigated how intraindividual variability and interindividual heterogeneity around modeled trajectories might prove to be informative in defining classroom behavior typologies. Finally, because both prior studies used the BASC TRS-C, students were classified using measures of a range of constructs, many of which are used for purposes of screening and diagnosis of a disability (Reynolds & Kamphaus, 2006).

Miller et al. (2015) compared the performance of five behavioral measures to screen social, emotional, and behavioral risk among school age children. Their focus was identification of at-risk students based upon the DBR-SIS, the BESS, the Social Skills Improvement System (SSiS; Elliott & Gresham, 2007). ODRs, and teacher nominations. They found that the BESS, ODRs and teacher nominations were the most conservative identification method with 18% identified as at-risk on the BESS, from 7-12% identified using ODRs, and 4-5% identified using

school nomination. The DBR-SIS and SSIS are both broader range screeners identifying 36-39% at risk using the DBR-SIS and 31-36% identified using the SSiS. Cut scores for identifying students at risk on the DBR-SIS used an average score calculated using from 6 to 10 measures. The cut scores were set using receiver operating curves with the BESS as the gold standard measure of at-risk behavior (Johnson et al., in press). These results provide a benchmark for evaluating the reasonability of a proposed typology by indicating what proportion of the population has behavior that is most problematic.

In summary, the research on student behavior described above focuses on defining the shape and scale of student behavior change and identifies groups of students based upon behavioral risk. None of these studies has identified the characteristics of student classroom behavior patterns over time, quantifying patterns of initial status, change and variability, grouped by students with similar behavior patterns. Nonetheless, these studies provided valuable resources for the current study to validate findings of proportions of students in various categories, use of covariates, and other outcome measures.

Contextual effects of student behavior. For the current study, observed classroom behavior patterns were used to identify typologies of student behavior. Often in single and multilevel regression contextual variables such as demographic groups and concurrent outcomes are used as covariates to study variation in the outcome variable. Although this approach was not taken for the current study, validation of the classes requires consideration of whether patterns in these context variables follow expectations based upon what is known from prior research. The results suggesting a set of typologies for student classroom behavior should include an investigation of the construct validity of the latent classes using concurrent and distal outcome measures.

Race, gender, ethnicity, special education status all have been shown to be significant predictors of variability in classroom behavior problems. For example, boys exhibit more externalizing behaviors than girls (Achenbach, 1991; Bongers et al., 2003; Kamphaus et al., 1997; Schaeffer et al., 2006). Students with disabilities are also more likely to be at higher risk for problem behavior (Lane, Carter, Pierson, & Glaeser, 2006). Students in minority race and ethnic backgrounds are also overrepresented in the higher risk groups (Kamphaus et al., 1997). Further, the BESS T-score was used as a criterion measure of classification in terms of level of behavioral risk, with scores above 60 indicating student risk level and scores above 70 indicating elevated risk level. In any typology, the expectation is that students with BESS scores above 60 would be classified into higher risk groupings.

Office Disciplinary Referrals (ODRs). ODRs represent a rich data source for screening and monitoring progress of student behavior in schools and provide many benefits (McIntosh, Frank & Spaulding, 2010). Data on ODRs are typically already tracked and available, and many schools use online data collection systems such as the School-wide Information System (SWIS; May et al., 2003). ODRs measure problem behavior in that they count the number of times a student has been identified as breaking a school rule. Data collection of ODRs also typically include information on place, time, and context to help identify patterns of behavior problems. Positive Behavioral Intervention and Supports (PBIS) suggests that 0 to 1 ODR represents low level risk for problem behavior. McIntosh, Frank and Spaulding (2010) and Pas, Bradshaw and Mitchell (2011) investigated whether ODRs could be used to screen students for problem behaviors. Both studies used receiver operating curves (ROC) analysis to identify optimal cut scores that identify students with problem behaviors. Both McIntosh et al. and Pas et al. suggested that two ODRs were predictive of elevated risk for problem behavior in the classroom.

The expectation is that students with two or more ODRs would be classified into higher risk groupings in a behavior typology.

Response to Intervention and Single-case design. Response to Intervention (RTI) represents a broad initiative of theory and tools used to identify and provide early intervention to students who are at risk (Fuchs & Fuchs, 2006). Tools with strong psychometric properties are needed to aid researchers and practitioners for both screening and progress monitoring (Glover & Albers, 2007). RTI is used as part of a problem solving model in schools to identify an intervention that is tailored to an individual student based upon their needs assessment. The student is observed for a baseline period during which repeated measures are taken and plotted on a time series graph. Established standards delineate the point when the intervention can be administered to maximize the validity of considering changes in the outcome measure as evidence of the effectiveness of the intervention.

Improvement in student behavior may be operationalized by a favorable change to the trend line, change in the mean, or a change in variability (Kazdin, 1982). The effectiveness of classroom behavior intervention is often evaluated using single case design techniques including a visual examination of a graph of repeated measures as a time series (Riley-Tillman & Burns, 2009). Visual analysis of single case research results includes identification and assessment of six visual features of the time series graph: (1) level, (2) trend, (3) variability, (4) immediacy of the effect, (5) overlap, and (6) consistency of data patterns across similar phases (Kratochwill et al., 2013, p. 31). Variability refers to both the statistical definition which can be measured using range, variance or standard deviation and to data that are so inconsistent that it precludes predicting a pattern for baseline and delays administration of the intervention. Practitioners may also evaluate the effectiveness of an intervention using a measure of effect size. One example of

an effect size measures of the amount of change in the level such as standard mean difference, although there are many others (Horner, Swaminathan, Sugai, & Smolkowski, 2012).

Within person variability can confound efforts to identify students for RTI and evaluate the effectiveness of interventions. In some cases, the time series plot may show so much variability that there is no identifiable trend subsequent to the intervention that would indicate a change, resulting in a "failure to establish a consistent pattern within a phase" (Kratochwill, 2013, p. 32). Kazdin (1982) and others suggests that averaging multiple observation points is sometimes a good strategy to improve detection of changes in level and trend across time. However, if the variability in the data has substantive meaning, that information is lost when taking averages.

Although the characteristics of level, trend and variability are used in LCG modeling to sort students into groups, there may also be distinct profiles of residual within student variability. In other words, we are often looking for the signal in the noise, but what if the noise is the signal? That is, classroom behavior may be a constructs for which variations in the level of variability can tell us as much or more than the level of the means and trend lines. There is a need for applied researchers to focus attention on modeling and interpretation of this variability. An example of using location scale models to examine this type of construct is explored next.

Location Scale Models. There are techniques that could be used to augment the current study to reflect other ways of modeling residual variance. In the current study, because students were measured 30 times over a period of one year, it would be of interest to identify factors which contribute to differences in both means and variances. In Hedeker and Mermelstein (2007) and Li and Hedeker (2011), the authors studied the effect of smoking on affect in adolescence using location scale (LS) models, modeling the random effects of means and random effects of

variances. These models were used to identify variables which acted to increase or decrease within subject and between subject variance in measures of positive affect and negative affect in an ecological momentary assessment study. For example, when subjects use cigarettes to help regulate negative mood, this produces changes in both negative affect (e.g. feelings of sadness or anger) and positive affect (e.g. feelings of confidence or happiness). The action of smoking to help regulate negative mood decreased the mean level of negative affect but also decreased both between and within subject variability in negative affect. Further, when subjects used cigarettes to regulate negative mood, this resulted in increased positive affect and decreased variability in both within and between level variance in positive affect. This type of information greatly improves understanding of the mechanism which triggers the use of cigarettes and is not available under normal multilevel modeling. An analysis similar to LS models used in Li and Hedeker (2011) applied to the current problem would greatly enhance our understanding of the effect of covariates on variability of the DBR. Unfortunately, it is not known what effect this type of analysis would have on the class assignment if incorporated into the growth mixture model, but it is likely to change the class composition. Mplus does not currently support this methodology.

Technical Considerations in Multivariate Modeling

Using sophisticated modeling techniques such as GM modeling requires the knowledge of how to make many decisions regarding model fit, class enumeration, and how to handle model convergence problems. The research literature provides a solid knowledge base to guide these decisions.

Model building. Ram and Grimm (2009) developed a model building heuristic for GM models to aid in the identification of well specified models that identify the proper number of

latent classes. Ram and Grimm emphasize the importance of identifying the patterns represented in the outcomes, focusing on the shape, the level, and the differences. Shape refers to the growth curve shape, level refers to the means, and differences reflects the level of variation between elements in the model. They recommend a four step approach which includes problem definition, model specification, model estimation, and model selection and interpretation. The model selection and interpretation phase involves using an iterative process to identify a best model. This process considers "(I) optimal number unobserved groups, and (ii) the type and extent of differences between and within those groups (Means, Means + Covs, Means + Covs + Pattern) (Ram & Grimm, 2009, p. 567). When a model is identified as problematic, they recommend reformulating the specifications for the model, but in some cases, the model cannot be remedied and must be discarded from consideration.

Model Fit. Model fit for statistical models tells us whether the model accurately represents the observed data and is a necessary part of identifying the preferred model. For structural models containing latent variables, model fit cannot be assessed on an absolute basis, but only using relative fit information. Relative indicators such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the sample size adjusted BIC (SABIC) help to identify the best fitting model among a variety of choices (Kline, 2011; McCoach & Black, 2008; Tofighi & Edwards, 2008). The formulas for the AIC, BIC, and SAIBC are as follows:

AIC= -2 LL + 2P

BIC= -2LL + ln (N) * 2P

SABIC = -2LL + ln ((N+2)/24)*2P

Where LL= log likelihood, P=number of parameters, and N=sample size. For these information criteria, the model with the lowest value is considered to be the best fitting model.

In some cases, the information criteria continue to decrease as more classes are added, obscuring the choice of an optimal model. Morin et al. (2011) found that models continued to converge after increasing the number of classes up to 7 classes. Petras and Masyn (2010) identified a technique to aid in identifying a threshold for model selection. They employed the technique of “elbow plots” similar to scree plots in principal components analysis (Netemeyer, Bearden, & Sharma, 2003). This technique plots the number of classes on the x-axis and information criteria on the y-axis. As more classes are added, the marginal change in the information criteria is high but at some point diminishes. At this point, there will be an angle from near vertical to more horizontal slope. The model represented by the elbow indicates a candidate for the optimal model. This technique has not been tested methodologically through simulations studies to see if it would systematically identify the optimal number of classes.

Accuracy of classification. Each individual is classified by the model in the class with the highest posterior probability. We are interested in assessing whether these classifications are accurate, and entropy allows us to do this. The entropy maximization principle tells us that the model that provides the highest value for entropy best represents the most accurate depiction of the information contained in the observations (Akaike, 1977). Higher values of entropy represent more accurate assignment of subjects to classes, and values of .80 or higher are considered optimal. This can be interpreted as saying that individuals are classified accurately 80% of the time (Clark & Muthén, 2009).

Class enumeration. As described in the introduction, class enumeration represents the techniques needed to identify the number of classes in the mixture. Tofighi and Enders (2008)

and Nylund et al., (2007) conducted simulation studies to explore the performance of information criteria and likelihood ratio tests. Likelihood ratio tests perform similarly to a chi-square difference test in which the difference is compared to a chi-square statistic with the degrees of freedom equal to the difference in estimated parameters of the two models. When comparing nested models, it is possible to use likelihood ratio tests, although the test that assumes the difference is distributed as a chi-square does not work for latent class analyses. For GM models, there are two alternatives. The Lo, Mendell, and Rubin (LMR) test uses an approximate distribution, comparing the fit between the kth class and the k-1 class. If this difference is significant, then the kth class improves model fit. Similarly, the bootstrap likelihood ratio test (BLRT) uses a bootstrap sample to develop the distribution rather than use a deterministic distribution which is assumed for the likelihood ratio. Again, if the test is significant, this means that the addition of the extra class has improved model fit. Tofighi and Enders (2008) found the SABIC and LMR were most effective in identification of the correct number of classes. Nylund et al. (2007) found that the BLRT outperformed other techniques, although the BLRT adds substantial computing time. Nylund et al. also suggest taking a multi-step approach to identify the random seed used to generate the proper model first and then run the BLRT as a second step, an approach taken in the current study.

Kim (2015) investigated the performance of information criteria on class enumeration under conditions of parallel process and piecewise growth models. Varying conditions of missing values, class probability, class separation, number of indicators, and sample size, Kim found that with more indicators and large sample sizes, the BIC performed well in class enumeration. Specifically, even with 20% missing data and uneven class probabilities, when class separation is high, four indicators are used, and the sample size is 500, the BIC identified

the correct number of classes 100% of the time. The factor that caused the worst deterioration is low class separation. For the same conditions as above, the BIC only identified the correct number of classes 5% of the time.

Convergence. Liu and Hancock (2014) also conducted a simulation study to identify the efficacy of an unrestricted mixture model in class enumeration. One of their most important findings was that the rates of non-convergence and inadmissible solutions were much higher in GM modeling research than what has been reported in prior research. Many of research studies examined for the current study identified problems with model convergence (Chen, Kwok, Luo, & Willson, 2010; Enders & Tofighi, 2008, Tofighi & Enders, 2008). Liu and Hancock recommend non-convergent models and inadmissible solutions should be either considered evidence for fewer classes or not used at all.

Local solutions. Mixture models often are susceptible to local solutions from estimation procedures. Using an applied study and a simulation study, Hipp and Bauer (2006) found that increasing the complexity of GM models resulted in deterioration in the model estimation efficacy. In the applied study, they used both a LCG model and a GM model, and found that the proportion of models that converge and the proportion of converged solutions that produced the best log-likelihood decreased as the number of classes increased. They also found that the proportion of solutions that converge that also represent the best solution, as indicated by the highest log-likelihood, decreases as the number of classes increases. They recommend increasing the number of random starts to counteract this effect. They also suggest comparing solutions to determine whether they result in dramatically different substantive interpretations.

Use of covariates. Tofighi and Enders (2008) examined the performance of various fit indices and likelihood ratio tests in class enumeration, including the impact of the covariates.

They discovered that including covariates dramatically reduced class enumeration accuracy. Using the SABIC, which outperformed other fit indices in class enumeration, with a sample size of 400, the number of replications which identified the correct number of classes was reduced from 81% to 54% when covariates were added. Using a fixed sample size of 1000, the authors also compared models with low and high class separation and found class enumeration accuracy decreased from 88% to 53% for low separation when covariates were included. Prior research recommended the inclusion of covariates in class identification (Muthén, 2004), making this a surprising result.

Validation of latent class structure. Validation of the latent classes involves examination of information outside the outcome measures to test how well the latent classes align with substantive theory, covariates and other outcome measures. When examining other outcome measures and covariates, these are referred to by Muthén (2004) as auxiliary information. Auxiliary data can include antecedent measures, concurrent measures, and distal outcomes. Antecedent measures refer to measures of states or traits that may predict class membership. Concurrent measures can be considered similarly to their treatment in standard growth modeling, in that they can be used to develop validity evidence. Consequences, also known as distal outcomes, refer to measures that theory suggests vary according to the level of the construct delineated in the latent class. For example, the students in a class representing higher risk may be more likely to have other types of disciplinary problems in school.

Auxiliary variables can be incorporated into the model using the methodology outlined in Asparouhov and Muthén (2013). These methods include the pseudo-class method, the 3-step method, and the Lanza method (Lanza, Tan & Bray, 2013). Unfortunately, not all of these methods are available for use in a multilevel GM model in Mplus, but the pseudo-class method is

available for use in a multilevel model. As described in Asparouhov and Muthén, and Clark and Muthén (2009), the pseudo-class draw involves identification of the latent class in a first step. Using the posterior probability distribution of being assigned to a particular class, the latent class is assigned to subjects to create a number of sets of data referred to as pseudo-class draws using a process similar to multiple imputation. The pseudo-class draw method, using option auxiliary (e) is superior to using class membership or the posterior probability as an observed variable because it more accurately reflects the uncertainty and variability of class assignment by using the distribution of posterior class probabilities and repeated sampling to calculate the class specific means and variances. Mplus (Muthén & Muthén, 1998-2014) uses 20 pseudo-class draws.

Summary

A review of the literature on characteristics of student behavior in school suggests that although many students exhibit behavior conducive to success in the classroom throughout the year, there is still much that needs to be learned with regard to the students who struggle. Research developing student behavior typologies have provided many insights and several methods for classifying students, but none of these classifications reflect the characteristics of student behavior in the classroom. The DBR-SIS are uniquely designed to measure three constructs shown to support student classroom success: academic engagement, disruptive behavior and respectful behavior.

The use of growth mixture modeling allows us to capture the behavior trajectory heterogeneity of different groups of students using patterns of observed variables. Methodological and applied research discussed in this literature review using growth mixture modeling provides clear guidelines for developing and testing models by increasing the number

of classes and relaxing parameter constraints. No studies were identified that examined the effect of changing between class parameterization regarding invariance of intercepts, slopes and residual variances, and none were multilevel. In particular, no studies provided detailed descriptions of how trajectories, class proportions, and class characteristics varied as parameterization changed. Chapter 3 provides information on how growth mixture modeling was applied to the DBR-SIS behavior ratings.

Chapter 3: Research Methodology

The purpose of this study was to gain an understanding of how invariance assumptions in multilevel GM modeling can result in differences in latent class composition and growth curve characteristics. Using the context of student classroom behavior, the study disentangles how mean levels and variability in the intercepts, slopes, random effects, and residual variances within and between students and teachers affect latent class profiles. From a substantive perspective, this study seeks to develop a typology of student classroom behavior using differences in means and variation over one year using the DBR-SIS. The study utilizes data from a longitudinal validation study of the DBR-SIS in which students in grades 1, 2,4,5,7 and 8 were assessed on their classroom behavior by teachers twice daily for five days, during each of three data collection periods, fall, winter, and spring, totaling 30 data points. Grades 1-2 are considered lower elementary (LE), grade 4-5 upper elementary (UE), and grades 7-8 middle school (MS). During each rating period, ratings were structured such that five students were concurrently rated, using the DBR-SIS, administered twice daily for five consecutive days.

Participants

Three universities in three different states, Connecticut, Missouri and New York, referred to as sites, collaborated in the data collection for this research. Teachers were recruited from schools in these states. To be eligible for this study, the teacher had to provide consent and administrators of the schools had to provide approval. All students within the classrooms of participating teachers were eligible for participation, but only students whose parents refused participation were excluded from the pool of students available for the study. From the available

pool of students, up to 10 students per classroom were selected at random to participate. Participants in the study included 1975 public school students enrolled in 202 1st, 2nd, 4th, 5th, 7th, and 8th grade classrooms. The distribution of students and teachers are displayed in Tables 1 and 2. As identified in the fall, 52.2% of student participants were male. The racial identity of a majority of participants was White (82.5%), with 13.0% of the participants identified as African-American and 1.7% as Asian. The ethnicity of most participants was non-Hispanic (92.6%). Thirteen percent of students received special education supports as part of a formal special education identification.

Table 1
Student Demographic Distribution by Data Collection Point and Grade Group

Time Point and Grade Group	Male	White	Hispanic	Students with Disabilities
Fall (N=1945)				
LE (N=644)	52%	82%	9%	11%
UE (N=711)	50%	81%	5%	11%
MS (N=590)	54%	82%	9%	17%
Winter (N=1862)				
LE (N=627)	52%	83%	9%	11%
UE (N=693)	51%	82%	5%	11%
MS (N=542)	54%	84%	9%	16%
Spring (N=1822)				
LE (N=619)	52%	83%	9%	11%
UE (N=675)	51%	81%	5%	11%
MS (N=528)	54%	84%	9%	17%

Table 2
Teacher Demographic Distribution by Data Collection Period and Grade Group

Time Point and Grade Group	Male	White	Hispanic
Fall (N=201)			
LE (n = 66)	8%	97%	2%
UE (n = 73)	7%	99%	1%
MS (n = 62)	27%	97%	0%
Winter (N=196)			
LE (n = 65)	8%	97%	2%
UE (n = 72)	7%	99%	1%
MS (n = 59)	29%	97%	0%
Spring (N=193)			
LE (n = 65)	8%	97%	2%
UE (n = 70)	7%	99%	1%
MS (n = 58)	28%	97%	0%

Measures

DBR-SIS. The direct behavior rating format reflects the teacher's perception of the proportion of time a student is observed engaged in a target behavior from 0 (*never*) to 10 (*always*). The DBR-SIS focused on three target behaviors (e.g., three single item scales): academically engaged (AE), disruptive behavior (DB) and respectful behavior (RS). AE is defined as active or passive participation in the classroom activity, DB as behavior that results in distraction or interruption of classroom activity, and RS as behavior that is polite and agreeable towards the teacher and other students.

Although the DBR-SIS used in this study consists of three single item scales, classroom behavior is a construct represented by a combination of the three measures. Prior research has examined the psychometric properties of the DBR-SIS as a composite, combining the three scales into a single score by adding AE, RS, and the reverse score of DB (10 minus DB) (Johnson et al., in press). In order to assess whether this type of equal weighting aligned with the

factor structure, a one-factor exploratory factor analysis was run using AE, RS, and DB from all data collection periods. Pattern coefficients were in the range from .67 to .78 for all measures, providing evidence in support of the practice of equal weighting of the three scales. Therefore, for this study, a composite score was computed by taking the sum of AE, RS, and (10 minus DB), averaged over two ratings for each day resulting in a maximum of 15 ratings per student, five in each of the fall, winter, and spring.

The reliability of the DBR-SIS composite score was assessed using data from the current study reflecting the structure of the data with observations nested within students, and students nested within classrooms. Students were not crossed with teachers, such that student observations for each day were used to decompose the variance in the DBR-SIS into two parts: variance of student measures nested within teachers and error variance. Using Varcomp in SPSS (2009) 60.3% of the variance in the DBR-SIS is within students nested within teachers. Using the Spearman Brown prophecy formula, (Crocker & Algina, 2008), the DBR-SIS has adequate internal consistency reliability when averaging two measures per day, at .75.

BESS. The BESS is a norm-referenced, brief teacher rating scale useful in screening for behavioral and emotional strengths and weaknesses in children and adolescents. The BESS teacher form for children and adolescents consists of 27 items scored using combined sex norms to obtain a T-score. The BESS was found to have strong internal consistency reliability (split-half reliability of .97 and test-retest correlation of .91) (Kamphaus & Reynolds, 2007). The BESS is often considered the gold standard for child and adolescent behavior assessment, making it useful for several purposes in this study. It was used as a criterion measure of classification in terms of level of behavioral risk, with scores above 60 indicating elevated risk

level. It may also be used as a concurrent outcome measure, providing a source of evidence for validation of the latent class outcomes from GM models.

Descriptive statistics for the DBR-SIS and BESS are reported in Table 3. The DBR-SIS means describe a pattern in which students are on average 90% engaged with very little change across the year. Older student mean levels are higher than the younger students. Variability in the measures is highest during the fall data collection. The DBR-SIS are negatively skewed with positive kurtosis of sufficient magnitude to warrant concern over the lack of normality. The BESS T scores have means near 50 and standard deviations near 10, which aligns with reported norm groups (Kamphaus & Reynolds, 2007).

Table 3
Descriptive Statistics by Data Collection Point and Grade Group

	Fall (N=1945)						
	DBR-SIS					BESS	
Grade Group (N)	*M*	*SD*	*Median*	*Skewness*	*Kurtosis*	*M*	*SD*
LE (N=644)	26.16	4.45	28	-1.77	3.98	50.54	10.32
UE (N=711)	26.98	4.06	28	-2.14	5.37	49.31	10.47
MS (N=590)	27.04	4.31	29	-2.24	6.10	51.28	11.17
	Winter (N=1862)						
	DBR-SIS					BESS	
Grade Group (N)	*M*	*SD*	*Median*	*Skewness*	*Kurtosis*	*M*	*SD*
LE (N=627)	26.38	4.21	28	-1.65	3.24	50.69	10.49
UE (N=693)	27.39	3.43	29	-2.27	6.75	49.57	10.1
MS (N=542)	27.33	3.94	29	-2.36	7.09	51.63	11.31
	Spring (N=1822)						
	DBR-SIS					BESS	
Grade Group (N)	*M*	*SD*	*Median*	*Skewness*	*Kurtosis*	*M*	*SD*
LE (N=619)	26.99	3.67	28	-1.79	4.19	50	10.38
UE (N=675)	27.5	3.66	29	-2.27	8.86	49.85	10.32
MS (N=528)	27.34	4.05	29	-2.44	7.39	51.5	11.15

Demographic variables. Demographic information on student participants was obtained through the use of a separate student record review form using school records. Traditional demographic variables such as gender, race, ethnicity, age, and grade were collected and used in this study. Other demographic variables necessary for validation include ODRs, suspensions, expulsions, attendance data, disability status, and variables identifying the types of disability services received. These variables were not used as covariates to explain portions of the variability; rather, they were used to validate the latent classes, testing whether the latent classes predict these indicators as distal outcomes (Asparouhov & Muthén, 2013). Particularly useful to

this study are ODRs, suspensions, and expulsions because these are clear indicators that the student is having behavior problems in school. In addition, variables which indicate the type of disability and for students receiving behavioral supports, the type of support, provide sources for describing the features of the latent class outcomes. This is discussed further in the section on research question 1.

Growth Curve Shape. The graph of the average values for MS at the 5 data collection points in the fall, winter, and spring is shown in Figure 6. This shape was similarly found in UE and LE. Kooken (2014) identified that the DBR-SIS exhibits a piecewise linear growth pattern supporting the selection of a piecewise growth specification (see Figure 7). The model represents three parallel processes, which provides a comparable depiction of the data collection process in which the three collection periods were separated by on average 50 days. This allowed intercepts and slopes to vary and covary across the data collection periods. Initially, the plan was to estimate intercept and slope variances and covariances, but due to the limited amount of variation in the slopes, they were estimated as fixed effects. Although the DBR-SIS are skewed and kurtotic, with a large proportion of the measures at the endpoint of the scale, results from Kooken (2014) found the use of the maximum likelihood estimator with robust standard errors provided the best model fit. Using the skewed distribution options in Mplus did not produce admissible solutions due to a non-positive definite Psi matrix (Asparouhov & Muthén, 2014a).

Figure 6. Graph of Average DBR-SIS Composite MS Sample.

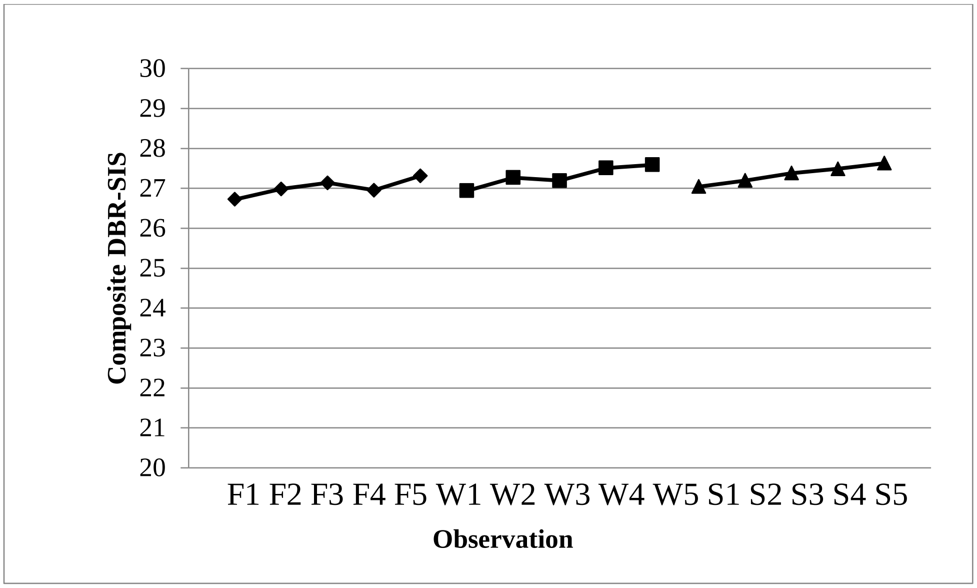


Figure 6. Shape of the average values for the 5 data collection points in the fall, winter, and spring for Middle School students.

Figure 7. Longitudinal Structural Equation Modeling Path Diagram for Piecewise Latent Growth Model.

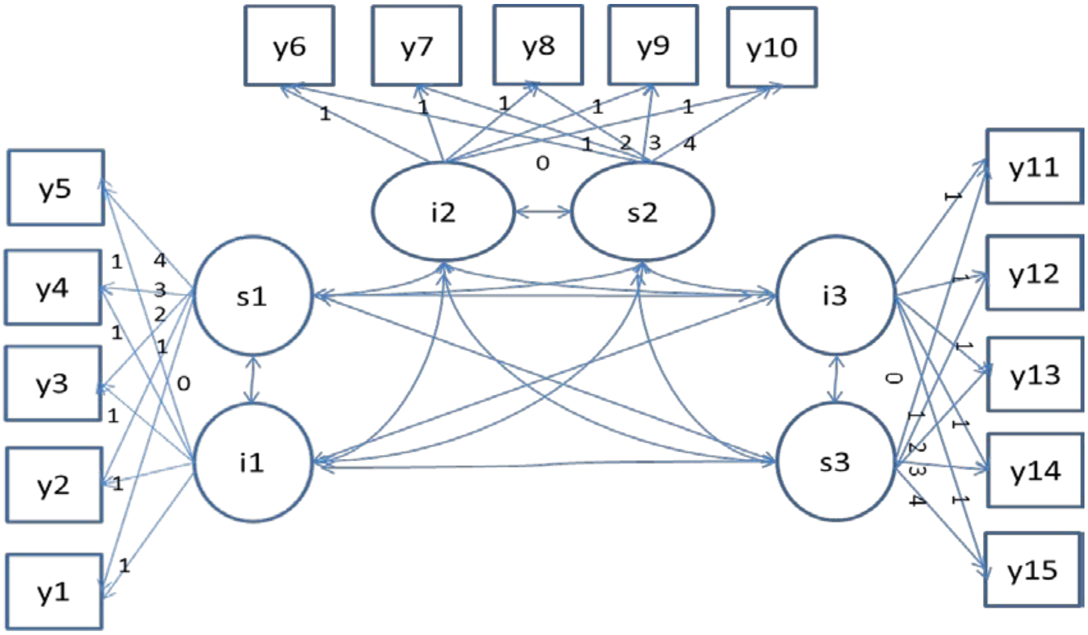


Figure 7. Piecewise growth model for the DBR-SIS composite using three non-contiguous data collection periods. Y1-Y5 were collected in the fall, Y6-Y10 were collected in the winter, and Y11-Y15 were collected in the spring.

Missing Data. Although it is often the case that an intensive longitudinal study may be plagued with problems associated with missing values, the level or pattern of missing data was not problematic in this study. Data were collected for each of the three measures twice daily for five days in the fall, winter, and spring, resulting in 30 measures per student. The data were checked thoroughly after each data collection period, and data anomalies and incidences of data that were missing were returned to the data collection sites for additional quality checks and follow up. A total of 1975 students participated in the study during at least one of the three data collection periods. Of the 1975 students, seven students did not participate in the fall, 93 did not participate in the winter, and 186 did not participate in the spring. On average for the entire year, 18% of the measures of AE, RS, and DB were missing (range 8% to 33%) with higher missing data rates during the spring data collection. Student attrition was due to families of students moving out of the district or the teacher's choice not to participate in the study in the later data collection periods. On individual days, missing values were due to student absences.

Using Mplus (Muthén & Muthén, 1998-2014), 823 missing data patterns were identified, including one pattern with 423 students having no missing values. Results were analyzed using SPSS Missing values analysis (Little, 1988). Results were significant indicating rejection of the null hypothesis of missing completely at random, $\chi 2$ (19,525, N = 1975) = 21,782, $p < .01$. Although not missing completely at random, the analysis of the causes for missing data presented above suggested data were missing at random, supporting the use of Mplus with the default Full Information Maximum Likelihood Estimator (FIML; Muthén & Muthén, 1998-2014). In addition, similar to studies reported in Enders (2010), the current study included a substantial effort put forth to go back and collect missing data or document reasons for its missingness. The conclusion supported by the protocol of this study was that the missing data in

this study could be assumed to be missing at random, and not causing the introduction of bias to the study. Research supports the use of FIML when data are missing completely at random or missing at random, with parameter estimates that are unbiased and efficient (Enders & Bandalos, 2001).

Analyses

Multilevel GM modeling was used to capture the behavior trajectory heterogeneity of different groups of students using patterns of observed variables at the student level, adjusting for correlated outcomes due to the teacher level. The procedures outlined in Ram and Grimm (2009), Jung & Wickrama (2008), Nylund, Asparouhov, and Muthén (2007), and using Mplus 7.3 (Muthén & Muthén, 1998-2014), provided guidelines in the development of an iterative process to test models with successively greater complexity from an unconditional multilevel model to a multilevel GM model.

Research Question 1. When selecting the multilevel growth GM model that best represents the characteristics of a real data sample of student classroom behavior using the DBR-SIS, in what ways and to what extent does noninvariance affect decision making, including application of enumeration indices, and resulting latent class characteristics?

To answer this question, a wide format database was used with one record per student which includes the repeated behavior measures, the clustering variable which represents the classroom, and demographic variables. A succession of increasingly complex mixture models were built and tested for substantive and methodological accuracy and model fit. The steps are outlined below and were followed for the composite DBR-SIS for each of the three grade groups resulting in three sets of models. The model specifications are reported in Tables 4 and 5; the Mplus syntax is reported in Appendix G.

Table 4

Final Models Specifications - Within Teacher

	Within LC			Between LC			Covariance
Model Number	Intercept Variance	Slope Variance	Residual Variance	Intercept Variance	Slope Variance	Residual Variance	I1, I2, I3
1	0	0	Invariant within F, W, S; non-invariant between F, W, S	0	0	Invariant	0
2	Estimated	0	Invariant within F, W, S; non-invariant between F, W, S	Invariant	0	Invariant	Estimated; invariant
3	Estimated	0	Invariant within F, W, S; non-invariant between F, W, S	Non-invariant	0	Invariant	Estimated; non-invariant
4	Estimated	0	Invariant within F, W, S; non-invariant between F, W, S	Invariant	0	Noninvariant	Estimated; invariant
5	Estimated	0	Invariant within F, W, S; non-invariant between F, W, S	Non-invariant	0	Varies	Estimated; non-invariant

Table 5

Final Models Specifications - Between Teacher

	Within LC		Between LC		Covariance
Model Number	Intercept Variance	Slope Variance	Intercept Variance	Slope Variance	I1, I2, I3
1	0	0	0	0	0
2	Estimated	0	Invariant	Invariant	Estimated
3	Estimated	0	Invariant	Invariant	Estimated
4	Estimated	0	Invariant	Invariant	Estimated
5	Estimated	0	Invariant	Invariant	Estimated

1. A series of LCG models, referred to as Model 1, with invariant classes were estimated, and the number of classes increased in steps to determine the model of best fit following the procedures outlined in Jung and Wickrama (2008), Asparouhov and Muthén (2013) using Mplus 7.3 (Muthén & Muthén, 1998-2014). To identify the best fitting model, information criteria, entropy, and two likelihood ratio tests were compared. When using the AIC, BIC, and SABIC, models with lower values were favored. Jung and Wickrama (2008) also recommend favoring the model with the lowest Lo, Mendell and Rubin likelihood ratio test (LMR) and the bootstrap likelihood ratio test (BLRT). When the BLRT and the LMR-LRT indicate divergent results, findings from Nylund, Asparouhov, and Muthén (2007) was applied which found that the BIC and BLRT outperformed the other fit indices and information criteria. In most cases the default number of starts in Mplus were used, which was 20 initial stage random sets of starting values and 4 final stage optimizations. However, if the model output contained evidence that the best log likelihood was not replicated, the number of starts were increased to 100 initial stage and 20 final stage optimizations (Muthén & Muthén, 1998-2014).

2. If models did not run properly, either due to non-convergence or inadmissible solutions, steps were taken to improve model specification, convergence, and replication of the best log-likelihood using recommendations from Hipp and Bauer (2006), Jung and Wickrama (2008), and Muthén and Muthén (1998-2014). In the current study, inadmissible solutions and non-convergent models were tracked, analyzed and reported. Revisions to model specification were considered only when model convergence presented a greater challenge, and these cases were identified.
3. After identifying the best number of latent classes in Model 1, the next steps involve a systematic approach to freely estimate variances, examining the impact on model fit and the number of latent classes. This technique builds upon a model building heuristic outlined in Ram and Grimm (2009). Based upon the preliminary analysis, there appeared to be variability in the intercepts and residual variances, but limited variability in the growth slopes.

   A series of models were built by expanding the parameterization to model in the variability in the population as follows:

   Model 2: Model 2 reflects the Mplus default settings with intercept random effects, covariances that are invariant across classes, and time varying but class invariant residual variances. Minor modifications were necessary to align the Mplus default settings with this study. As with all models, slope variances were constrained to 0 at both levels. Residual variances were constrained to be equal within the fall, winter, and spring data collection periods but to vary between the data collection periods.

Model 3: This model reflects intercept random effects that vary within and across classes and residuals that vary across fall, winter, and spring but are invariant across classes. Intercepts were allowed to covary.

Model 4: This model reflects intercept random effects that are invariant across classes and residuals that vary across fall, winter and spring and across classes. Intercepts were allowed to covary.

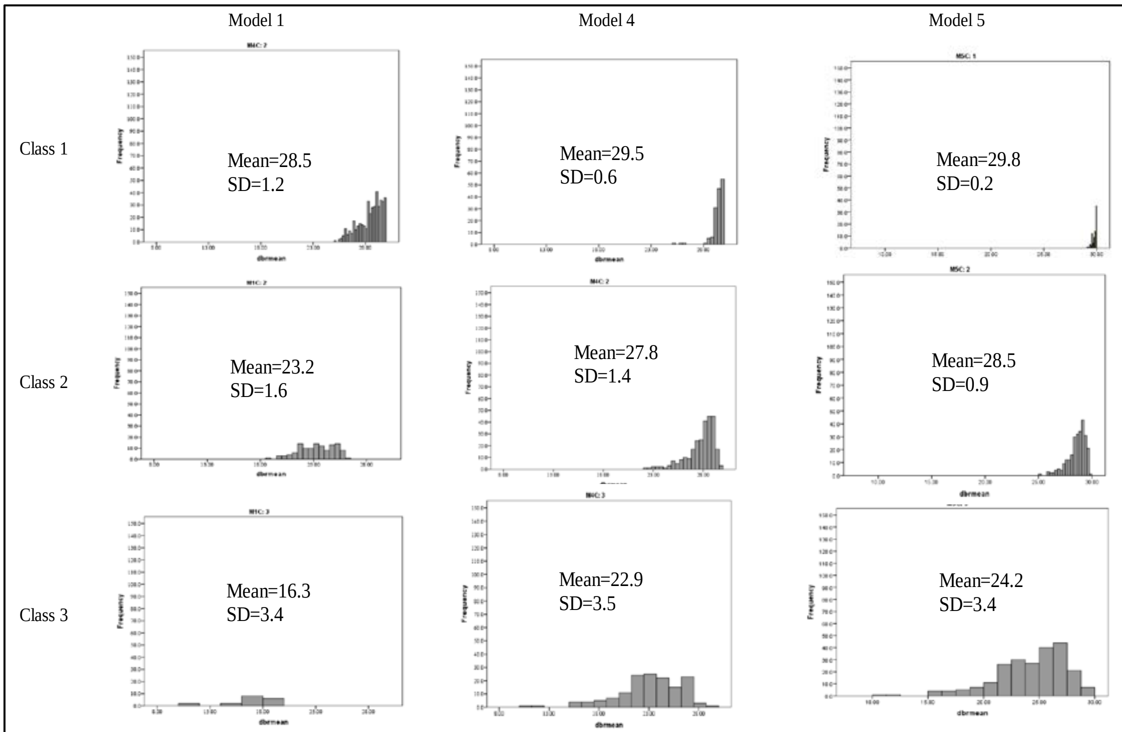
Model 5: This model reflects intercept random effects that vary across classes and residuals that vary by data collection point and across classes. Intercepts were allowed to covary.

For Models 2-5, random effects were estimated for the intercept at the between teacher level. These reflect the portion of the variability in the DBR-SIS mean is attributable to the teacher. The random effects were invariant between classes. Slopes were estimated as fixed effects. Intercepts were allowed to covary.

4. A comparison of latent trajectory class assignment, class characteristics, shape of the trajectory, and model fit criteria is provided between the latent trajectory classes in Models 1 and 5 and Models 4 and 5. An important part of this analysis involves tracking how students changed from one class to another as the parameterization changed. Tables and histograms are provided to trace how students are reassigned to different classes as model specifications change. The distribution characteristics of student level means and standard deviations of observed DBR-SIS scores provide insight into the distribution of classroom behavior of students in each latent class. The student level distribution is developed by first calculating the mean of the DBR-SIS for each student across all observations. The student level means and standard deviations are then analyzed by

displaying their frequency distribution through a histogram, with overall means and standard deviations. Figure 8 provides an example of these histograms.

Figure 8. Comparison of DBR-SIS Student Means by Latent class



	Model 1	Model 4	Model 5
Class Proportions			
Class 1	76%	29%	17%
Class 2	21%	45%	44%
Class 3	3%	26%	39%
Skewness			
Class 1	-0.75	-3.89	-1.16
Class 2	-0.36	-1.44	-1.21
Class 3	-1.65	-1.45	-1.52
Kurtosis			
Class 1	-0.34	21.29	1.06
Class 2	-0.55	2.11	1.53
Class 3	-3.00	3.68	4.00
BESS T scores (SD)			
Class 1	48.1 (8.3)	43.1 (6.0)	40.3 (4.0)
Class 2	62.1 (8.4)	51.8 (8.5)	49.2 (7.6)
Class 3	71.6 (9.3)	61.8 (10.0)	59.6 (9.9)

Figure 8. Example of histogram displaying the distribution of student means of the DBR-SIS by latent class, Model 1 versus Model 5.

5. The observed versus predicted trajectories were compared graphically as depicted in Figure 9. This technique is comparable to that suggested in Asparouhov & Muthén (2014b). Plots of the observed values versus predicted values allow us to visually examine the amount of variability in the observed values around the predicted trajectory. These plots provide insights in comparing how the different models depicted heterogeneity.

The results from Research Question 1 provide a rather voluminous amount of information comparing competing models that differ in their invariance assumptions. By looking at how class enumeration, class assignment, distribution of criterion measures and distribution of the DBR-SIS change as a function of the different models, we can identify the impact of changing invariance assumptions, The selection of a best model must be guided by model fit, but it must also be influenced by substantive theory (Muthén, 2004).

Figure 9. Comparison of Predicted Growth Trajectories versus Sample Values by Model and Class

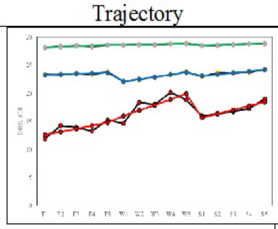
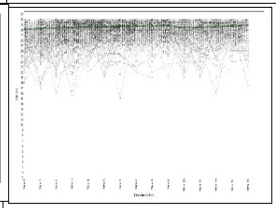
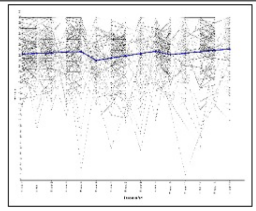
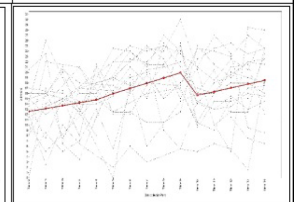
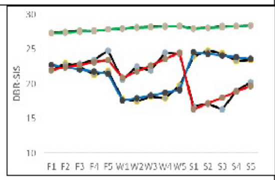
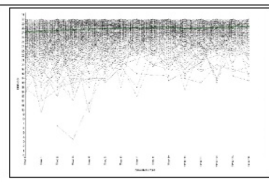
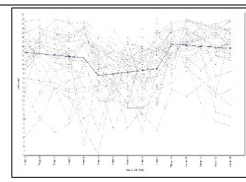
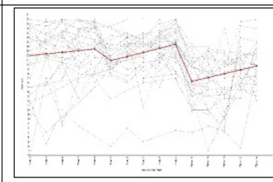
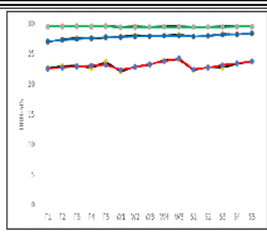
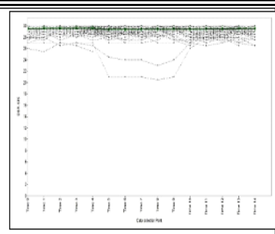
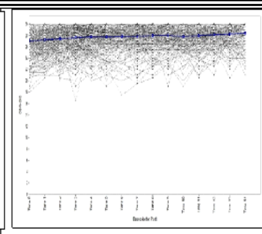
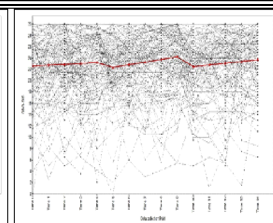
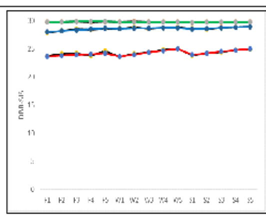
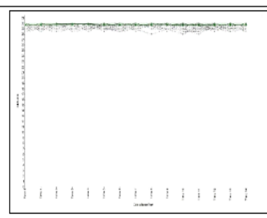
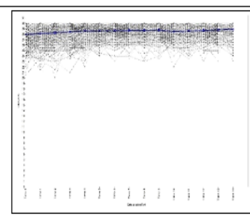
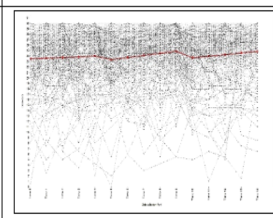
Model	Average Trajectory	Class 1	Class 2	Class 3
1				
2				
4				
5				

Figure 9. This sample graph shows variability of sample values around latent class mean trajectory for three-class GM model solution for middle school students.

Research Question 2. To what extent does the final trajectory class membership selected from research question 1 predict distal outcomes of student behavior using other measures such as using the BESS risk classification, ODRs, and suspensions, and provide validity evidence for the latent trajectory classes?

The purpose of this question is to investigate the validity of the latent classes in predicting other behavior outcome measures. The latent classes represent different levels on a continuum of behavior from adaptive to maladaptive. The goal of this question is to identify whether the assignment of students to classes predicts the level of disciplinary actions and level of BESS scores. Finding that students assigned to a maladaptive class had more disciplinary actions or higher BESS scores would provide evidence for the validity of the class assignment. Similarly, if students assigned to an adaptive class had fewer disciplinary actions and lower BESS scores, this would provide evidence for the validity of the assignment to the adaptive class.

Muthén (2004) emphasized the importance of using a variety of information sources in the form of covariates and outcomes to provide evidence for the validity of the latent classes. This suggestion is made with the caveat that covariates differ in their substantive position relative to the latent trajectory classification, and as a result, this affects the choice of how to position the covariate in the model. Covariates can be used as predictors of intercepts and slopes of the growth model. They can also be used to predict class membership. They can also be used to test the model by using the latent trajectory class as a predictor of the covariate as an outcome.

In this study, the GM modeling process classifies students into different latent trajectory classes based upon the repeated outcome measures using the DBR-SIS. Although there are other measures of student behavior (e.g., BESS, SSIS, and ODRs) neither theory nor prior analyses

provide a clear picture of whether these measures are useful for distinguishing the latent trajectory classes of interest. This clarity of theory or analysis is necessary to support the use of these as antecedent covariates. Therefore, the effect of these other measures was explored by testing them as distal outcomes of the latent trajectory class.

When using Mplus 7.3 (Muthén & Muthén, 1998-2014) distal outcomes can be incorporated into the model as auxiliary variable using the methodology outlined in Asparouhov and Muthén (2013). The auxiliary variables that were tested using this approach include risk status using the BESS, ODR, suspension and expulsion. Given that prior research found that males, minorities and students with disabilities are often disproportionately high in more severe behavior classes, these variables were also examined using the auxiliary (e) option. Given that this study contained many additional demographic variables, history of grade retention and absences for the year were also examined. The expectation is that students who are considered higher risk, with higher BESS scores, more than 2 ODRs, suspensions or expulsions will be assigned at a disproportionately higher rate to the mal-adaptive classes than the normative classes. In addition, based upon prior research, there is an expectation that males, minorities, and disabled students may be overrepresented in maladaptive classes.

Research Question 3: What are the characteristics of the students comprising latent classes of classroom behavior trajectories that emerge from the use of GM modeling, and in what ways do the latent classes vary in intercepts, slopes and residual variances?

The results from the analyses conducted to answer research questions 1 and 2 was fully examined in this step. Descriptive statistics provide the composition of each latent class by gender, race, ethnicity, disability status, and level of behavioral support. In addition, information on parameter estimates and a plot of the average trajectory is provided and discussed with

emphasis on the substantive interpretation. Variation in intercepts, slopes, and residual variances were connected with substantive theory and practice. Specifically, recommendations and insights aim to connect these results with single case design research, especially using the DBR-SIS.

Using the descriptive information and shape of the trajectory, the classes are described in terms of their typology of student behavior. These trajectory classes were compared to those found in the behavior typology literature. Latent trajectory classes were described by their level of risk for behavioral problems in school and compared to findings from other literature (Dowdy et al., 2014; Dever, Dowdy, Raines, & Carnazzo, 2015).

Summary

This chapter provided a description of the sample, the measures, and the statistical techniques used in this study of the DBR-SIS using growth mixture modeling. Chapter 4 presents the results of these analyses.

Chapter 4: Results

Each of three research questions were completed for each of the three samples, MS, UE, and LE. The research questions investigated in this study were as follows:

1. When selecting the multilevel GM model that best represents the characteristics of a real data sample of student classroom behavior using the DBR-SIS, in what ways and to what extent does noninvariance affect decision making, including application of enumeration indices, and resulting latent class characteristics?
2. To what extent does the trajectory class membership selected from research question 1 predict distal outcomes of student behavior measured using alternatives measures such the Behavior Assessment System for Children-2 Behavioral and Emotional Screen System (BESS; Kamphaus & Reynolds, 2007) risk classification, office disciplinary referrals (ODRs), and suspensions, providing validity evidence for the latent trajectory classes?
3. What are the characteristics of the students comprising latent classes of classroom behavior trajectories that emerge from the use of GM models, and in what ways do the latent classes vary in intercepts, slope, and variances?

Introduction

The parameterization for all models within teacher and between teachers is reported in Chapter 3, Tables 4 and 5. Full invariance except for intercepts and slopes is assumed in Model 1 and progressively relaxed for Models 2-5. For all models, the residual variance was constrained to be equal within each data collection period and vary between data collection periods. In other words, the residual variance for fall was constrained to be equal for all five time points. The residual variance for the winter and spring were also constrained to be equal across all five time

points, but fall was allowed to vary from winter, winter from spring and fall from spring. Initially, the parameterization followed the Mplus default which allows the residual variance to vary over time requiring 15 parameters, but this parameterization resulted in convergence problems. Using the estimated parameters, a series of simulations were run to test whether the parameterization was the modeling problem or the low cluster versus high parameter numbers, and it was confirmed that the models contained too many parameters for the cluster sizes (e.g. simulations with cluster sizes greater than the number of parameters properly converged). Therefore, the decision was made to restrict the residual variances to invariant within data collection periods but noninvariant between data collection periods.

For purposes of consistency, Class 1 is always used to represent the class with students who have the highest scores. Class 2 and higher represent students with progressively lower scores and diminishing levels of school behavior that is conducive to learning.

Middle School

Research Question 1. When selecting the multilevel GM model that best represents the characteristics of a real data sample of student classroom behavior using the DBR-SIS, in what ways and to what extent does noninvariance affect decision making, including application of enumeration indices, and resulting latent class characteristics?

Model fit results for Models 1-5 for MS are discussed in order and reported in Appendix A, Table A1. Beginning with Model 1 using AIC, BIC, SABIC, and entropy, the 3-class model had the best fit. Although fit indices indicated preference for the 4-class model over the 3-class model, the BLRT p-value for the 4-class model indicated that the fourth class did not improve model fit over the 3-class model. The 3-class model was selected as the best fitting LCG model. Students were distributed among the classes with 76% in Class 1, 21% in Class 2, and 3% in the

Class 3. Class 1 has initial status of 28.5, and increases at a very small rate over the year, ending at 29.0. Class 2 has initial status of 23.3 increasing at a very small rate across the year, but it drops at each node representing the change from the fall to the winter and winter to the spring data collections. Class 2 ended at 24.2 representing only a small amount of change over the year. This pattern of a shift in the level from the end of one data collection period to the beginning of another, referred to as a node, was seen in many of the models. Class 3 had initial status of 12.6 and increasing steadily in the fall and winter with positive slopes. The model reflected a drop from the winter to spring node, and the year ends at 18.5. The ending DBR-SIS predicted score of 18.5 can be interpreted as meaning that on average, students in this class exhibit the targeted behavior 62% of the time.

Models 2-5 represent a series of increasingly complex models, and in each case the 3-class model represented the best fitting model. Model 2 represented the simplest GM model and the Mplus default model, with class invariant intercept random effects, slope fixed effects, and class invariant residual variances. The BLRT for the 3-class model indicated an improvement in fit over the 2-class model ($p < .001$), and the 4-class model did not converge due to a saddle point. The distribution by class was 87% for Class 1, 7% for Class 2, and 6% for Class 3. For the Model 2 3-class model, Class 1 has initial status of 27.4, and increases at a very small rate over the year, ending at 28.5. Class 2 has initial status of 22.7 and increases across the year. Class 2 exhibits a very large drop in the node from fall to winter and a large increase in the node from winter to spring, ending the year at 23.6. Class 3 has initial status of 22.0 increasing at a very small rate across the year. This class drops abruptly at each node representing the change from the fall to the winter and winter to the spring data collections ending at 19.7.

Model 3 represented class varying intercept random effects, but all other specifications remained the same as for Model 2. Surprisingly, Model 3 was problematic in that the solution had a non-positive definite latent variable covariance matrix. Model 4 represented class invariant intercept random effects and class varying residual variances. This model provided improved model fit for the 3-class model, but the 4-class model was not estimated due to a non-invertible matrix in one class. The distribution of the 3-class Model 4 by class changed substantially from Models 1 and 2 to 29% for Class 1, 45% for Class 2, and 26% for Class 3. Model 4 Class 1 had initial status of 29.5 and no change throughout the year. Model 4 Class 2 had initial status of 27.1 with a very small increase throughout the year, ending at 28.4. Model 4 Class 3 had initial status of 22.6, again with only a small change throughout the year, ending at 23.7.

Model 5 with three classes represented the model of best fit using the AIC, BIC, SABIC, Entropy, and BLRT. The 4-class model was not estimated due to a non-invertible matrix in one class. The 3-class model had statistically significantly better fit than the 2-class model as indicated by the BLRT ($p < .001$). The distribution by class was 17% for Class 1, 46% for the Class 2, and 39% for Class 3. Model 5 Class 1 starts very high at 29.8 and stays at that level throughout the year, with statistically non-significant slope estimates. Model 5 Class 2 starts at 28.0 increasing only slightly to 28.9 by the end of the year. Model 5 class 3 starts at 23.7 and increases slightly during the fall, drops at the beginning of the winter followed by a slight increase, and then drops again for the spring followed by a slight increase, ending at 25.1. The intercepts and slopes and corresponding trajectories did not vary much between Models 4 and 5, but the proportion of students assigned to each class did.

***Latent trajectory comparisons among competing models*.** For all MS models, the best fitting model was always a 3-class model. The model trajectories are reported in Figure 10. For

each model, Class 1 was characterized by near perfect scores, close to 30 with very little change across the year. For Class 2, the scores were lower, in the range of 25 to 30 for most models, with a small positive slope across the year. Class 3 scores were low, in the 20 to 25 range for most models. Class 3 experienced greater improvement over the year as indicated by a larger positive slope, but Class 3 also experienced sizable changes in the intercepts in the nodes between fall and winter and winter and spring.

Figure 10. Comparison of MS Model Estimated Trajectory to Student Sample Trajectories by Model and Class

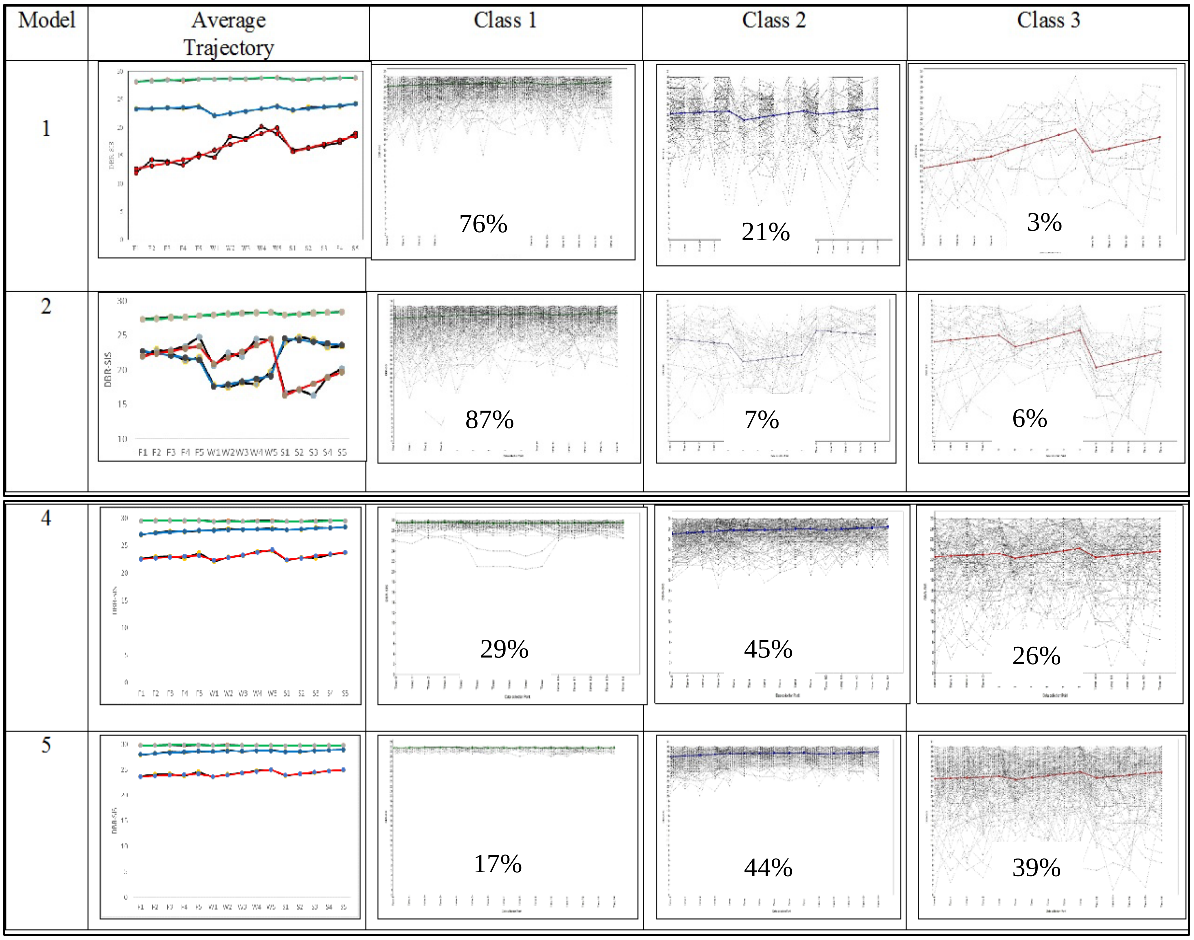


Figure 10. Comparison of MS model estimated trajectories for Models 1, 2, 4, and 5 and sample data for each of classes 1, 2, and 3. Percentages represent proportion of students assigned to each class. For each graph, the x axis represents the fifteen data collection points and the y axis represents the DBR-SIS score from 0 to 30.

The growth curve parameters for Model 5 are provided in Appendix B, Table B1. The growth factor variances were statistically significant for the Classes 2 and 3, but not for Class 1 indicating greater variability in the scores. In particular the magnitude of the variance for the intercepts for Class 3 was substantially higher than the other classes. For example, the variance of the intercept for the fall was not statistically significant for Class 1, but was 1.00 ($p < .01$) for Class 2 and 16.52 ($p < .01$) for Class 3. In addition, residual variances were statistically significant for all time periods, and once again substantially larger in magnitude for Class 3 than

the other classes. Covariances for Classes 2 and 3 were statistically significant at the within teacher level. Neither variances nor covariances of the intercepts were statistically significant at the between teacher level.

All models identified a large amount of variability in the intercepts and the residuals and across the classes. To better understand this, comparisons of the average trajectories for each class and model to the individual sample values are displayed in Figure 10. Given that Model 1 represents a set of class trajectories assuming that all heterogeneity is reflected in the mean intercepts, slopes and residual variances, the amount of variability around the average trajectories is large for all three classes and represented completely in the model by the class invariant residual variances. Model 1 Class 1 is very large, representing 76% of the students. Model 1 Class 3 is quite small with only 19 students (3%) characterized by a lower average DBR-SIS and a greater rate of increase across the year. For Model 2, Class 1 is also very large, and the trajectories for Classes 2 and 3 reflect fluctuations in the intercepts and slopes that are difficult to interpret. The trajectories in Models 4 and 5 have smaller slopes indicating very little change across the year, but a large difference in the level of variability around the average trajectories. For Class 1, there is very little variability, for Class 2, there is more and for Class 3, the variability is quite large. Models 4 and 5 also reflect a greater assignment of students to Class 3. It appears that Models 1 and 2 represent a classification system based upon mean levels, as reflected in the intercept and slope estimates, and Models 4 and 5 represents a classification system based upon both mean levels and variability.

Now that three candidate models have been identified, the next steps involve comparing these models on proportions of assigned students, demographic composition, model shape and descriptive statistics to identify the preferred model. Even though Model 2 represents the Mplus

default, due to the unusual fluctuations in the Model 2 trajectories, the focus of the remainder of this study for MS compares Models 1, 4 and 5. For comparative purposes, Model 2 tables have been provided in Appendix C.

Comparisons of change in class among competing models. In addition to differences in parameter estimates and predicted trajectories, variations among the five models also reflect a shift in the assignment of students to each class. Characteristics of the shifting of students provided insights into the nature of these models' resultant mixtures. A comparison of the change in latent class assignment from Model 1 to Model 5 and Model 4 to Model 5 are shown Table 6 and 7. For example, Table 6 displays a matrix identifying the number of students who either changed class or remained in the same class when comparing Models 1 and 5. Table 8 provides demographic characteristics of students that change class between Models 1 and 5 in each of the 9 cells, although in Table 8, there are only students in 6 of the 9 cells.

Out of 451 students assigned to Class 1 in Model 1, 22% remained in Class 1, 57% were assigned to Class 2 and 21% were assigned to Class 3 in Model 5. Of the 121 students who were in Class 2 in Model 1, only 2% remained in Class 2 and 98% were assigned to Class 3 in Model 5. The 19 students assigned to Class 3 in Model 1 remained there in Model 5.

Considering also the shift from Model 4 to Model 5, of the 176 students in Class 1 in Model 4, 57% remained in Class 1, 43% shifted to Class 2, and less than 1% shifted to Class 3. Of the 268 students in Model 4 Class 2, 70% remained in Class 2 and 30% shifted to Class 3. All of the 147 students in Model 4 Class 3 remained in Class 3 for Model 5.

Table 6.
Comparison of Latent Class Assignment Model 1 versus Model 5, MS

		Model 5 Latent Class				
		1	2	3	Model 1 Sum	Model 1 Proportion
Model 1 Latent Class	1	100	259	92	451	76%
	2	0	3	118	121	21%
	3	0	0	19	19	3%
	Model 5 Sum	100	262	229	591	
	Model 5 Proportion	17%	44%	39%		100%

Table 7.
Comparison of Latent Class Assignment Model 4 versus Model 5, MS

		Model 5 Latent Class				
		1	2	3	Model 4 Sum	Model 4 Proportion
Model 4 Latent Class	1	100	75	1	176	30%
	2	0	187	81	268	45%
	3	0	0	147	147	25%
	Model 5 Sum	100	262	229	591	
	Model 5 Proportion	17%	44%	39%		100%

Demographic comparison of change in class among competing models. It is also informative to understand the demographic composition of the students shifting class assignment from Model 1 to Model 5 and Model 4 to Model 5 as shown in Tables 8 and 9. Conventional demographic categories including race and gender were considered. In addition, the proportion of students who were retained at least one grade, who were special education, who had emotional or behavior disability (EBD), received behavior supports, received at least one expulsion or suspension, were rated as basic risk on the BESS, were rated as elevated risk on the BESS, and received 2 or more ODRs were compared. Comparing Model 1 to Model 5, students shifting from Class 1 to Class 2 are disproportionately white, female, special education, and have fewer ODRs, suspensions and expulsions. Students shifting from Class 1 to Class 3 are disproportionately female, black, special education, and Hispanic. They are also more likely to receive behavior supports, to be at risk based upon the BESS, and to be at risk based upon the number of ODRs. Students shifting from Class 2 to Class 3 were disproportionately higher on all the identified demographic risk factors.

Table 8
Demographic distribution by change group—MS—Model 1 versus Model 5

Category	All MS	Always Class 1	Always Class 2	Always Class 3	Class 1,2	Class 1,3	Class 2,3
Male	.54	.32	.67	.89	.51	.52	.75
White	.82	.95	1.00	.79	.85	.79	.65
Black	.13	.01	.00	.16	.08	.16	.31
Retained	.06	.01	.33	.00	.06	.07	.10
Special Education	.18	.03	.00	.21	.19	.22	.23
EBD	.02	.00	.00	.00	.01	.01	.05
Behavior	.04	.01	.00	.26	.02	.05	.07
Days Absent	7.62	5.32	7.50	8.47	7.50	5.67	11.72
Suspension or Expulsion	.09	.01	.00	.16	.07	.07	.20
Brisk	.19	.00	.00	.21	.14	.21	.43
Erisk	.11	.00	.00	.74	.02	.08	.34
Hispanic	.09	.06	.00	.11	.07	.11	.14
ODRRISK	.20	.06	.00	.68	.12	.20	.42
DBR mean	27.07	29.80	25.60	16.50	28.60	27.10	23.17
DBR SD	3.22	.20	.45	3.38	.79	.97	1.62
BESS T score mean	51.70	40.30	53.10	71.60	49.10	53.60	62.30
BESS T score SD	10.71	3.99	.84	9.33	7.64	7.77	8.42
n	591	100	3	19	259	92	118

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS,
ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 1 and Class b in Model 5.
Only nonempty change classes are reported.

Table 9.
Demographic distribution by change group—MS—Model 4 versus Model 5

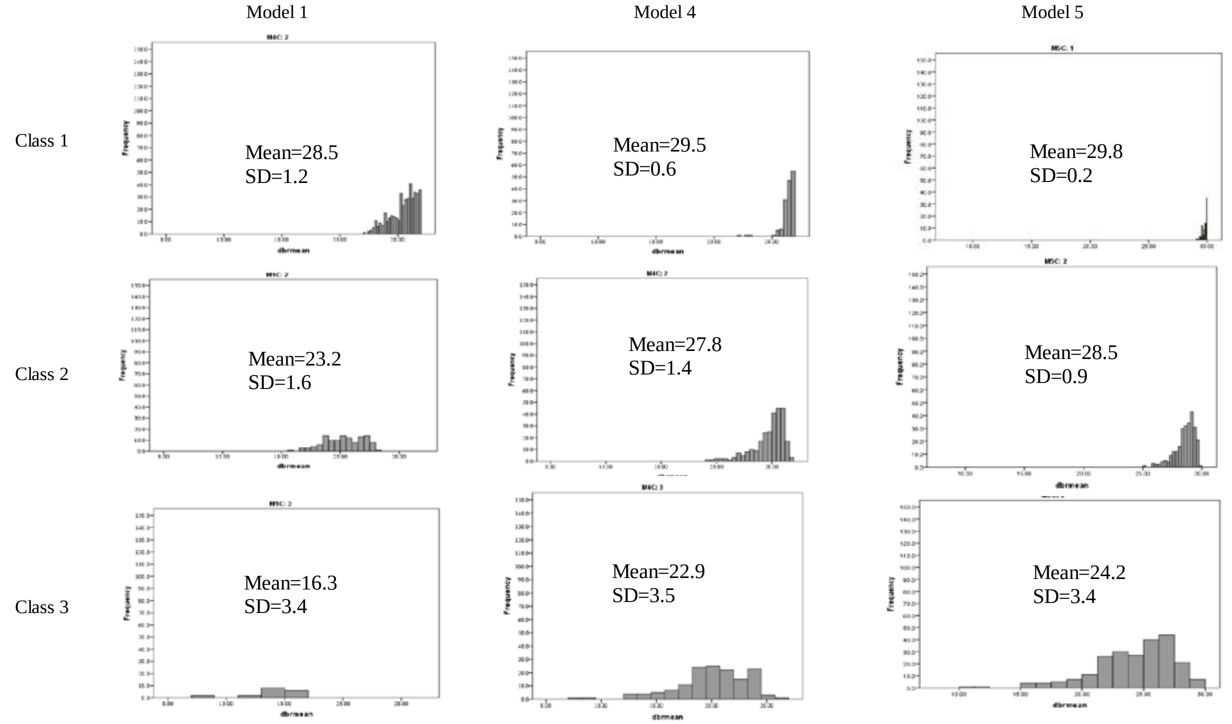
Category	All MS	Always Class 1	Always Class 2	Always Class 3	Class 1,2	Class 1,3	Class 2,3
Male	.54	.32	.55	.72	.43	1.00	.57
White	.82	.95	.86	.71	.85	1.00	.74
Black	.13	.01	.09	.24	.05	.00	.23
Retained	.06	.01	.07	.08	.03	.00	.07
Special Education	.18	.03	.21	.21	.13	.00	.25
EBD	.02	.00	.01	.03	.01	.00	.04
Behavior	.04	.01	.02	.08	.01	.00	.07
Days Absent	7.62	5.32	7.86	10.20	6.58	.00	6.96
Suspension or Expulsion	.09	.01	.07	.18	.04	.00	.09
Brisk	.19	.00	.18	.35	.04	1.00	.26
Erisk	.11	.00	.03	.34	.00	.00	.14
Hispanic	.09	.06	.06	.14	.11	.00	.10
ODRRISK	.20	.06	.13	.41	.09	1.00	.23
DBR Mean	27.07	29.81	28.31	22.91	29.14	25.10	26.49
DBR SD	3.22	.20	.82	3.49	.62	-	1.69
BESS T-score Mean	51.70	40.30	50.21	61.77	46.61	60.33	55.59
BESS T score SD	10.71	3.99	7.94	9.95	6.03	-	8.51
n	591	100	187	147	75	1	81

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS,
ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 4 and Class b in Model 5.
Only nonempty change classes are reported.

Comparing Models 4 and 5 as shown in Table 9, students shifting from Class 1 to Class 2 were disproportionately female, white, and in general less representative of demographic characteristics not associated with behavioral risk. The 81 students that shifted from Class 2 to Class 3 were disproportionately male, black, assigned to special education, had higher BESS risk status, and higher ODR risk.

Comparisons of descriptive statistics among competing models. To understand more about the characteristics of the students by Class in Model 1 versus Model 5 and Model 4 versus Model 5, it is also useful to look at characteristics of within student means and standard deviations by model. The within student mean represents the average DBR-SIS across all observations. The within student standard deviation represents the standard deviation of the DBR-SIS across all observations for a single student. The distribution of the DBR-SIS within student means and within student standard deviations across students are shown in Figures 11 and 12. Also provided are the mean and standard deviations across all students. To provide some context, a comparison of means provides us with information on the average behavior of the group of students assigned to that class. The higher the mean, the better the student behavior. A comparison of the standard deviations tells us how similar or different the students are in the class. When the mean of the student DBR-SIS standard deviations for a class is low, this implies the students in that class are very similar in their level of variability. Higher levels imply students are less similar to one another in the variability of the DBR-SIS measures across time. Also, the means and standard deviation histograms for AE, RS, and DB separately are provided in Appendix E.

Figure 11. MS-DBR-SIS student level means by model and class



	Model 1	Model 4	Model 5
Class Proportions			
Class 1	76%	29%	17%
Class 2	21%	45%	44%
Class 3	3%	26%	39%
Skewness			
Class 1	-0.75	-3.89	-1.16
Class 2	-0.36	-1.44	-1.21
Class 3	-1.65	-1.45	-1.52
Kurtosis			
Class 1	-0.34	21.29	1.06
Class 2	-0.55	2.11	1.53
Class 3	-3.00	3.68	4.00
BESS T scores (SD)			
Class 1	48.1 (8.3)	43.1 (6.0)	40.3 (4.0)
Class 2	62.1 (8.4)	51.8 (8.5)	49.2 (7.6)
Class 3	71.6 (9.3)	61.8 (10.0)	59.6 (9.9)

Figure 11. Histogram displays the distribution of the within student mean DBR-SIS score by model and class for MS. The proportion of students assigned to each class, skewness, kurtosis and average BESS T scores are displayed at the bottom.

Figure 12. MS-DBR-SIS student level standard deviations by model and class

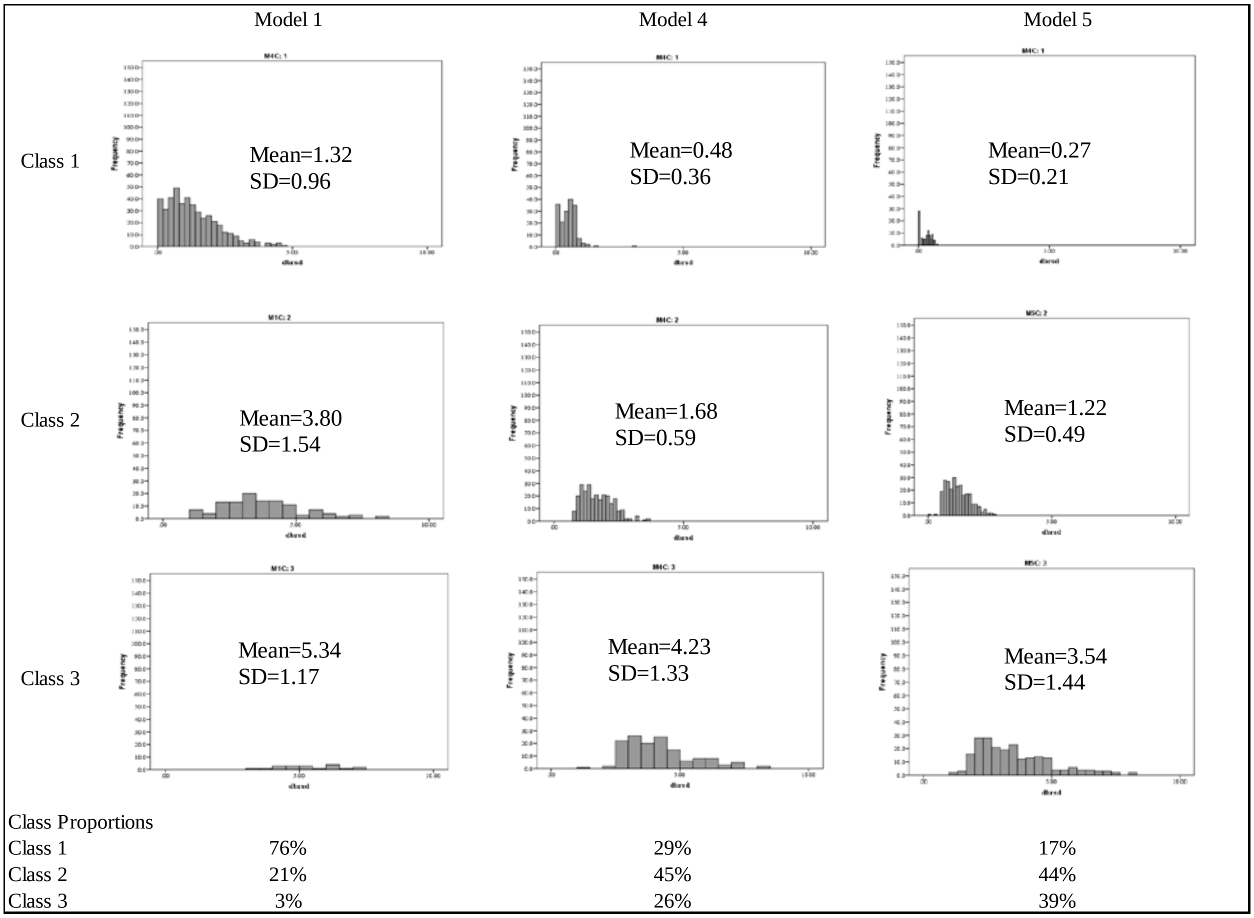


Figure 12. Histogram displays the distribution of the within student DBR-SIS standard deviation by model and class for MS. The proportion of students assigned to each class is displayed at the bottom.

Starting with Class 1, the means for Class 1 are very similar between Model 1 at 28.5, Model 4 and 29.5, and Model 5 at 29.8; however, the amount of variability in the means and in the standard deviations is different. As can be seen in the graph or the values of the standard deviations, the variability within Model 1 Class 1 is much higher (M = 1.32) than either Model 4 (M = 0.48) or Model 5 Class 1 (M = 0.27). Similarly with Class 2, Model 1 reflects greater variability than Model 4 and Model 5. However, in this case, the mean values are also lowest in Model 1 (M = 23.2). The standard deviations for Model 1 Class 2 indicate greater variation in the

means for Model 1 versus Models 4 and 5 (Model 1 Class 2 M=3.80; Model 2 Class 2 M = 1.68, Model 5 Class 2 M = 1.22. Finally, the most dramatic difference can be seen when comparing Class 3 in Model 1 versus Model 5. The mean of Model 1 is much lower than Model 5 (M = 16.3 versus M = 24.2), but the standard deviations are much higher (M = 5.34 versus M = 3.54).

As mentioned earlier, Models 4 and 5 are very similar in the shape of the growth trajectories. Model 4 reflects class differences in residual variances, and Model 5 reflects class differences in both intercept random effects and residual variances, resulting in the reassignment of students from Classes 1 and 2 to Class 2 and Class 3 in Model 5. Eighty-one students shifted from Class 2 in Model 4 to Class 3 in Model 5. These students have a mean DBR-SIS score of 26.5 and standard deviation of 2.30, placing them in the higher, more variable end in Model 4 Class 2. A few students had mean DBR-SIS above 28.0 representing a pattern of nearly all values in the range of 28-30, but with one or two scores of 24 or below. In summary, the key difference between Model 4 and Model 5 is that Model 5 assigns students with lower means and higher variability to higher classes than Model 4, resulting in a set of 3 classes that are more homogeneous in their response patterns.

The skewness and kurtosis can also tell us a lot about the relative performance of the three models. As stated in earlier chapters, GM modeling is based upon the assumption that the full distribution is not normally distributed but that it is comprised of a mixture of normally distributed latent subgroups. The classes in Model 1 are slightly negatively skewed and leptokurtic. Only Class 3 would be considered of concern in its skewness (-1.65) or kurtosis (-3.00). Model 4 on the other hand resulted in classes that are more negatively skewed with positive kurtosis. Model 5 is an improvement over Model 4 with smaller negative skewness and smaller positive kurtosis.

These comparisons provide a picture of the differences between Model 1, 4, and 5. Model 1 represents a classification with a very large class of well-behaved students and a small average and poorly behaved classes of students. Model 1 sorts the sample based more on the mean levels. If the purpose of this classification was to identify students with severe problems with great precision, Model 1 has some merits. Models 4 and 5 seem to sort the sample based upon not just means but also based upon differences in the variability over time. The class proportions are more evenly distributed for both Models 4 and 5. Models 4 and 5 are also very similar in their means and standard deviations; however, consideration of the skewness and kurtosis gives preference to Model 5.

Research Question 2. To what extent does the trajectory class membership selected from research question 1 predict distal outcomes of student behavior measured using alternatives measures such the BESS risk classification, office disciplinary referrals (ODRs), and suspensions, providing validity evidence for the latent trajectory classes?

Using the auxiliary (e) option in Mplus, the latent trajectory classes were analyzed to determine if they predicted differences in other known outcomes and demographic variables. This function provides a technique for the inclusion of covariates in the model without allowing them to influence the assignment of students to latent classes. It also provides a method for testing whether the classes are statistically different in their composition based upon characteristics which are associated with higher risk of behavioral problems. Two types of variables were analyzed under auxiliary (e). The first are variables representing outcomes that occurred throughout the school year representing behavior problems. These outcomes, at risk according to the BESS, ODR’s, suspensions and expulsions, are represented as dichotomous variables such that the means represent the proportion of the subjects assigned to that class that

are predicted by the model in each demographic category. Other variables represent demographic characteristics, attributes such as gender, race, special education status and whether or not the student is currently receiving behavioral supports, which research and practice suggest indicate increased risk of behavior concerns. Once again, the mean represents the proportion of that class predicted by the model in each category. These means can be compared across classes and models and also compared to the proportion in each category for the entire sample. The auxiliary (e) option was completed for Models 1, 4 and 5 with results displayed in Tables 10 – 12.

Table 10.
Means (SE) and chi-square statistics for distal outcomes and demographic variable composition, MS Model 1

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.05 (.01)	1 vs. 2	13.56**
	2	.19 (.04)	1 vs. 3	1.52
	3	.16 (.08)	2 vs. 3	0.16
BESS basic risk	1	.12 (.02)	1 vs. 2	36.59**
	2	.42 (.05)	1 vs. 3	0.86
	3	.21 (.09)	2 vs. 3	3.99*
BESS elevated risk	1	.03 (.01)	1 vs. 2	46.92**
	2	.33 (.04)	1 vs. 3	48.95**
	3	.74 (.10)	2 vs. 3	13.98**
ODR risk	1	.12 (.02)	1 vs. 2	36.52**
	2	.41 (.05)	1 vs. 3	26.86**
	3	.68 (.11)	2 vs. 3	5.41*
Demographic Category				
Male	1	.47 (.02)	1 vs. 2	28.14**
	2	.73 (.04)	1 vs. 3	31.27**
	3	.90 (.07)	2 vs. 3	3.95*
White	1	.86 (.02)	1 vs. 2	18.63**
	2	.66 (.04)	1 vs. 3	0.66
	3	.79 (.10)	2 vs. 3	1.44
Special Education	1	.16 (.02)	1 vs. 2	1.95
	2	.22 (.04)	1 vs. 3	0.29
	3	.21 (.09)	2 vs. 3	0.01
Behavioral Supports	1	.08 (.01)	1 vs. 2	24.35**
	2	.30 (.04)	1 vs. 3	0.93
	3	.16 (.09)	2 vs. 3	2.06
Hispanic	1	.08 (.01)	1 vs. 2	3.52
	2	.14 (.03)	1 vs. 3	0.15
	3	.11 (.07)	2 vs. 3	0.23

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether the two classes are statistically significantly different in that category.
**p < .01, *p < .05.

Table 11.
Means (SE) and chi-square statistics for distal outcomes and demographic variable composition, MS Model 4

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.02 (.01)	1 vs. 2	6.11*
	2	.08 (.02)	1 vs. 3	20.39**
	3	.18 (.03)	2 vs. 3	7.68**
BESS basic risk	1	.02 (.01)	1 vs. 2	36.66**
	2	.20 (.03)	1 vs. 3	62.55**
	3	.35 (.04)	2 vs. 3	10.29**
BESS elevated risk	1	.00 (.00)	1 vs. 2	10.76**
	2	.05 (.01)	1 vs. 3	72.91**
	3	.34 (.04)	2 vs. 3	45.60**
ODR risk	1	.08 (.02)	1 vs. 2	5.82*
	2	.16 (.02)	1 vs. 3	52.60**
	3	.42 (.04)	2 vs. 3	29.20**
Demographic Category				
Male	1	.36 (.04)	1 vs. 2	15.65**
	2	.56 (.03)	1 vs. 3	42.49**
	3	.72 (.04)	2 vs. 3	9.45**
White	1	.92 (.02)	1 vs. 2	7.59**
	2	.82 (.02)	1 vs. 3	22.67**
	3	.70 (.04)	2 vs. 3	6.80*
Special Education	1	.07 (.02)	1 vs. 2	20.82**
	2	.22 (.03)	1 vs. 3	12.79**
	3	.21 (.04)	2 vs. 3	.07
Behavioral Supports	1	.03 (.01)	1 vs. 2	16.59**
	2	.13 (.02)	1 vs. 3	31.80**
	3	.24 (.04)	2 vs. 3	7.34**
Hispanic	1	.08 (.02)	1 vs. 2	.13
	2	.07 (.02)	1 vs. 3	3.14
	3	.14 (.03)	2 vs. 3	4.65*

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether the two classes are statistically significantly different in that category.
**p < .01, *p < .05.

Table 12.
Means (SE) and chi-square statistics for distal outcomes and demographic variable composition, MS Model 5

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.01 (.01)	1 vs. 2	7.359**
	2	.06 (.02)	1 vs. 3	26.961**
	3	.15 (.02)	2 vs. 3	8.803**
BESS basic risk	1	.00 (.00)	1 vs. 2	38.53**
	2	.13 (.02)	1 vs. 3	103.08**
	3	.33 (.03)	2 vs. 3	25.43**
BESS elevated risk	1	.00 (.00)	1 vs. 2	4.86*
	2	.02 (.01)	1 vs. 3	81.40**
	3	.27 (.03)	2 vs. 3	62.97**
ODR risk	1	.06 (.02)	1 vs. 2	3.05
	2	.12 (.02)	1 vs. 3	52.87**
	3	.35 (.03)	2 vs. 3	38.96**
Demographic Category				
Male	1	.32 (.05)	1 vs. 2	10.75**
	2	.51 (.03)	1 vs. 3	39.14**
	3	.32 (.05)	2 vs. 3	13.35**
White	1	.95 (.02)	1 vs. 2	7.96**
	2	.86 (.02)	1 vs. 3	35.29**
	3	.72 (.03)	2 vs. 3	12.51**
Special Education	1	.03 (.02)	1 vs. 2	29.24**
	2	.19 (.03)	1 vs. 3	34.70**
	3	.22 (.03)	2 vs. 3	0.61
Behavioral Supports	1	.01 (.01)	1 vs. 2	53.48**
	2	.08 (.02)	1 vs. 3	21.62**
	3	.23 (.03)	2 vs. 3	53.48**
Hispanic	1	.06 (.02)	1 vs. 2	0.21
	2	.07 (.02)	1 vs. 3	3.99*
	3	.13 (.02)	2 vs. 3	3.49

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether the two classes are statistically significantly different in that category.
**p < .01, *p < .05.

Model 1. The results for Model 1 are shown in Table 10. For Model 1, the proportion of students in risk categories in Class 1 is always lower than the Class 3. In five categories, Class 2 had a higher proportion of students in the risk category than Class 3. BESS basic risk was higher for Class 2 (M = .42) versus Class 3 (M = .21) and there were a larger proportion of students receiving behavioral supports in Class 2 (M = .30) as compared to Class 3 (M = .16). This is not the expected pattern, although it is possibly due to small sample size due to Class 3 containing only 19 students. In Model 1, BESS elevated risk was highest in Class 3 (M = .74) as compared to Class 1 (M = .03). ODR risk was highest in Class 3 (M = .68) versus Class 2 (M = .41). Looking at the Chi-Square test of differences in the means, Class 1 was statistically significantly different from Class 2 in 7 categories, Class 1 was different from Class 3 in 3 categories, and Class 2 and 3 were different in 4 categories.

Model 4. The results for Model 4 are shown in Table 11. For Model 4, the proportion of students in all demographic groups associated with higher risk of behavioral problems ranked from lowest to highest for Class 1 to Class 3. In addition, the three groups were statistically significantly different from each other in mean levels in almost all cases using the Chi-Square test. Each pair of classes was different in eight out of the nine demographic groupings tested.

Model 5. The results for Model 5 are shown in Table 12. The results for Model 5 were very similar to Model 4. The proportion of students in all demographic groupings associated with increased risk of behavioral problems ranked from lowest to highest for Class 1 to Class 3. Once again, the three groups were statistically significantly different from each other in mean levels in almost all cases using the Chi-Square test. Class 1 and 3 were statistically significantly different in all nine risk categories, Classes 1 and 2 in seven categories, and Class 2 and 3 in seven categories.

Comparing Models 1, 4 and 5. Model 1 classifications contain more students in Class 1 and fewer students in Class 3, and the proportion of students in risk categories is higher in Class 3. In Class 3, the mean proportion of students in the ODR risk category was .68 (.11) for Model 1 and .42 (.04) for Model 5. Although Models 4 and 5 are very similar, Model 5 contains a larger Class 3. This resulted in a shifting of students from Class 1 to Class 2 and Class 2 to Class 3 in the comparison of Models 4 and 5. The students that shifted were less likely to be identified in one of the groups reflecting higher risk; rather, these students were identified by the model based upon the patterns in the DBR-SIS, a feature explored in the earlier section.

In summary, Model 1 assumed that all variability in the measures was reflected in the class intercepts and slope fixed effects and time varying but class invariant residual variances. Because residual variances were consistent across the latent trajectory classes, the class assignment represented differences in mean levels in the intercepts and slopes. Model 4 assumed the variability in the measures was different across the classes in the residuals but not in the intercepts and slopes. Model 5 reflected variability in the measures in three ways: class varying fixed effects, class varying random effects, and class varying residual variances. Model 5 reflects a larger proportion of students assigned to Class 3, reflecting a more sensitive screening of students who are at risk than the other models. Based upon the more favorable fit statistics, class composition, descriptive statistics, and results from auxiliary (e), Model 5 is the preferred model. In the next section, characteristics of the class composition of Model 5 are explored.

Research Question 3. What are the characteristics of the students comprising latent classes of classroom behavior trajectories that emerge from the use of GM models, and in what ways do the latent classes vary in intercepts, slope, and variances?

Table 13 provides a demographic breakdown of student composition by Class for Model 5, and Figure 10 provides the graph of the latent class trajectories. Class 1 is characterized by students who are consistently academically engaged and respectful almost all the time and disruptive almost none of the time. Class 1 is disproportionately lower on demographic characteristics associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, Hispanic and ODRs. It is comprised of 32% Male, 95% White, 1% Black, and 9% Hispanic.

Table 13

Demographic distribution of Model 5 by class—MS

Category	All MS (n = 591)	Class 1 Optimal (n = 100)	Class 2 Average (n = 262)	Class 3 Lowest (n = 229)
Male	.54	.32	.51	.67
White	.82	.95	.85	.72
Black	.13	.01	.08	.24
Retained	.06	.01	.06	.08
Special Education	.18	.03	.19	.22
EBD	.02	.00	.01	.03
Behavior	.04	.01	.02	.08
Days Absent	7.62	5.32	7.50	8.93
Suspension or Expulsion	.09	.01	.06	.14
Brisk	.19	.00	.14	.32
Erisk	.11	.00	.02	.27
Hispanic	.09	.06	.07	.13
ODRRISK	.20	.06	.12	.35

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS, ODRRISK=two or more office disciplinary referrals.

On average, students in Class 1 have DBR-SIS above 29 (see Appendix B, Table B1). The slope for the fall was very small, and the slopes for the winter and spring were not statistically significant. Intercept variances at the within level were not statistically significant. Residual variances were low but increasing over time, ranging from .06 for fall, .09 for winter, and .12 for spring.

Class 2 is characterized by students who are not consistently academically engaged and respectful and are occasionally disruptive. Class 2 has a higher proportion of students compared to Class 1 in demographic groups associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, Hispanic and ODRs. It is comprised of 51% Male, 85% White, 8% Black, and 12% Hispanic.

Students in Class 2 still have high average DBR-SIS, between 28 and 29 (see Appendix B, Table B1). Slopes were positive and statistically significant for all three data collection periods. Intercept variances at the within level were moderate and statistically significant. Residual variances were larger than for Class 1, ranging from 2.12 for fall, 0.97 for winter, and 1.05 for spring.

Class 3 is characterized by students who are highly variable in their classroom behavior, indicated by a wide spread of sample values around the average trajectory in Figure 10. Class 3 has a much higher proportion of students compared to Class 1 and 2 in demographic groups associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, Hispanic and ODRs. It is comprised of 67% Male, 72% White, 24% Black, and 13% Hispanic. Class 3 represents the class of students at-risk for failure due to behavior problems. Using the DBR-SIS and Model 5, 39% of MS student are in class 3 while using the BESS, 19% of MS students are identified as at-risk.

Students Class 3 have average DBR-SIS score around 24 (see Appendix B, Table B1). Slopes were positive and statistically significant for all three data collection periods. Intercept variances at the within level were much larger than the other Classes (e.g., fall intercept variance 16.52 compared to 0.02 for Class 1). Residual variances were very large compared to Class 1 and 2, ranging from 10.32 for fall, 8.28 for winter, and 9.42 for spring.

Variability due to Teacher. A limitation in Mplus restricts variances at the between level from varying across classes, also limiting the interpretability of the between teacher variance. Techniques exist to allow between level variances to vary across class, but they require the use of an additional latent factor at the between level which changes the model specifications. This change would make Model 5 not comparable to the other models. Future research should investigate the impact of this technique on model results.

For the MS grade group, intercepts at the between teacher level were not statistically significant. Given that other variances were larger for Class 2 and Class 3, the between teacher variance may have been larger for these classes too. As such, the statistically non-significant between teacher variances may be due to class invariance in the model. Nonetheless, there are a few comparisons that are meaningful. We can consider the change in variability from the beginning of data collection to the final data collection within each class, but not across classes. The variance (SE) in the intercept between teachers was .05 (.03) for fall, .06 (.04) for winter, and .15 (.08) for spring. Using the intraclass correlation (ICC) which compares the intercept variance between teachers to the sum of the intercept variance within teachers, between teachers, and the residual variance, for Class 1, 38.5% of the intercept variability was between teachers in the fall, 37.5% in the winter, and 51.7% in the spring. For Class 2, 1.6% of the intercept variability was between teachers in the fall, 3.7% in the winter, and 9.3% in the spring. For Class

3, 0.2% of the intercept variability was between teachers in the fall, 0.3% of the variability in the winter, and 0.6% of the variability in the spring. For all classes, the variability between teachers increased from the fall to the spring data collections.

Another way to look at the ICC involves examining the intercept variances in Model 4. Model 4 reflects a parameterization that constrains all intercept variances to equal across classes. The ICC for this comparison is calculated as the ratio of the between teacher variances to the within and between variance in the intercept. For MS, the variability in the intercept between teachers changed from 40.7% in the fall, to 39.8% in the winter, and 50.1% in the spring. The increase from fall to spring may be due to teacher familiarity with students or the positive effect of classroom management on producing more homogeneous classroom behavior patterns.

Looking further into patterns of observations at the teacher level can provide evidence to support the validity of the method taken in assigning the latent classes. The alternative method to identifying the latent classes at the student level is to identify the classes at the teacher level. An examination of the class distribution of students by teacher provided some insights into whether modeling at the teacher level would make sense. If 8 or more students within a teacher are assigned to a single class, this would support consideration of a teacher level model. Of the 62 teachers in the MS grade group sample, four teachers had 8 or more students assigned to Class 2, and seven teachers had 8 or more assigned to Class 3. The distribution of the majority of teachers followed the overall distribution with the majority of students in Class 2, supporting the use of the student level class assignment.

Summary. Five model parameterizations were explored for the MS sample using GM modeling, testing progressive increases in the number of classes and relaxing model parameters to test for potential noninvariance. Using model fit criteria, maximum likelihood ratio tests, class

demographic composition, and descriptive statistics, Model 5 with three classes represents the preferred model with the greatest screening sensitivity. The model classes can be described as follows: Class 1 is optimal and consistent, Class 2 as average behavior and less consistent, and Class 3 as poor, highly variable classroom behavior.

Upper Elementary

Research Question 1. When selecting the multilevel GM model that best represents the characteristics of a real data sample of student classroom behavior using the DBR-SIS, in what ways and to what extent does noninvariance affect decision making, including application of enumeration indices, and resulting latent class characteristics?

Model fit for Models 1-5 for UE is provided in Appendix A, Tables A2 and A3. The model specifications previously described for Models 1 and 2 were executed without incidence for the 2-class and 3-class models. For Model 1, although the information criteria indicated preference for the 4-class model over the 3-class model, the BLRT p-value for the 4-class model indicated that the fourth class did not improve model fit over the 3-class model. For the 3-class model, students were distributed among the classes with 72% in the Class 1, 23% in Class 2, and 5% in the Class 3. Class 1 has a model predicted initial status of 28.2, and increases at a very small rate over the year, ending at 29.0. Class 2 had model predicted initial status of 24.1 with a slight drop in the node from winter to spring, followed by slight increase ending at 25.2. The model predicted initial status of Class 3 was 16.3, again with a very small slope and a drop in the node from winter to spring, ending the year at 17.4.

The results for Model 2 were similar to Model 1, with more classes improving model fit up to 4 classes. The 4-class model was inadmissible due to a non-positive definite first order derivative product matrix. Therefore, the 3-class model was favored, and a statistically

significant BLRT provided evidence that the third class improved model fit over the 2-class model. The 3-class Model 2 represented 92% in Class 1, 5% in Class 2 and 3% in Class 3. Intercepts and slopes for the Class 2 and 3 predicted erratic growth patterns not easily interpretable making Model 2 an unsupportable model. Class 1 had initial status of 27.2 with a small growth rate across the year ending at 28.1. Class 2 had an initial status of 20.0 with no change across the fall. For the winter, Class 2 exhibited a large increase at the node between fall and winter followed by a small positive slope. This was followed by a large drop in the node between winter and spring, ending the year at 11.6.

Similar to MS, the Model 3 specification ran for the 2-class model but would not execute properly for the 3-class or 4-class models due to a non-positive definite latent variable matrix. Identifying a defensible sequence of models for UE proved to be more challenging than for MS for Models 4-5. Model 4 represented class invariant intercept random effects and class varying residual variances. The Model 4 parameterization executed properly for the 2-class model only and would not execute for the 3-class or 4-class models due to a non-invertible matrix for one of the classes. The model steps were followed in an attempt to execute Model 5, and similar problems were encountered. The 2-class model ran properly, but neither the 3-class nor the 4-class models were executable, again due to non-invertible matrices for one of the classes.

At this point, the goal was to identify a change in the parameterization that could be carried through to all the model steps and would be defensible based upon the substantive theory and findings from the other grade level models. A comparison of the residual variances across time points indicated that more variability existed between classes than between time points. Therefore, a change was made to constrain the residual variances to be equal across time. Residual variances were still allowed to vary between latent classes for Models 4 and 5. In

addition, for Model 5, covariances between the intercepts were constrained to be equal across classes. This parameterization resulted in executable Model 4 and 5, and the Model 2 and 3 steps were repeated using the new parameterization. The Model 2 revision aligned very closely to the original parameterization, and once again, Model 3 would not run correctly. The results for the original parameterization, UE-1, are displayed in Table B2 and the revised parameterization, UE-2, are displayed in Table B3. All further remarks on model results refer to the UE-2 parameterization.

The 3-class model for the Model 4 fit the data statistically significantly better than the 2-class model as indicated by the BLRT ($p < .001$), and the 4-class model was not estimated due to a non-invertible matrix for one of the classes. Class 1 consisted of 28% of students, Class 2 53%, and Class 3 was 19%. The model estimates for the Class 1 start at 29.3 and ended at 29.6, exhibiting almost no change across the year. For Class 2, the initial status in the fall was estimated at 27.0 and increased slightly to 28.0 by the end of the year. For Class 3, the initial status was 21.8, increasing at a greater rate than the other two classes, ending the year at 23.5.

The results for Model 5 were very similar to Model 4 with the 3-class model representing the best fitting model. Similar to Model 4, the 4-class model was not estimated due to a non-invertible matrix for one of the classes. The 3-class model had statistically significantly better fit than the 2-class model as indicated by the BLRT ($p < .001$) and lower information criteria. The distribution by class was 22% for Class 1, 50% for Class 2, and 28% for Class 3. The model estimates for the Class 1 start very high at 29.5 and stay at that level throughout the year, with statistically non-significant slope estimates. Class 2 starts at 27.6 increasing only slightly to 28.4 by the end of the year. Class 3 starts at 22.8 and increases slightly during the fall, also with very little change across the year, ending at 28.4.

Latent trajectory class comparisons among competing models. For UE, the 3-class model fit best for Models 1, 2, 4 and 5. The model trajectories for Models 1, 2, 4 and 5 are reported in Figure 13.

Figure 13. Comparison of UE Model Estimated Trajectory to Student Sample Trajectories by Model and Class.

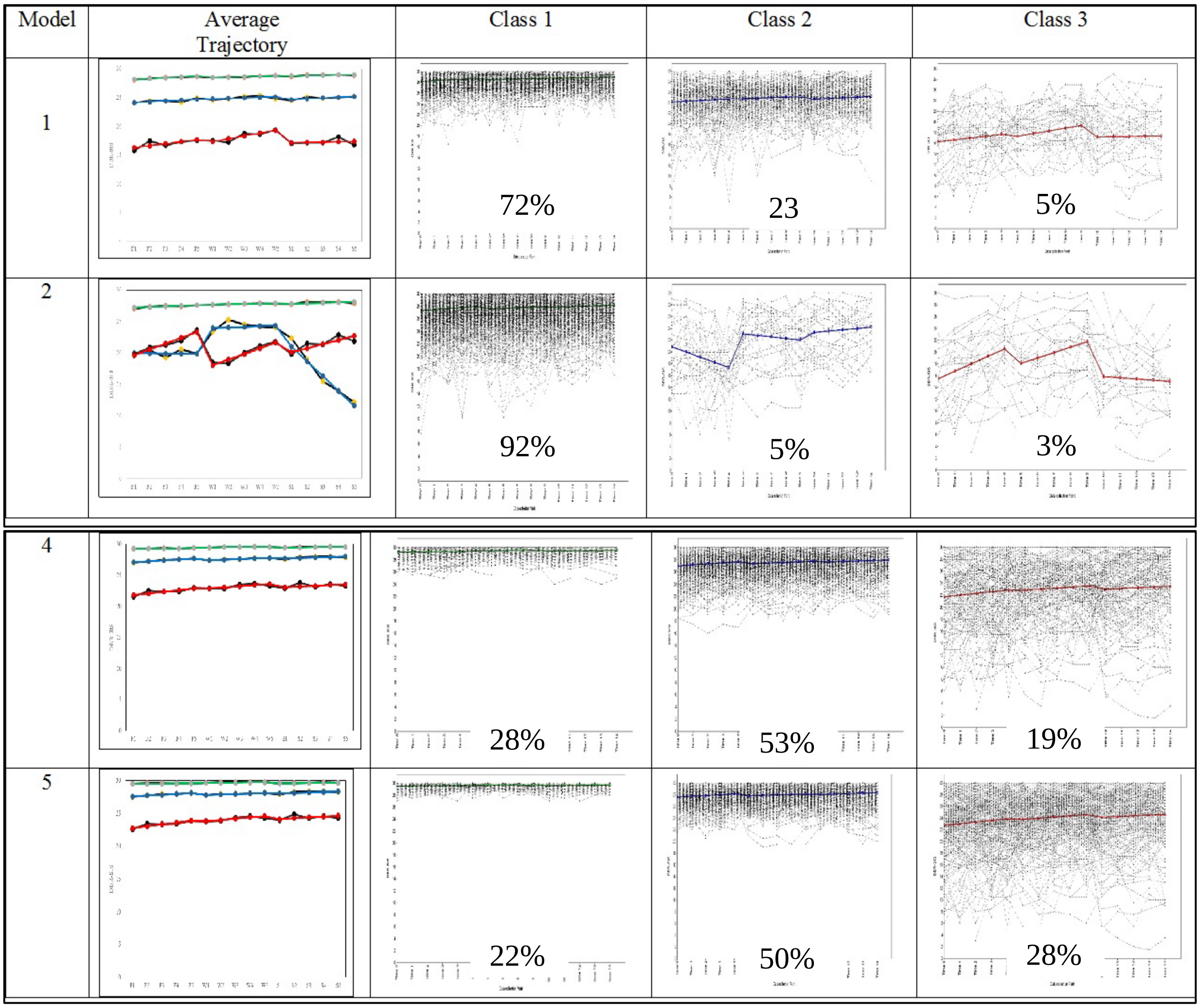


Figure 13. Comparison of UE model estimated trajectories for Models 1, 2, 4, and 5 and sample data for each of classes 1, 2, and 3. Percentages represent proportion of students assigned to each class. For each graph, the x axis represents the fifteen data collection points and the y axis represents the DBR-SIS score from 0 to 30.

For each model, Class 1 was characterized by near perfect scores, close to 30 with minimal change across the year. Class 2 for all models except Model 2 reflected a lower score

than Class 1 and minimal change across the year. For Models 1, 4 and 5, the Class 2 predicted initial value was in the range of 24 to 27, with a small positive slope across the year. The results for Model 2 in the Classes 2 and 3 were erratic, with an initial status of 20 and large fluctuations. It should be noted that these two classes only represent 5% each of the sample. The Class 3 scores started at 16.0 for Model 1, but unlike MS, the UE Class 3 did not experience much improvement over the year. Similar to the MS grade group, Class 2 in UE experienced a drop in the intercept between the winter and spring for Model 1. For Models 4 and 5 the patterns for all classes were relatively flat with the Class 3 experiencing some small amount of growth over the year.

The growth curve parameters for the best model, which is the 3-Class Model 5, are provided in Appendix B, Table B2. The growth factor variances were statistically significant for fall, winter, and spring for the Class 2 and 3 but only for winter for Class 1, indicating greater variability in the scores for classes containing students with more behavior problems. In particular the magnitude of the variance for the intercepts for Class 3 was substantially higher than the other classes. For example, the variance of the intercept for the fall was not statistically significant for Class 1, was estimated at 1.36 ($p < .01$) for Class 2, and estimated at 14.96 ($p < .01$) for Class 3. In addition, residual variances were statistically significant, and once again substantially larger in magnitude for Class 3 than the other classes. For the UE models, residual variances were invariant across time. Similar to MS, covariances, also invariant across classes, were not statistically significant at the within teacher level. Neither variances nor covariances of the intercepts were statistically significant at the between teacher level, also similar to MS.

Comparisons of the average trajectories for each class and model to the individual sample values are shown in Figure 13. Given that Model 1 represents a set of class trajectories assuming

that all heterogeneity is reflected in the mean intercepts, slopes and residual variances, the amount of variability around the average trajectories is consistently large for all three classes and represented completely in the model by class invariant residual variances. The variability around the average trajectory for Model 2 is also very large, and the trajectories for Classes 2 and 3 are difficult to interpret with a great deal of overlap.

The trajectories in Models 4 and 5 have smaller slopes indicating very little change across the year, but a large difference in the level of variability around the average trajectories. Given that Models 4 and 5 allow for class varying residual variances, the variability around the average trajectory is visibly quite different. For Class 1, there is very little variability, for Class 2, there is more and for Class 3, the variability is quite large. Similar to the MS grade group, the UE results appear to indicate that Model 1 represents a classification system based upon mean levels and Models 4 and 5 represents a classification system based upon means and variability.

Comparisons of change in class among competing models. Similar to MS, students shifted from one class to another as the models changed, and an analysis of these patterns provided insights into the nature of these models' resultant mixtures. A comparison of the change in latent class assignment from Model 1 to Model 5 is provided in Table 14. Out of 522 students assigned to Class 1 in Model 1, 32% stayed in Class 1, 60% were assigned to Class 2 and 8% were assigned to Class 3 in Model 5. Of the 166 students who were in Class 2 in Model 1, 26% stayed in Class 2 and 74% were assigned to Class 3 in Model 5. Of the 37 students that were assigned to Class 3 in Model 1, all remained in Class 3 for Model 5. It is also informative to understand the demographic composition of the students shifting class assignment from Model 1 to Model 5 as shown in Table 16. Students shifting from Class 1 to Class 2 are disproportionately white, female, and are less likely to have 2 or more ODRs. Students shifting from Class 1 to

Class 3 are disproportionately male and white. They are also more likely to be at risk based upon the BESS in the basic risk range and to be at risk based upon the number of ODRs. Students shifting from Class 2 to Class 3 were disproportionately male, black, have higher proportions of students who were retained, are more likely to receive behavior supports, are absent more, and are more likely to be at risk based upon the BESS and based upon ODRs.

Table 14.
Comparison of Latent Class Assignment Model 1 versus Model 5, UE

		Model 5 Latent Class				
		1	2	3	Model 1 Sum	Model 1 Proportion
Model 1 Latent Class	1	165	316	41	522	72%
	2	0	43	123	166	23%
	3	0	0	37	37	5%
	Model 5 Sum	165	359	201	725	
	Model 5 Proportion	22%	50%	28%		100%

Table 15 reports the change in latent class assignment from Model 4 to Model 5. Of the 203 students assigned to Class 1 in Model 4, 81% stayed in Class 1 and 19% shifted to Class 2 in Model 5. Of the 388 students in Model 4 Class 2, 82% stayed in Class 2 and 18% shifted to Class 3. All 133 students assigned to Model 4 Class 3 stayed in Class 3 for Model 5. Students who shifted from Class 1 to Class 2 were disproportionately white with lower than average proportions of students in all the behavior risk categories. Students who shifted from Class 2 to Class 3 were disproportionately male, white, special education, and had higher than average BESS basic risk.

Table 15.
Comparison of Latent Class Assignment Model 4 versus Model 5, UE

		Model 5 Latent Class				
		1	2	3	Model 4 Sum	Model 4 Proportion
Model 4 Latent Class	1	165	38	0	203	28%
	2	0	320	68	388	53%
	3	0	0	133	133	19%
	Model 5 Sum	165	358	201	725	
	Model 5 Proportion	22%	50%	28%		100%

Table 16.
Demographic distribution by change group—UE—Model 1 versus Model 5

Category	All UE	Always Class 1	Always Class 2	Always Class 3	Class 1,2	Class 1,3	Class 2,3
Male	.51	.38	.44	.76	.47	.68	.66
White	.81	.86	.74	.51	.85	.83	.75
Black	.13	.08	.23	.35	.11	.05	.20
Retained	.04	.01	.07	.14	.02	.00	.09
Special Education	.11	.05	.21	.22	.09	.05	.20
EBD	.00	.01	.02	.00	.00	.00	.01
Behavior	.03	.01	.07	.19	.02	.00	.06
Days Absent	5.55	4.72	5.85	7.50	5.64	4.46	6.15
Suspension or Expulsion	.03	.01	.00	.00	.03	.05	.09
Brisk	.16	.02	.28	.38	.10	.20	.38
Erisk	.07	.00	.07	.51	.01	.00	.21
Hispanic	.05	.05	.05	.14	.03	.02	.08
ODRRISK	.06	.01	.00	.22	.03	.15	.14
DBR mean	27.22	29.65	25.79	17.50	28.30	27.72	24.46
DBR SD	3.05	.36	.84	3.10	.77	.67	1.47
BESS T score mean	49.66	41.32	54.32	66.55	47.86	49.73	58.73
BESS T score SD	9.83	5.42	8.51	7.57	7.06	6.50	8.37
n	725	165	43	37	316	41	123

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS, ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 1 and Class b in Model 5.
Only nonempty change classes are reported.

Table 17.
Demographic distribution by change group—UE—Model 4 versus Model 5

Category	All UE	Always Class 1	Always Class 2	Always Class 3	Class 1,2	Class 2,3
Male	.51	.38	.47	.68	.50	.68
White	.81	.86	.82	.66	.95	.84
Black	.13	.08	.13	.24	.05	.12
Retained	.04	.01	.03	.12	.03	.00
Special Education	.11	.05	.12	.17	.03	.16
EBD	.00	.01	.00	.00	.00	.01
Behavior	.03	.01	.03	.08	.00	.04
Days Absent	5.55	4.72	5.65	6.43	5.80	5.26
Suspension or Expulsion	.03	.01	.03	.10	.00	.00
Brisk	.16	.02	.13	.37	.05	.29
Erisk	.07	.00	.02	.30	.00	.07
Hispanic	.05	.05	.03	.07	.05	.10
ODRRISK	.06	.01	.03	.20	.00	.07
DBR mean	27.22	29.65	27.93	22.72	28.56	26.05
DBR SD	3.05	.36	1.14	3.91	.73	1.97
BESS T score mean	49.66	41.32	48.79	60.44	47.20	54.22
BESS T score SD	9.83	5.42	7.73	9.54	5.60	7.82
n	725	165	320	133	38	68

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS, ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 4 and Class b in Model 5.
Only nonempty change classes are reported.

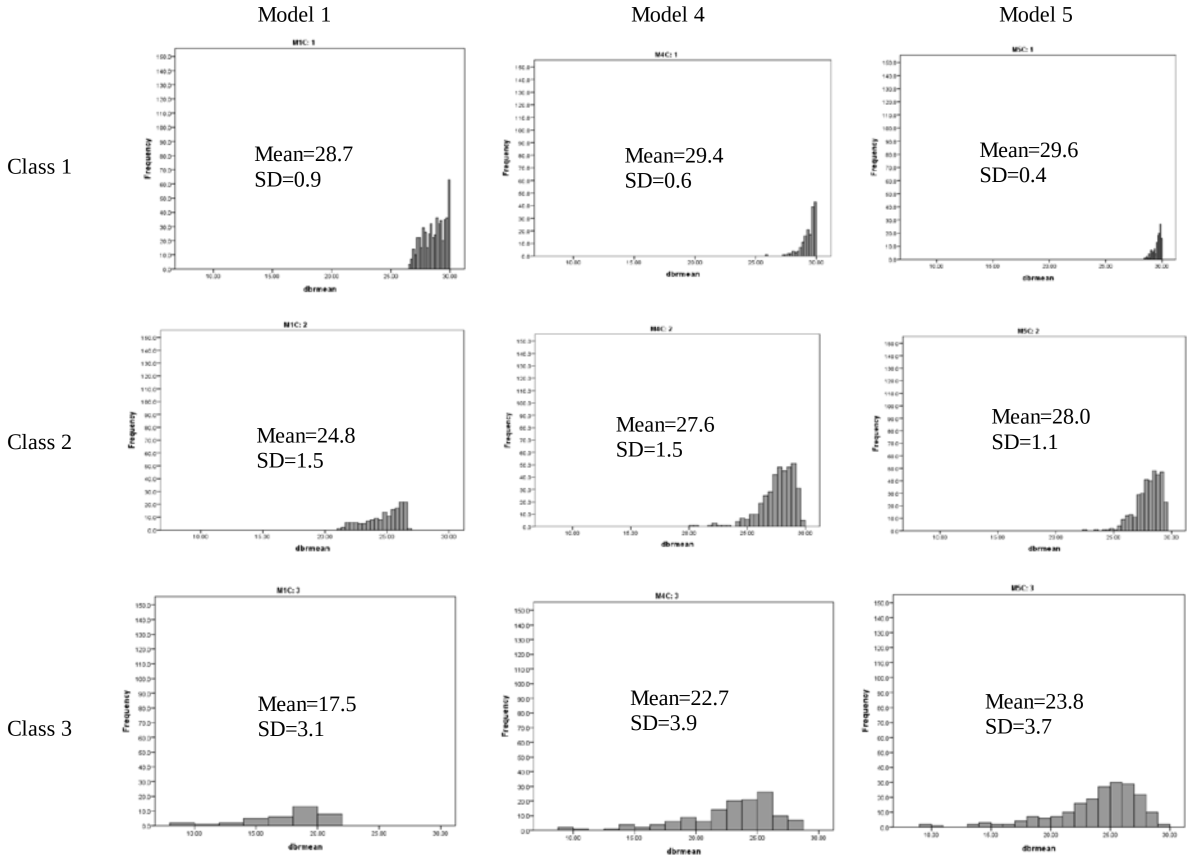
Comparisons of descriptive statistics among competing models. To understand more about the characteristics of the students by class between Models 1, 4 and 5, it is also useful to look at the distribution of means and standard deviations by model. These distributions for UE are shown in Figures 14 and 15. Starting with Class 1, the means for Class 1 are very similar between Model 1 at 28.7, Model 4 at 29.4, and Model 5 at 29.6; however, the amount of variability in the means and in the standard deviations is different. As can be seen in the graph or the values of the standard deviations, the variability within Model 1 Class 1 is much higher (M = 1.12) than Model 4 Class 1 (M=0.49) or Model 5 Class 1 (M = 0.39). Similarly with Class 2, Model 1 reflects greater variability than Model 4 or 5, and the mean values are lowest for Model 1 and highest for Model 5. The standard deviations for Model 1 Class 2 (M = 2.71) vary more than the standard deviations of Model 4 Class 2 (M = 1.55) or Model 5 Class 2 (M = 1.35). Big differences exist between the three models for Class 3. The mean of Model 1 Class 3 is quite low (M = 17.5) and the standard deviation is quite high (M = 3.99). The mean for Model 4 Class 3 is low but not as low as Model 1 (M = 22.7). The standard deviation for Model 4 Class 3 is also high (M = 3.60). The mean of Model 5 Class 3 is higher (M = 23.8) and the average standard deviation is lower (M = 3.15). Similar to MS, these comparisons provide a picture of the differences between Model 1, 4, and 5. The classes in Model 1 are more heterogeneous as indicated by higher mean standard deviations for each class. Models 4 and 5 classes are more homogeneous, with Model 5 the most homogeneous.

Similar to MS, Models 4 and 5 are very similar in the shape of the growth trajectories. Model 4 reflects class differences in residual variances, and Model 5 reflects class differences in both intercept random effects and residual variances, resulting in the reassignment of students from Classes 1 and 2 to Class 2 and Class 3 in Model 5. Sixty-eight students shifted from Class 2

in Model 4 to Class 3 in Model 5. These students have a mean DBR-SIS score of 26.0 and standard deviation of 2.28, placing them in the higher, more variable end in Model 4 Class 2. This group contained a few students whose scores were mostly very high but with 1 or 2 points that were much lower. Similar to MS, the key difference between Model 4 and Model 5 in UE is that Model 5 assigns students with lower means and higher variability to higher classes than Model 4, resulting in a set of 3 classes that are more homogeneous in their response patterns.

The skewness and kurtosis also provide insights on the relative performance of the three models. As mentioned earlier, GM modeling assumes that the distribution is comprised of a mixture of normally distributed subgroups. The classes in Model 1 are slightly negative skewed and with kurtosis that is slightly different from zero but not in the range of concern. Model 4 resulted in classes that are more negatively skewed for Class 1 and 2 and positively skewed for Class 3. Model 4 also represents a deterioration in the kurtosis for all classes. Model 5 is an improvement over Model 4 with smaller negative skewness and small positive kurtosis.

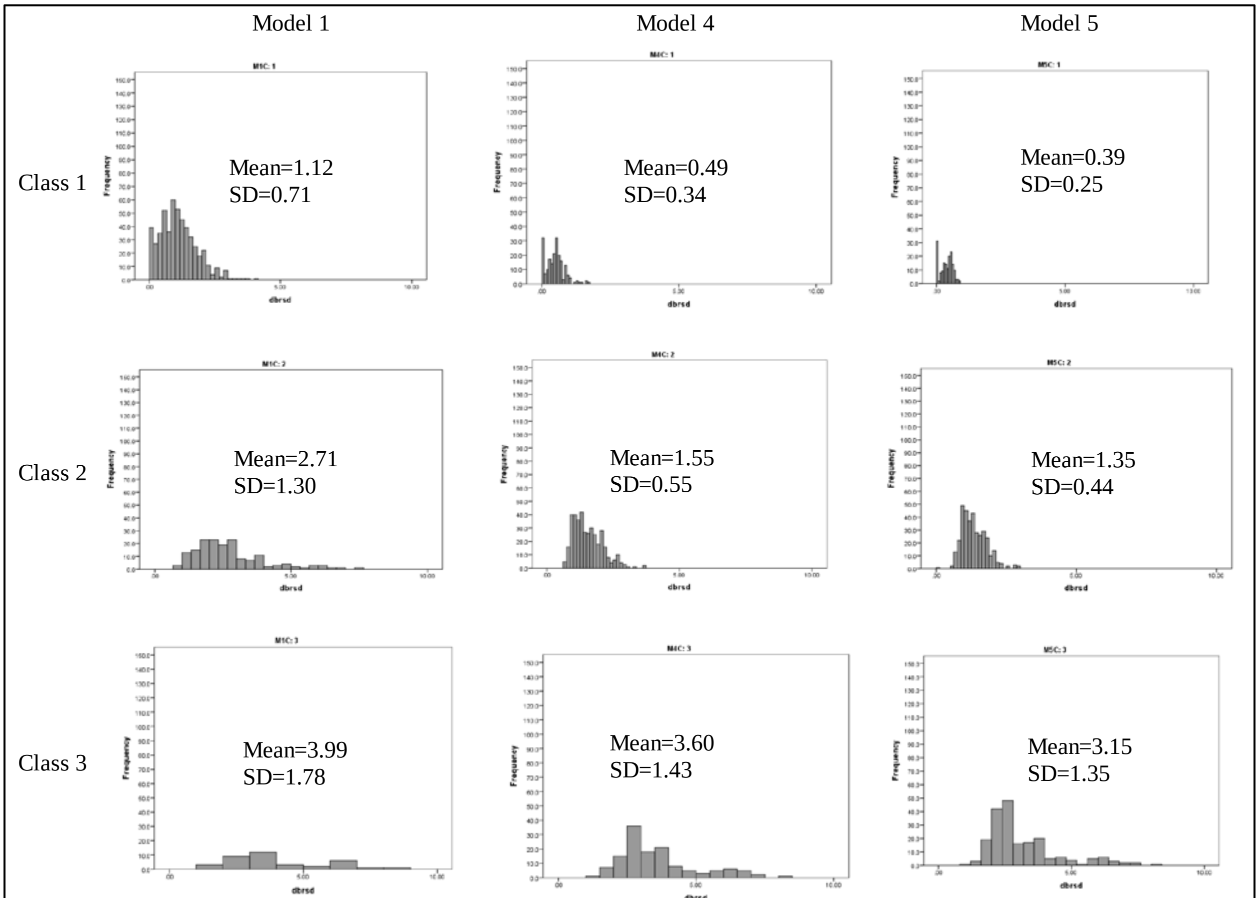
Figure 14. UE - DBR-SIS student level means by model and class



	Model 1	Model 4	Model 5
Class Proportions			
Class 1	72%	28%	22%
Class 2	23%	53%	50%
Class 3	5%	19%	28%
Skewness			
Class 1	-0.32	-1.99	-1.23
Class 2	-0.69	-1.53	-1.22
Class 3	-1.24	1.20	-1.40
Kurtosis			
Class 1	-1.02	6.00	0.91
Class 2	-0.65	3.58	1.91
Class 3	0.92	1.45	2.28
BESS T scores (SD)			
Class 1	45.9 (7.3)	42.4 (5.9)	41.3 (5.4)
Class 2	57.6 (8.6)	49.7 (8.0)	48.5 (7.5)
Class 3	66.6 (7.6)	60.4 (9.5)	58.3 (9.4)

Figure 14. Histogram displays the distribution of the within student mean DBR-SIS score by model and class for UE. The proportion of students assigned to each class, skewness, and kurtosis, are displayed at the bottom.

Figure 15. UE - DBR-SIS student level standard deviations by model and class.



Class Proportions			
Class 1	72%	28%	22%
Class 2	23%	53%	50%
Class 3	5%	19%	28%

Figure 15. Histogram displays the distribution of the within student DBR-SIS standard deviation by model and class for UE. The proportion of students assigned to each class is displayed at the bottom.

Similar to MS, Model 1 represents a classification with a very large class of well-behaved students and small average and poorly behaved classes of students. Model 1 sorts the sample based more on the mean levels. If the purpose of this classification was to identify students with severe problems with great precision, Model 1 has some merits. Models 4 and 5 sort the sample

based upon not just means but also based upon differences in the variability over time. The class proportions are more evenly distributed for both Models 4 and 5. Models 4 and 5 are also very similar in their means and standard deviations; however, consideration of the skewness and kurtosis gives preference to Model 5.

Research Question 2. To what extent does the trajectory class membership selected from research question 1 predict distal outcomes of student behavior measured using alternatives measures such the BESS risk classification, office disciplinary referrals (ODRs), and suspensions, providing validity evidence for the latent trajectory classes?

Model 1. The auxiliary (e) option in Mplus was completed for Models 1,4, and 5 and Table 18, 19, and 20 provide the results of these analyses. For Model 1, the proportion of students in each risk category increases from the Class 3 to Class 1. In Model 1, BESS elevated risk was highest in Class 3 (M = .51) as compared to Class 1 (M=.01). Looking at the Chi-Square test of differences in the means, Class 1 was statistically significantly different from Class 2 in 8 categories, Class 1 was different from Class 3 in 8 categories, and Class 2 and 3 were different in 5 categories.

Model 4. The results for Model 4 are shown in Table 19. For Model 4, the proportion of students in all demographic groups associated with higher risk of behavioral problems ranked from lowest to highest for Class 1 to Class 3. In addition, the three groups were statistically significantly different from each other in mean levels in almost all cases using the Chi-Square test. Each pair of classes was different in eight out of the nine demographic groupings tested.

Table 18. *Means (SE) and chi-square statistics for distal outcomes and demographic variable composition, UE Model 1*

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.02 (.01)	1 vs. 2	4.52*
	2	.07 (.02)	1 vs. 3	12.26**
	3	.00 (.00)	2 vs. 3	11.71**
BESS basic risk	1	.09 (.01)	1 vs. 2	44.20**
	2	.35 (.04)	1 vs. 3	13.14**
	3	.38 (.08)	2 vs. 3	0.08
BESS elevated risk	1	.01 (.00)	1 vs. 2	30.20**
	2	.17 (.03)	1 vs. 3	37.83**
	3	.51 (.08)	2 vs. 3	15.14**
ODR risk	1	.04 (.01)	1 vs. 2	7.22**
	2	.10 (.02)	1 vs. 3	6.98**
	3	.22 (.07)	2 vs. 3	2.43
Demographic Category				
Male	1	.46 (.02)	1 vs. 2	9.10**
	2	.60 (.04)	1 vs. 3	16.04**
	3	.76 (.07)	2 vs. 3	3.92*
White	1	.85 (.02)	1 vs. 2	6.77**
	2	.75 (.03)	1 vs. 3	16.06**
	3	.51 (.08)	2 vs. 3	6.99**
Special Education	1	.08 (.01)	1 vs. 2	10.53**
	2	.19 (.03)	1 vs. 3	3.94*
	3	.22 (.07)	2 vs. 3	0.12
Behavioral Supports	1	.10 (.01)	1 vs. 2	10.11**
	2	.21 (.03)	1 vs. 3	10.33**
	3	.35 (.08)	2 vs. 3	2.87
Hispanic	1	.03 (.01)	1 vs. 2	3.49
	2	.08 (.02)	1 vs. 3	3.21
	3	.14 (.06)	2 vs. 3	0.96

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether the two classes are statistically significantly different in that category.
**$p < .01$, *$p < .05$.

Table 19.
Means (SE) and chi-square statistics for distal outcomes and demographic variable composition, UE Model 4

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.01 (.01)	1 vs. 2	1.29
	2	.02 (.01)	1 vs. 3	10.27**
	3	.10 (.03)	2 vs. 3	7.46**
BESS basic risk	1	.03 (.01)	1 vs. 2	31.04**
	2	.16 (.02)	1 vs. 3	55.86**
	3	.37 (.04)	2 vs. 3	18.36**
BESS elevated risk	1	.00 (.00)	1 vs. 2	9.51**
	2	.03 (.01)	1 vs. 3	51.82**
	3	.29 (.04)	2 vs. 3	38.87**
ODR risk	1	.01 (.01)	1 vs. 2	6.43*
	2	.04 (.01)	1 vs. 3	26.54**
	3	.19 (.04)	2 vs. 3	16.82**
Demographic Category				
Male	1	.40 (.04)	1 vs. 2	5.34*
	2	.50 (.03)	1 vs. 3	28.90**
	3	.69 (.04)	2 vs. 3	14.45**
White	1	.87 (.02)	1 vs. 2	2.47
	2	.82 (.02)	1 vs. 3	17.64**
	3	.67 (.04)	2 vs. 3	10.71**
Special Education	1	.06 (.02)	1 vs. 2	5.91*
	2	.12 (.02)	1 vs. 3	9.69**
	3	.18 (.03)	2 vs. 3	2.34
Behavioral Supports	1	.07 (.02)	1 vs. 2	4.14*
	2	.13 (.02)	1 vs. 3	13.97**
	3	.23 (.04)	2 vs. 3	6.10*
Hispanic	1	.05 (.02)	1 vs. 2	0.36
	2	.04 (.01)	1 vs. 3	0.76
	3	.07 (.02)	2 vs. 3	1.88

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether the two classes are statistically significantly different in that category.
**p < .01, *p < .05.

Table 20. *Means (SE) and chi-square statistics for distal outcomes and demographic variable*

composition, UE Model 5.

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.01 (.01)	1 vs. 2	0.65
	2	.02 (.01)	1 vs. 3	7.19**
	3	.06 (.02)	2 vs. 3	4.99*
BESS basic risk	1	.02 (.01)	1 vs. 2	20.64**
	2	.12 (.02)	1 vs. 3	76.95**
	3	.34 (.03)	2 vs. 3	32.36**
BESS elevated risk	1	.00 (.00)	1 vs. 2	6.59*
	2	.02 (.01)	1 vs. 3	56.52**
	3	.22 (.03)	2 vs. 3	43.69**
ODR risk	1	.01 (.01)	1 vs. 2	2.14
	2	.03 (.01)	1 vs. 3	27.26
	3	.15 (.03)	2 vs. 3	20.23**
Demographic Category				
Male	1	.38 (.04)	1 vs. 2	3.82
	2	.47 (.03)	1 vs. 3	33.88**
	3	.68 (.03)	2 vs. 3	22.26**
White	1	.72 (.03)	1 vs. 2	0.48
	2	.84 (.02)	1 vs. 3	11.24**
	3	.86 (.03)	2 vs. 3	9.62**
Special Education	1	.05 (.02)	1 vs. 2	4.57*
	2	.11 (.02)	1 vs. 3	12.43**
	3	.17 (.03)	2 vs. 3	3.68
Behavioral Supports	1	.08 (.02)	1 vs. 2	2.47
	2	.12 (.02)	1 vs. 3	12.24**
	3	.20 (.03)	2 vs. 3	5.58*
Hispanic	1	.05 (.02)	1 vs. 2	0.79
	2	.03 (.01)	1 vs. 3	1.2
	3	.08 (.02)	2 vs. 3	4.60*

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether the two classes are statistically significantly different in that category.
**p < .01, *p < .05.

Model 5. The results for Model 5 are shown in Table 20. The results for Model 5 were very similar to Model 4. The proportion of students in all demographic groupings associated with

increased risk of behavioral problems ranked from lowest to highest for Class 1 to Class 3. Once again, the three groups were statistically significantly different from each other in mean levels in almost all cases using the Chi-Square test. Class 1 and 3 were statistically significantly different in eight risk categories, Classes 1 and 2 in three categories, and Class 2 and 3 in seven categories.

Based upon the more favorable fit statistics, class composition, descriptive statistics, and results from auxiliary (e), Model 5 is the preferred model. In the next section, characteristics of the class composition of Model 5 are explored.

Research Question 3. What are the characteristics of the students comprising latent classes of classroom behavior trajectories that emerge from the use of GM models, and in what ways do the latent classes vary in intercepts, slope, and variances?

Table 21 provides a demographic breakdown of student composition by Class for Model 5. Refer to Figure 13 provides the graph of the latent class trajectories. Just like MS, Class 1 is characterized by students who are consistently academically engaged and respectful almost all the time and disruptive almost none of the time. Class 1 is disproportionately lower on demographic characteristics associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, Hispanic and ODRs. It is comprised of 38% Male, 86% White, 8% Black, and 5% Hispanic. Students in this class have average DBR-SIS scores that are near 30 and do not change much over time (see Appendix B, Table B3). None of the Class 1 slopes were statistically significant. Intercept variances at the within and between level for the fall and spring were not statistically significant and residual variances were low, 0.15 (0.03).

Table 21.
Demographic distribution of Model 5 by class—UE

Category	All UE (n = 725)	Class 1 Optimal (n = 165)	Class 2 Average (n = 359)	Class 3 Lowest (n = 201)
Male	.51	.38	.47	.68
White	.81	.86	.84	.72
Black	.13	.08	.12	.20
Retained	.04	.01	.03	.08
Special Education	.11	.05	.11	.17
EBD	.00	.01	.00	.00
Behavior	.03	.01	.02	.07
Days Absent	5.55	4.72	5.67	6.03
Suspension or Expulsion	.03	.01	.02	.06
Brisk	.16	.02	.12	.34
Erisk	.07	.00	.02	.22
Hispanic	.05	.05	.03	.08
ODRRISK	.06	.01	.03	.15

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS, ODRRISK=two or more office disciplinary referrals.

Class 2 is characterized by students who are not consistently academically engaged and respectful and are occasionally disruptive. Class 2 has a higher proportion of students compared to Class 1 in demographic groups associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, and ODRs. It is comprised of 47% Male, 84% White, 12% Black, and 3% Hispanic. Students in this class have high average DBR-SIS scores near 27.6 that increase slightly over the year. Slopes were positive

and statistically significant for all three data collection periods. Intercept variances at the within level were moderate and statistically significant but larger than Class 1. The residual variance was estimated at 1.33 (0.17), which was larger than Class 1.

Class 3 is characterized by students who are highly variable in their classroom behavior, indicated by a wide spread of sample values around the average trajectory in Figure 13. Class 3 has a much higher proportion of students compared to Class 1 and 2 in demographic groups associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, Hispanic and ODRs. It is comprised of 68% Male, 72% White, 20% Black, and 8% Hispanic. Students in this class have much lower average DBR-SIS score that increase slightly over the year, starting at 22.9 and ending the year at 24.5. Slopes were positive and statistically significant for all three data collection periods. Intercept variances at the within level were much larger than the other classes (e.g., fall intercept variance 14.96 compared to 0.19 for Class 1 and 1.36 for Class 2). The intercept variances at the between level were not statistically significant. The estimated residual variance was 6.95, large compared to Class 1 and 2. Using the DBR-SIS and Model 5, 28% of UE student are in class 3 while using the BESS, 16% of UE students are identified as at-risk.

The findings for UE were similar to that of MS. Model specifications were explored for the UE sample using GM modeling, testing progressive increases in the number of classes and relaxing model parameters to reflect potential noninvariance. Using model fit criteria, maximum likelihood ratio tests, class demographic composition, and descriptive statistics, Model 5 with three classes represents the preferred model with the greatest screening sensitivity. The model classes can be described as follows: Class 1 is optimal and consistent, Class 2 as average behavior and less consistent, and Class 3 as poor, highly variable classroom behavior.

Variability due to Teacher. The variances of the intercepts between teachers were not statistically significant. The variance (SE) in the intercept between teachers was .37 (.24) for fall, .65 (.66) for winter, and 1.25 (.74) for spring. For all classes, the amount of variability between teachers increased substantially from the fall to the spring data collection. Using the intraclass correlation which compares between teacher variability to within, between, and residual variances, for Class 1, the proportion of variability between teachers is 52.1% in the fall, 72.2% in the winter, and 77.2% in the spring. For Class 2, 12.1% of the variability was between teachers in the fall, 21.7% in the winter, and 35.6% in the spring. For Class 3, 1.7% of the variability was between teachers in the fall, 3.9% in the winter, and 6.0% in the spring.

Another way to look at the ICC involves examining the intercept variances in Model 4. Model 4 reflects a parameterization that constrains all intercept variances to equal across classes. The ICC for this comparison is calculated as the ratio of the between teacher variances to the within and between variance in the intercept. For UE, the between teacher variability increased from 32.5% in the fall, to 53.3% in the winter, and 59.3% in the spring. As mentioned previously, this phenomenon could be due to teacher familiarity with students or the positive effect of classroom management on producing more homogeneous classroom behavior patterns.

Looking further into patterns of observations at the teacher level can provide evidence to support the validity of assigning latent classes at the student level as compared to the teacher level. An examination of the class distribution of students by teacher provided some insights into whether modeling at the teacher level would make sense. Similar to what was done for the MS grade group, if 8 or more students within a teacher are assigned to a single class, this would support consideration of a teacher level model. Of the 73 teachers in the UE grade group sample, one teacher had all 10 students assigned to Class 1, twelve teachers had 8 or more students

assigned to Class 2, and five teachers had 8 or more assigned to Class 3. The distribution of the majority of teachers followed the overall distribution with the majority of students in Class 2, supporting the use of the student level class assignment.

Summary. Five model parameterizations were explored for the MS sample using GM modeling, testing progressive increases in the number of classes and relaxing model parameters to test for potential noninvariance. Using model fit criteria, maximum likelihood ratio tests, class demographic composition, and descriptive statistics, Model 5 with three classes represents the preferred model with the greatest screening sensitivity. The model classes can be described as follows: Class 1 is optimal and consistent, Class 2 as average behavior and less consistent, and Class 3 as poor, highly variable classroom behavior.

Lower Elementary

Research Question 1. When selecting the multilevel GM model that best represents the characteristics of a real data sample of student classroom behavior using the DBR-SIS, in what ways and to what extent does noninvariance affect decision making, including application of enumeration indices, and resulting latent class characteristics?

The model fit results for Models 1-5 for LE are provided in Appendix A, Table A4. For Model 1 using AIC, BIC, SABIC, and entropy, the 5-class model had the best fit. Although fit indices indicated preference for the 5-class model over the 4-class model, the BLRT p-value for the 5-class model indicated that the fifth class did not improve model fit over the 4-class model. The 4-class model was selected as the best fitting Model 1. In Model 1, students were distributed among the classes with 49% in the Class 1, 28% in Class 2, and 15% in Class 3 and 8% in the Class 4. Class 1 had a model predicted initial status of 28.7, and increased at a very small rate over the year, ending at 29.2. Class 2 has a model predicted initial status of 25.2 increasing at a

very small rate across the year and ending at 27.1, and it drops at the node representing the change from the fall to the winter data collections. Class 3 had model predicted initial status of 22.6 with a very large drop in the node from fall to winter, followed by slight increase ending at 23.7. The model predicted initial status of the Class 4 was 16.0. Class 4 had the largest rate of change, ending the spring at 20.9.

Models 2-5 represent a series of increasingly complex models estimated using from two to four classes. For each of Models 2 and 4, only the 2-class and 3-class models ran properly, and similar to MS and UE, Model 3 did not run correctly for the 3-class model due to a non-positive definite latent variable matrix. The BLRT for the 3-class model for both Models 2 and 4 indicated an improvement in fit over the 2-class model ($p < .001$). Combined with the non-convergence of the 4-class model, the 3-class models were selected as the best model for each. For Model 2, the distribution by class for the 3-class model was 90% for Class 1, 5% for Class 2, and 5% for Class 3. The model estimates for Model 2 Class 1 have initial status at 26.6 and increase at a small rate across the year ending at 27.5. The model estimates for Class 2 and 3 indicate a very erratic pattern, similar to MS and UE, reducing the interpretability of the results. Class 2 starts at 21.2 and increases at a low rate during the fall. Class 2 drops from 23.6 to 15.4 from the fall to winter nodes, and then increases at an increased rate over the winter. Class 2 then increases from the winter to the spring nodes and ends at 23.6. Class 3 starts very low at 15.3 and increases steadily across the year, ending at 23.9, slightly higher than Class 2.

Model 4 represented class invariant intercept random effects and class varying residual variances. This model provided improved model fit for the 3-class model, but the 4-class model was not estimated due to a non-positive definite covariance matrix. The distribution by class changed substantially to 17% for Class 1, 57% for Class 2, and 26% for Class 3. The model

estimates for Class 1 start very high at 29.4 and stays at that level throughout the year, with statistically non-significant slope estimates. Class 2 starts at 26.6 increasing only slightly to 27.6 by the end of the year. Class 3 starts at 21.5 and increases slightly ending at 24.4.

Model 5 with three classes represented the best fitting model. The 4-class model was not estimated due to a non-positive definite covariance matrix. The 3-class model had statistically significantly better fit than the 2-class model as indicated by the BLRT ($p < .001$). The distribution by class was 16% for Class 1, 50% for Class 2, and 34% for Class 3. The model estimates for Class 1 start very high at 29.5 and stays at that level throughout the year, with statistically non-significant slope estimates. Class 2 starts at 27.2 increasing only slightly to 28.1 by the end of the year. Class 3 starts at 22.0 and increases slightly during the fall, drops at the beginning of the winter followed by a slight increase through the spring, ending at 24.5.

Comparison of Model 1 3-class to 4-class. For LE, the 4-class model for Model 1 had the best model fit, which was not typical when compared to the other grade groups. Comparing the 3-class to the 4-class models, some students shifted from each class of the 3-class Model 1 to the next lower class in the 4-class Model 1. Table 22 provides the breakdown of which students shifted. Of the 371 LE students originally in Class 1, 320 stayed in Class 1 and 51 shifted from Class 1 to Class 2. Of the 202 students originally assigned to Class 2, 133 stayed in Class 2 and 69 moved to Class 3. Of the 83 students originally assigned to Class 3, 32 remained in Class 3 and 51 shifted to Class 4. In each case, students with the lowest DBR-SIS scores in the Class were the ones that shifted to the lower class in the 4-class model. Students that shifted down were more likely to be in a higher risk demographic group. This resulted in a 4th class with much lower initial status in comparison to the other grade group Class 3.

Table 22.
Demographic distribution by change group, Model 1 3-class to 4-class—LE

Category	All LE	Always Class 1	Always Class 2	Always Class 3	Class 1,2	Class 2,3	Class 3,4
Male	.52	.42	.62	.69	.58	.61	.61
White	.82	.84	.82	.88	.85	.78	.69
Black	.10	.08	.10	.09	.06	.16	.18
Retained	.06	.03	.08	.13	.02	.12	.16
Special Education	.11	.07	.14	.19	.00	.23	.20
EBD	.01	.00	.02	.00	.00	.03	.02
Behavior	.06	.00	.05	.13	.02	.16	.25
Days Absent	6.49	5.96	6.70	5.88	6.19	7.79	8.12
Suspension or Expulsion	.01	.01	.00	.00	.02	.00	.08
Brisk	.20	.05	.29	.41	.19	.39	.49
Erisk	.06	.01	.03	.19	.02	.20	.29
Hispanic	.09	.07	.12	.06	.13	.10	.06
ODRRISK	.11	.04	.13	.34	.10	.20	.25
n	657	320	133	32	51	69	51

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS, ODRRISK=two or more office disciplinary referrals.

Latent trajectory class comparisons among competing models. The model trajectories for Models 1, 2, 4 and 5 are displayed in Figure 16. For each model, Class 1 was characterized by near perfect scores, close to 30 with very little change across the year. Similar to the MS and UE grade group in Model 1, Class 3 in LE experienced a drop in the intercept between the fall and winter. The Model 2 results were once again highly erratic for Classes 2 and 3. The initial status of Model 2 class 2 was 21.3, but there was a large drop in the node between fall and winter, resulting in a drop in the winter initial status to 16.7. Class 3 scores started at 15.3 for Model 2, but experienced steady improvement over the year as indicated by a larger positive

slope. Models 4 and 5 exhibited patterns similar to each other and similar to MS and UE. Class 2 had initial status of 27 with almost no change across the year. Class 3 for both Models started from an initial status around 22, but with a downward shift in the node between fall and winter, followed by steady but small improvement over the remainder of the year.

Figure 16. Comparison of LE Model Estimated Trajectory to Student Sample Trajectories by Model and Class.

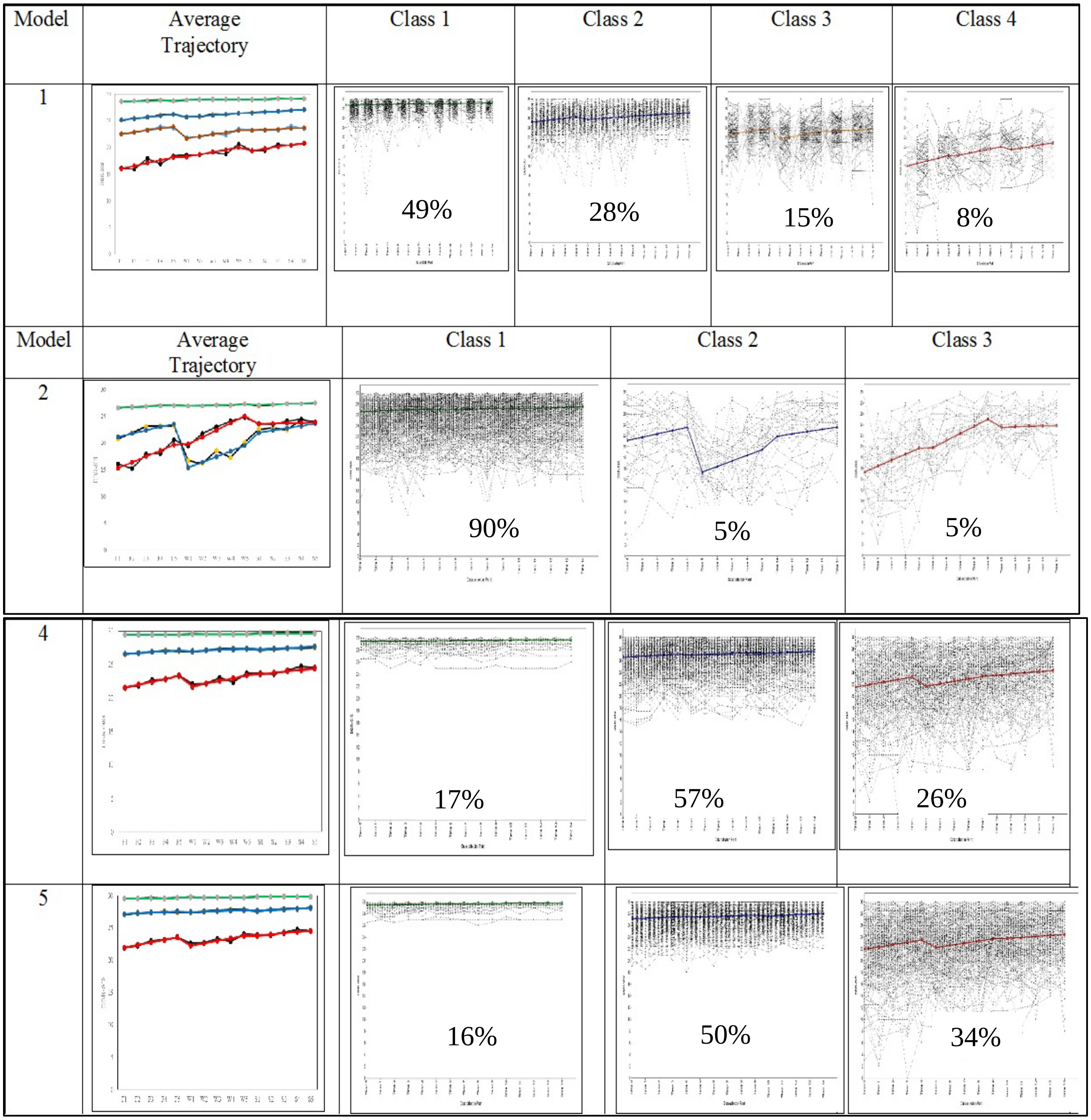


Figure 16. Comparison of LE model estimated trajectories for Models 1, 2, 4, and 5 and sample data for each of classes 1, 2, and 3. Percentages represent proportion of students assigned to each class. For each graph, the x axis represents the fifteen data collection points and the y axis represents the DBR-SIS score from 0 to 30.

The growth curve parameters for the best model, which is the 3-Class Model 5, are provided in Appendix B, Table B3. The growth factor variances were statistically significant for fall, winter, and spring for Classes 2 and 3. Growth factor variances were statistically significant only for winter for Class 1. This indicated greater variability in the scores for the more problematic classes than Class 1. In particular the magnitude of the variance for the intercepts of Class 3 was substantially higher than the other classes. For example, the variance of the intercept for the fall was not statistically significant for Class 1, estimated at 2.80 ($p < .01$) for Class 2, and 15.07 ($p < .01$) for Class 3. In addition, residual variances were statistically significant for all time periods, and once again substantially larger in magnitude for Class 3 than the other classes. Covariances for fall for Class 2 and all data collection periods for Class 3 were statistically significant at the within teacher level. Neither variances nor covariances of the intercepts were statistically significant at the between teacher level.

Comparisons of the average trajectories for each class and model to the individual sample values are shown in Figure 16. Given that Model 1 represents a set of class trajectories assuming that all heterogeneity is reflected in the mean intercepts, slopes and residual variances, the amount of variability around the average trajectories is large for all four classes and represented completely in the model by the residual variances. Model 2 is also very large, and the trajectories for Classes 2 and 3 are difficult to interpret with a great deal of overlap. The trajectories in Models 4 and 5 have smaller slopes indicating very little change across the year, but a large difference in the level of variability around the average trajectories. For Class 1, there is very little variability, for Class 2, there is more and for Class 3, the variability is quite large. Similar to the MS and UE grade groups, the LE results appear to indicate that Model 1 represents a

classification system based upon mean levels and Models 4 and 5 represent a classification system based upon mean and variability.

Comparisons of change in class among competing models. Tables 23 and 24 provide details of how students shifted in class assignment from Model 1 to 5 and Model 4 to 5. Considering Models 1 and 5, out of 320 students assigned to Class 1 in Model 1, 32% remained in Class 1, 63% were assigned to Model 5 Class 2 and 5% were assigned to Model 5 Class 3. Of the 185 students who were in Model 1 Class 2, 1% were assigned to Class 1 in Model 5 and 65% remained in Class 2, and 34% were assigned to Class 3 in Model 5. Of the 101 students were assigned to Class 3 in Model 1, 13% were assigned to Class 2 in Model 5 and 87% remained in Class 3. All 51 students in Model 1 Class 4 were assigned to Model 5 Class 3.

Table 23.
Comparison of latent class assignment Model 1 versus Model 5, LE.

		Model 5 Latent Class				
		1	2	3	Model 1 Sum	Model 1 Proportion
Model 1 Latent Class	1	103	201	16	320	49%
	2	2	120	63	185	28%
	3	0	13	88	101	15%
	4	0	0	51	51	8%
	Model 5 Sum	105	334	218	657	
	Model 5 Proportion	16%	50%	34%		100%

Considering Models 4 and 5 in Table 24, out of 112 students assigned to Model 4 Class 1, 92% remained in Class 1 and 8% were assigned to Model 5 Class 2. Of the 373 students who were in Model 4 Class 2, !% were assigned to Class 1 in Model 5, 87% remained in Class 2, and 12% were assigned to Class 3 in Model 5. All 172 students in Model 4 Class 3 were assigned to Model 5 Class 3.

Table 24.
Comparison of latent class assignment Model 4 versus Model 5, LE.

		Model 5 Latent Class				
		1	2	3	Model 4 Sum	Model 4 Proportion
Model 4 Latent Class	1	103	9	0	112	17%
	2	2	325	46	373	57%
	3	0	0	172	172	26%
	Model 5 Sum	105	334	218	657	
	Model 5 Proportion	16%	50%	34%		100%

It is also informative to understand the demographic composition of the students shifting class assignment from Model 1 to Model 5 and Model 4 to Model 5 as shown in Tables 25 and 26. Considering Model 1 versus Model 5, students shifting from Class 1 to Class 2 are disproportionately white, female, and have fewer ODRs, suspensions and expulsions than average for the entire group. Students shifting from Class 2 to Class 3 were disproportionately male, black, have higher proportions of students who were retained, are more likely to receive behavior supports, are absent more, and are more likely to be at risk based upon the BESS and based upon ODRs. Students shifting from Class 4 in Model 1 to Class 3 in Model 5 were also more likely to reflect higher risk characteristics.

Comparing students who shifted from Model 4 to Model 5. Students shifting from Class 1 to Class 2 are disproportionately white and male. Students shifting from Class 1 to 2 are more likely to be at risk based upon the BESS but less likely to be at risk based upon ODRs. Students shifting from Class 2 to Class 3 were disproportionately white, have higher proportions of students who were retained, are more likely to receive behavior supports, are absent more, and are more likely to be at risk based upon the BESS.

Table 25.
Demographic distribution by change group-LE-Model 1 versus Model 5.

Category	All LE	Always Class 1	Always Class 2	Always Class 3	Class 1,2	Class 1,3	Class 2,1	Class 2,3	Class 3,2	Class 4,3
Male	.52	.30	.63	.63	.48	.50	.00	.60	.69	.61
White	.82	.85	.88	.80	.84	.81	1.00	.73	.92	.69
Black	.10	.05	.08	.16	.09	.13	.00	.11	.00	.18
Retained	.06	.01	.06	.13	.04	.13	.00	.06	.08	.16
Special Education	.11	.06	.11	.20	.08	.00	.00	.08	.31	.20
EBD	.01	.00	.01	.02	.00	.00	.00	.02	.00	.02
Behavior	.06	.01	.03	.16	.00	.00	.00	.08	.08	.25
Days Absent	6.49	5.59	6.31	7.14	6.22	5.21	4.00	7.13	7.85	8.12
Suspension or Expulsion	.01	.00	.00	.00	.01	.00	.00	.02	.00	.08
Brisk	.20	.01	.23	.38	.07	.13	.00	.35	.54	.49
Erisk	.06	.00	.03	.22	.01	.00	.00	.03	.08	.29
Hispanic	.09	.06	.12	.09	.07	.00	.00	.14	.08	.06
ODRRISK	.11	.02	.10	.27	.04	.13	.00	.16	.08	.25
DBR mean	26.43	29.74	26.29	22.93	28.62	28.10	27.02	25.94	23.75	18.20
DBR SD	3.38	.27	.80	1.15	.60	.49	.09	.87	.71	2.79
BESS T score mean	50.41	40.04	52.30	59.28	46.17	45.73	46.50	53.50	57.38	64.28
BESS T score SD	9.88	5.42	6.88	7.31	6.64	6.73	3.06	7.81	7.52	8.28
n	657	103	120	88	201	16	2	63	13	51

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS, ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 1 and Class b in Model 5.
Only nonempty change classes are reported.

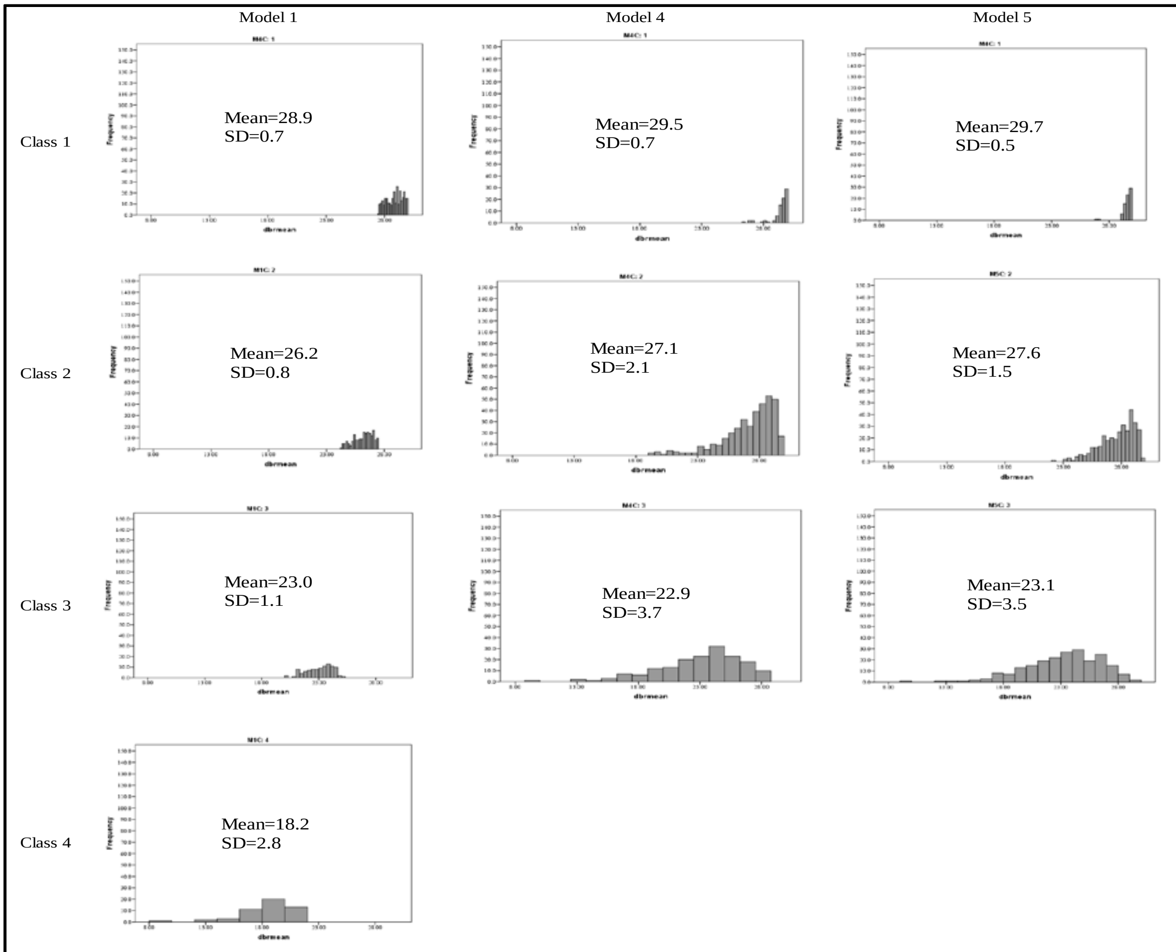
Table 26.
Demographic distribution by change group-LE-Model 4 versus Model 5.

Category	All LE	Always Class 1	Always Class 2	Always Class 3	Class 1,2	Class 2,1	Class 2,3
Male	.52	.30	.54	.63	.56	.00	.52
White	.82	.85	.86	.72	.89	1.00	.89
Black	.10	.05	.09	.17	.00	.00	.04
Retained	.06	.01	.05	.11	.00	.00	.13
Special Education	.11	.06	.10	.15	.22	.00	.15
EBD	.01	.00	.01	.02	.00	.00	.02
Behavior	.06	.01	.01	.15	.00	.00	.15
Days Absent	6.49	5.57	6.33	7.33	5.67	4.00	6.80
Suspension or Expulsion	.01	.00	.01	.03	.00	.00	.00
Brisk	.20	.01	.14	.40	.22	.00	.30
Erisk	.06	.00	.02	.20	.00	.00	.02
Hispanic	.09	.06	.09	.11	.11	.00	.02
ODRRISK	.11	.02	.07	.26	.00	.00	.11
DBR mean	26.43	29.70	27.59	22.86	27.80	29.53	23.84
DBR SD	3.38	.46	1.53	3.65	.86	.04	2.90
BESS T score mean	50.41	40.14	48.83	58.78	48.07	41.25	54.07
BESS T score SD	9.88	5.50	7.53	9.28	8.31	2.48	7.81
n	657	103	325	172	9	2	46

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS, ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 4 and Class b in Model 5.
Only nonempty change classes are reported.

Comparisons of descriptive statistics among competing models. A comparison of the distributions of the student means and standard deviations by class for Models 1, 4 and 5 for LE are displayed in Figures 17 and 18. Starting with Class 1, the means are very similar between Model 1 at 28.9, Model 4 at 29.5 and Model 5 at 29.7. Model 1 contains a larger proportion of students in Class 1 than either Models 4 or 5. As can be seen in the graph or the values of the standard deviations, the variability within Model 1 Class 1 is much higher (M = 1.01) than either Model 4 Class 1 (M = 0.39) or Model 5 Class 1 (M = 0.34). Similarly with Class 2, Model 1 reflects greater variability than Models 4 and 5. Class 2 mean values are lower in Model 1 (M = 26.2) versus Model 4 (M = 27.1) or Model 5 (M = 27.6). The average standard deviation for Class 2 indicates much more variability in Model 1 (M=2.21) compared to Models 4 and 5 (M=1.51 and M=1.38). Class 3 in Model 1 is not comparable to Class 3 in Models 4 and 5 because of the existence of Model 1 Class 4 which contains students with lower mean scores and greater variability. Yet surprisingly, the mean values are very similar. The Models 1, 4, and 5 mean DBR-SIS for Class 3 are as 23.0, 22.9, and 23.1. The mean of Model 1 Class 4 is quite low (M = 18.2) and the average standard deviation is quite high (M = 3.92). These comparisons provide a picture of the differences between Models 1, 4 and 5, where Model 1 sorts the sample based more on the mean levels and Models 4 and 5 sort the sample based upon both means but variability over time.

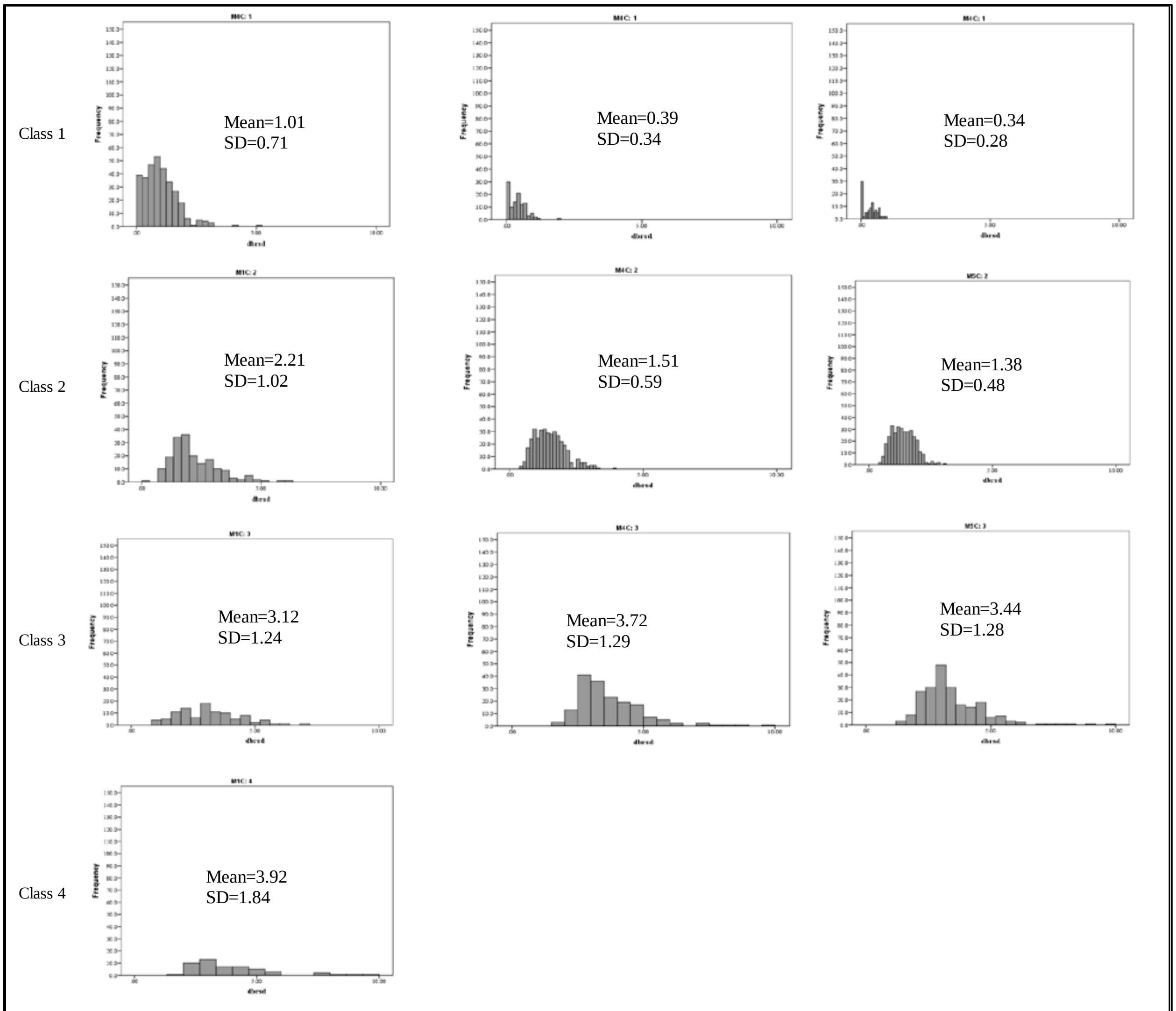
Figure 17. LE-DBR-SIS student level means by model and class.



	Model 1	Model 4	Model 5
Class Proportions			
Class 1	49%	17%	16%
Class 2	28%	57%	50%
Class 3	15%	26%	34%
Class 4	8%	-	-
Skewness			
Class 1	-0.26	-2.67	-3.87
Class 2	-0.43	-1.51	-0.85
Class 3	-0.48	-1.12	-1.05
Class 4	-2.06	-	-
Kurtosis			
Class 1	-1.10	7.31	20.13
Class 2	-0.79	2.04	0.22
Class 3	-0.54	2.08	2.08
Class 4	5.15	-	-
BESS T scores (SD)			
Class 1	44.2 (6.9)	40.8 (6.1)	40.2 (5.5)
Class 2	52.6 (7.2)	49.4 (7.8)	48.8 (7.5)
Class 3	59.0 (7.3)	58.8 (9.3)	57.8 (9.2)
Class 4	64.3 (8.2)		

Figure 17. Histogram displays the distribution of the within student DBR-SIS mean by model and class for LE. The proportions of students assigned to each class, skewness and kurtosis are displayed on the bottom.

Figure 18. LE-DBR-SIS student level standard deviations by model and class.



Class Proportions			
Class 1	49%	17%	16%
Class 2	28%	57%	50%
Class 3	15%	26%	34%
Class 4	8%	-	-

Figure 18. Histogram displays the distribution of the within student DBR-SIS standard deviations by model and class for LE. The proportions of students assigned to each class are displayed on the bottom.

Because Models 4 and 5 are very similar in their parameter estimates and growth trajectories, it is helpful to consider how they differ in their descriptive statistics. Comparatively fewer students in the LE grade group shifted from Model 4 Class 2 to Model 5 Class 3 relative to MS and UE. These students have a mean DBR-SIS of 23.9 and standard deviation of 2.42. Once again, these students are lower than the average for Model 4 Class 2 and also reflect higher score variability. This evidence provides more support for the statement above that the Model 4 classes are more heterogeneous than the Model 5 classes.

The skewness and kurtosis do not provide as clear a picture of the differences in the models for LE as the other two grade groups. The classes in Model 1 are slightly negatively skewed, Class 4 being the worse. The kurtosis for Model 1 is also not of concern except for Class 4 which has kurtosis = 5.15. Model 4 skewness and kurtosis deteriorates over Model 1, but Model 5 also does not show improvement over Model 4. Skewness and kurtosis statistics do not provide the clear picture of mixtures of normally distributed subgroups for LE as they did for the other two grade groups.

These comparisons provide a description of the differences between Models 1, 4 and 5. Model 1 contains 4 classes in total. Class 1 represents students with nearly perfect scores, and Class 2 represents students with scores near 27 that increase slightly through the year. Model 1 also contains two lower classes. Class 3 starts higher and drops in the node between fall and winter, and Class 4 starts very low and increases throughout the year. It appears that LE may have had a larger number of students that experienced a large drop between fall and winter, a phenomenon that could be due to the teachers or to student awareness of the rating process. If the purpose of this classification was to identify students with severe problems with great precision, once again Model 1 has some merits in identifying a group of 8% of the students with highly

inconsistent behavior. Models 4 and 5 sort the sample based upon not just means but also based upon differences in the variability over time. The class proportions are more evenly distributed for both Models 4 and 5. Models 4 and 5 are also very similar in their means and standard deviations. Unlike MS and UE, neither Models 4 nor 5 outrank the other for best model based upon student composition due to their similarity, and neither have stellar skewness and kurtosis statistics by class. Nonetheless, Model 5 represents a more homogeneous set of students within each class and improved model fit over all other models.

Research Question 2. To what extent does the trajectory class membership selected from research question 1 predict distal outcomes of student behavior measured using alternatives measures such the BESS risk classification, office disciplinary referrals (ODRs), and suspensions, providing validity evidence for the latent trajectory classes?

The auxiliary (e) option from Mplus was once again completed for Models 1, 4 and 5, and Tables 24-27 provide the results of these analyses. The results from auxiliary (e) for Model 1 was split into two tables, Table 24 and 25, because the Model 1 4-class model requires more class comparison, causing the table to be much larger. For all models, the proportion of students in each risk category increases from Class 1 to Class 3. In Model 1, BESS elevated risk was highest in Class 4 (M = .30) as compared to Class 1 (M=.01). In Model 5, BESS elevated risk was highest in Class 3 also (M = .16) versus Class 1 (M = 0), but each was lower than the Model 1 results. Model 1 classifications contain more students in Class 1 and fewer students in Class 3, so the proportion of students in risk categories is higher in Class 3.

The auxiliary (e) also allows us to determine if there are statistically significant differences in class composition by demographic variable across the three latent classes within a Model. For Model 1, classes differed in their composition based upon BESS elevated risk: Class

1 versus Class 3; Class 1 versus Class 4, Class 2 versus Class 4, and Class 2 versus 3. There were differences in class composition based upon ODR risk for each pair of classes and overall. For the remaining variables tested, Class 3 and Class 4 were not statistically significantly different in their class composition (e.g., special education status). For Model 5, the classes differed statistically significantly in their composition by class on all risk variables except Suspensions and Expulsions.

The results from the auxiliary (e) step can provide additional insight into comparing classification accuracy between the two Models. Model 5 has more statistically significant differences between the classes than Model 1, suggesting that Model 5 classes are more distinctly different on these variables than Model 1. In addition, Model 1 Class 3 and 4 did not have statistically significant differences except in Suspensions and Expulsions. This is perhaps why Classes 3 and 4 were combined in Model 5.

Table 24.

Means (SE) and chi-square statistics for distal outcomes, LE Model 1

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.01 (.01)	1 vs. 2	0.03
	2	.01 (.01)	1 vs. 3	1.99
	3	.00 (.00)	1 vs. 4	3.59
	4	.08 (.04)	2 vs. 3	0.93
			2 vs. 4	3.69
			3 vs. 4	4.32*
BESS basic risk	1	.06 (.01)	1 vs. 2	27.17**
	2	.25 (.03)	1 vs. 3	45.04**
	3	.41 (.05)	1 vs. 4	34.19**
	4	.48 (.07)	2 vs. 3	6.73**
			2 vs. 4	8.47**
			3 vs. 4	0.52
BESS elevated risk	1	.01 (.01)	1 vs. 2	3.03
	2	.03 (.01)	1 vs. 3	20.30**
	3	.19 (.04)	1 vs. 4	20.31**
	4	.30 (.07)	2 vs. 3	13.33**
			2 vs. 4	16.41**
			3 vs. 4	2.03
ODR risk	1	.04 (.01)	1 vs. 2	8.57**
	2	.12 (.03)	1 vs. 3	20.36**
	3	.25 (.04)	1 vs. 4	11.75**
	4	.25 (.06)	2 vs. 3	5.92*
			2 vs. 4	4.06*
			3 vs. 4	0.01

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether two classes are statistically significantly different in that category.

**p < .01, *p < .05.

Table 25.

Means (SE) and chi-square statistics for demographic variable composition, LE Model 1

Category	Class	M (SE)	Class Comparison	χ^2
Demographic Category				
Male	1	.42 (.03)	1 vs. 2	17.20**
	2	.62 (.04)	1 vs. 3	12.98**
	3	.63 (.05)	1 vs. 4	6.67*
	4	.62 (.07)	2 vs. 3	0.02
			2 vs. 4	0.01
			3 vs. 4	0.03
White	1	.84 (.02)	1 vs. 2	0.33
	2	.82 (.03)	1 vs. 3	0.30
	3	.82 (.04)	1 vs. 4	5.30*
	4	.69 (.07)	2 vs. 3	0.00
			2 vs. 4	2.01
			3 vs. 4	0.27
Special Education	1	.07 (.01)	1 vs. 2	1.55
	2	.10 (.02)	1 vs. 3	10.18**
	3	.21 (.04)	1 vs. 4	4.69*
	4	.19 (.06)	2 vs. 3	4.73*
			2 vs. 4	2.19
			3 vs. 4	0.08
Behavioral Supports	1	.08 (.02)	1 vs. 2	0.14
	2	.09 (.02)	1 vs. 3	2.31
	3	.14 (.04)	1 vs. 4	2.75
	4	.17 (.05)	2 vs. 3	1.31
			2 vs. 4	2.01
			3 vs. 4	0.27
Hispanic	1	.07 (.01)	1 vs. 2	4.09*
	2	.13 (.03)	1 vs. 3	0.07
	3	.08 (.03)	1 vs. 4	0
	4	.07 (.04)	2 vs. 3	1.72
			2 vs. 4	1.88
			3 vs. 4	0.84

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether two classes are statistically significantly different in that category.

**p < .01, *p < .05.

Table 26

Means (SE) and chi-square statistics for distal outcomes and demographic variable composition, LE Model 4

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.00 (.00)	1 vs. 2	3.11
	2	.01 (.00)	1 vs. 3	2.01
	3	.03 (.01)	2 vs. 3	5.12*
BESS basic risk	1	.03 (.02)	1 vs. 2	30.03**
	2	.16 (.02)	1 vs. 3	28.24**
	3	.40 (.04)	2 vs. 3	81.96**
BESS elevated risk	1	.00 (.00)	1 vs. 2	30.14**
	2	.02 (.01)	1 vs. 3	7.35**
	3	.20 (.03)	2 vs. 3	41.27**
ODR risk	1	.02 (.01)	1 vs. 2	21.76**
	2	.08 (.01)	1 vs. 3	9.33**
	3	.25 (.03)	2 vs. 3	41.07**
Demographic Category				
Male	1	.32 (.05)	1 vs. 2	4.32*
	2	.53 (.03)	1 vs. 3	16.38**
	3	.63 (.04)	2 vs. 3	28.09**
White	1	.86 (.03)	1 vs. 2	11.92**
	2	.86 (.02)	1 vs. 3	0.01
	3	.72 (.04)	2 vs. 3	8.29**
Special Education	1	.07 (.02)	1 vs. 2	1.81
	2	.11 (.02)	1 vs. 3	1.88
	3	.15 (.03)	2 vs. 3	5.35*
Behavioral Supports	1	.05 (.02)	1 vs. 2	7.30**
	2	.08 (.010	1 vs. 3	2.25
	3	.17 (.03)	2 vs. 3	12.73**
Hispanic	1	.06 (.02)	1 vs. 2	0.45
	2	.08 (.02)	1 vs. 3	0.63
	3	.10 (.02)	2 vs. 3	1.53

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether two classes are statistically significantly different in that category.

**p < .01, *p < .05.

Table 27.
Means (SE) and chi-square statistics for distal outcomes and demographic variable composition, LE Model 5

Category	Class	M (SE)	Class Comparison	χ^2
Distal Outcome				
Suspension or Expulsion	1	.00 (.00)	1 vs. 2	1.99
	2	.01 (.00)	1 vs. 3	5.10*
	3	.02 (.01)	2 vs. 3	2.28
BESS basic risk	1	.01 (.01)	1 vs. 2	35.51**
	2	.14 (.02)	1 vs. 3	109.61**
	3	.37 (.03)	2 vs. 3	34.28**
BESS elevated risk	1	.00 (.00)	1 vs. 2	5.85*
	2	.02 (.01)	1 vs. 3	42.61**
	3	.16 (.03)	2 vs. 3	30.78**
ODR risk	1	.02 (.01)	1 vs. 2	5.80*
	2	.07 (.01)	1 vs. 3	41.45**
	3	.22 (.03)	2 vs. 3	32.43**
Demographic Category				
Male	1	.30 (.05)	1 vs. 2	19.55**
	2	.54 (.03)	1 vs. 3	29.69**
	3	.61 (.03)	2 vs. 3	2.40
White	1	.86 (.01)	1 vs. 2	0.00
	2	.86 (.02)	1 vs. 3	5.47*
	3	.75 (.03)	2 vs. 3	9.21**
Special Education	1	.05 (.02)	1 vs. 2	3.05
	2	.10 (.02)	1 vs. 3	9.20*
	3	.12 (.73)	2 vs. 3	2.71
Behavioral Supports	1	.05 (.02)	1 vs. 2	1.26
	2	.08 (.02)	1 vs. 3	9.58**
	3	.15 (.03)	2 vs. 3	5.77*
Hispanic	1	.06 (.02)	1 vs. 2	1.1
	2	.09 (.02)	1 vs. 3	1.61
	3	.10 (.02)	2 vs. 3	0.12

Note: The mean represents the predicted proportion of the distal outcome or demographic category in the trajectory class based upon the auxiliary (e) step. The χ^2 measures whether two classes are statistically significantly different in that category.
**p < .01, *p < .05.

Based upon the more favorable fit statistics, class composition, descriptive statistics, and results from auxiliary (e), Model 5 is the preferred model. In the next section, characteristics of the class composition of Model 5 are explored.

Research Question 3. What are the characteristics of the students comprising latent classes of classroom behavior trajectories that emerge from the use of GM models, and in what ways do the latent classes vary in intercepts, slope, and variances?

Table 28 provides a demographic breakdown of student composition by Class for Model 5, and Figure 16 provides the graph of the latent class trajectories. Just like MS, Class 1 is characterized by students who are consistently academically engaged and respectful almost all the time and disruptive almost none of the time. Class 1 is disproportionately lower on demographic characteristics associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, Hispanic and ODRs. It is comprised of 30% Male, 86% White, 5% Black, and 6% Hispanic.

Table 28

Demographic distribution of Model 5 by class—LE

Category	All LE (n = 657)	Class 1 Optimal (n = 105)	Class 2 Average (n = 334)	Class 3 Lowest (n = 218)
Male	.52	.30	.54	.61
White	.82	.86	.86	.75
Black	.10	.05	.08	.15
Retained	.06	.01	.05	.11
Special Education	.11	.06	.10	.15
EBD	.01	.00	.01	.02
Behavior	.06	.01	.01	.15
Days Absent	6.49	5.55	6.32	7.22
Suspension or Expulsion	.01	.00	.01	.02
Brisk	.20	.01	.14	.38
Erisk	.06	.00	.02	.17
Hispanic	.09	.06	.09	.09
ODRRISK	.11	.02	.07	.22

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS,
ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 1 and Class b in Model 5.
Only nonempty change classes are reported.

On average, students in Class 1 have nearly perfect DBR-SIS scores that do not change much over time (see Appendix B, Table B3). The slopes for the fall were not statistically significant, and the slopes for the winter and spring were very small. Intercept variances at the within level for the fall and spring were not statistically significant. Residual variances were low but decreasing over time, ranging from .23 for fall, .10 for winter, and .04 for spring.

Class 2 is characterized by students who are not consistently academically engaged and respectful and are occasionally disruptive. Class 2 has a higher proportion of students compared

to Class 1 in demographic groups associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, Hispanic and ODRs. It is comprised of 54% Male, 86% White, 8% Black, and 7% Hispanic. Students in this class have high average DBR-SIS scores representing the targeted behavior around 90% of the time, resulting in scores near 27.0. Slopes were small, positive and statistically significant for all three data collection periods. Intercept variances at the within level were moderate and statistically significant. Residual variances were larger than for Class 1, ranging from 1.68 for fall, 1.44 for winter, and 1.08 for spring.

Class 3 is characterized by students who are highly variable in their classroom behavior, indicated by a wide spread of sample values around the average trajectory in Figure 16. Class 3 has a much higher proportion of students compared to Class 1 and 2 in demographic groups associated with elevated risk such as Male, Black, Retained, Special Education, EBD, Days Absent, Suspension or Expulsion, BESS risk, Hispanic and ODRs. It is comprised of 61% Male, 75% White, 15% Black, and 22% Hispanic. On average, students in this class have much initial status DBR-SIS score of 22.0, with upward shifts at the nodes. Slopes were positive and statistically significant for all three data collection periods, and the yearend predicted value was 24.5. These DBR-SIS scores indicate that on average, students in this class are exhibiting the targeted behavior from 70-80% of the time. Intercept variances at the within level were much larger than the other classes (e.g., fall intercept variance 15.07 compared to 0.02 for Class 1). Residual variances were very large compared to Class 1 and 2, ranging from 9.30 for fall, 9.70 for winter, and 6.48 for spring. Residual variances were highest in the winter and lowest in the spring. Using the DBR-SIS and Model 5, 34% of LE student are in class 3 while using the BESS, 20% of LE students are identified as at-risk.

Variability due to Teacher. The variances of the intercepts between teachers were not statistically significant. The variance (SE) in the intercept between teachers was 0.33 (0.38) for fall, 0.45 (0.64) for winter, and 0.38 (0.64) for spring. Using the intraclass correlation which compares between teacher variability to within, between, and residual variances, for Class 1, the proportion of variability between teachers is 53.2% in the fall, 76.2% in the winter, and 86.4% in the spring. For Class 2, 6.8% of the variability was between teachers in the fall, 10.7% in the winter, and 11.9% in the spring. For Class 3, 1.3% of the variability was between teachers in the fall, 2.1% in the winter, and 2.1% in the spring. For Classes 1 and 2, the amount of variability between teachers increased over time, but for Class 3, variability increased from fall to winter, but not from winter to spring.

Another way to look at the ICC involves examining the intercept variances in Model 4. Model 4 reflects a parameterization that constrains all intercept variances to equal across classes. The ICC for this comparison is calculated as the ratio of the between teacher variances to the within and between variance in the intercept. For LE, the between teacher variability increased from 34.7% in the fall, to 43.8% in the winter, and 40.2% in the spring. As mentioned previously, this phenomenon could be due to teacher familiarity with students or the positive effect of classroom management on producing more homogeneous classroom behavior patterns.

Looking further into patterns of observations at the teacher level can provide evidence to support the validity of assigning latent classes at the student level as compared to the teacher level. An examination of the class distribution of students by teacher provided some insights into whether modeling at the teacher level would make sense. Similar to what was done for the MS and UE grade groups, if 8 or more students within a teacher are assigned to a single class, this would support consideration of a teacher level model. Of the 67 teachers in the LE grade group

sample, eleven teachers had 8 or more students assigned to Class 2, and seven teachers had 8 or more assigned to Class 3. The distribution of the majority of teachers followed the overall distribution with the majority of students in Class 2, supporting the use of the student level class assignment.

Summary. Five model parameterizations were explored for the LE sample using GM modeling, testing progressive increases in the number of classes and relaxing model parameters to test for noninvariance. Once again, the Model 5 3-class model represented the best fit and most homogeneous set of classes. Class 1 is optimal and consistent, Class 2 has average behavior and is less consistent, and Class 3 represents poor, highly variable classroom behavior.

Comparison of MS, UE, and LE

The five model parameterizations provided a technique for modeling the level and variability across the grade groups and the classes. Consideration of these results across the grade levels can also provide information on the characteristics of the three classes. Figure 19 displays a comparison of the latent class composition for all models and all grade groups. Model 1, which represents a classification mechanism by mean level, with class invariant residual variances and no random effects, places more students in Class 1 for all grade groups. Model 2 provides for random effects, but not class varying random effects. This model classifies even more students in Class 1. Model 3 did not have admissible solutions so is not shown here. Model 4 provided for class varying residuals and Model 5 provide for class varying residuals and intercept random effects. In these models, students from the Model 1 Class 1 were classified into Classes 2 and 3 based upon variability. In each sample, Model 5 classifies the majority of students into Class 2, fewer students in Class 1, and more students in the group of students with more problematic behavior, which is Class 3.

Figure 19. Proportion of students assigned to each class by model.

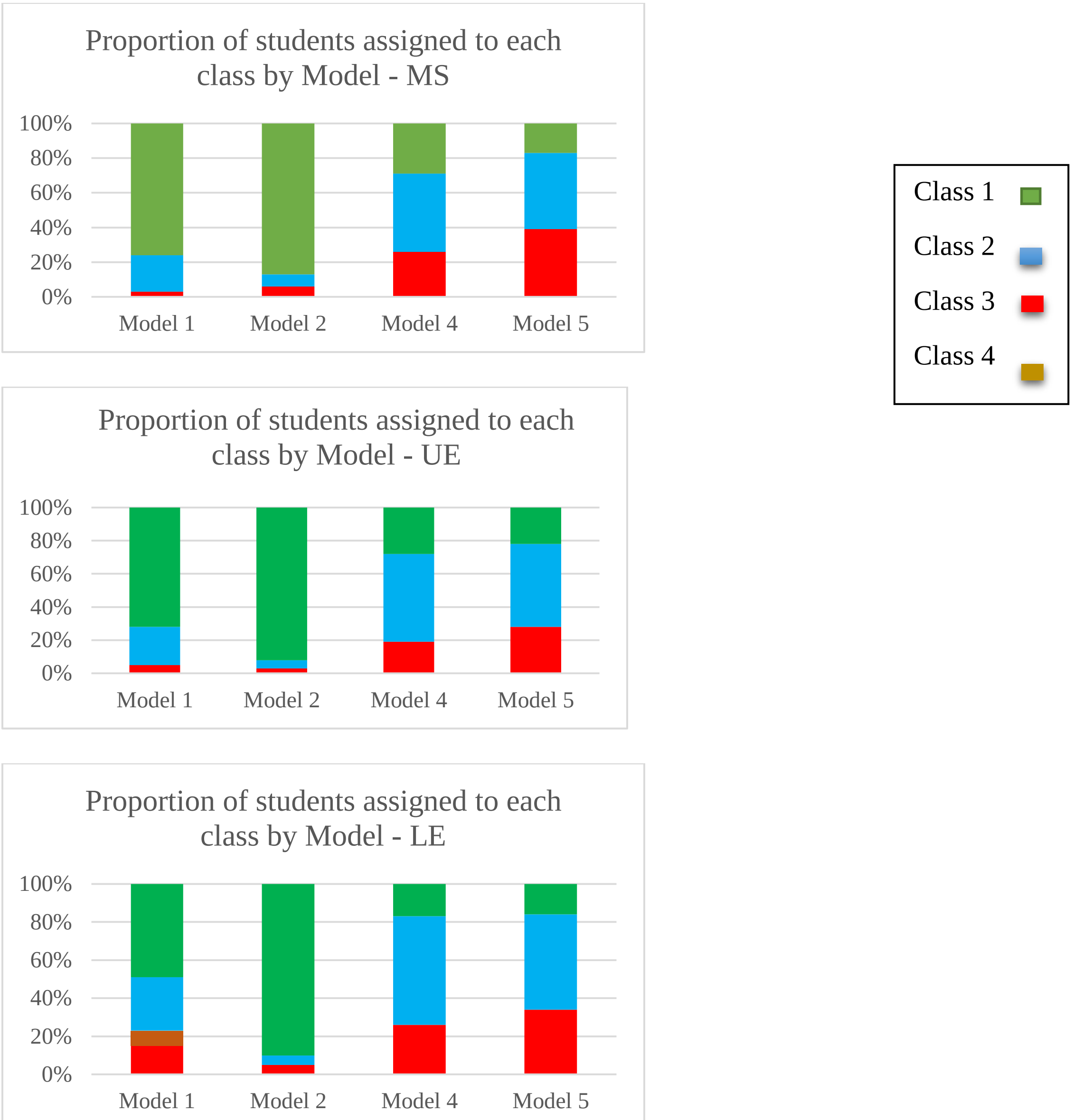


Figure 19. Proportion of students assigned to each class by model.MS=middle school, UE=upper elementary, LE=lower elementary. Red=problem behavior (Class 3 or 4), Blue=Class 2, average behavior, Green=Class 1, optimal behavior.

To explore the characteristics of these classes further, Figures 20 and 21 provides a set of histograms for the Model 5 for MS, UE, and LE. The proportion of students in risk categories increases across classes with the lowest percentage in Class 1 and the highest percentage in Class 3. Class 3 had a disproportionately large number of students who are male, black, Hispanic, disabled, have basic or elevated risk on the BESS, have had two or more ODRs, or have been suspended or expelled at least once. Class 3 also has a greater proportion of students who have been retained, and the average number of days absent is also the largest. For each grade group, Class 3 was the most distinct as measured by the chi-square test of mean differences from the auxiliary (e) test. The implications of these findings are explored more in chapter five from both a methodological and substantive perspective.

Figure 20. Proportion of students in demographic and behavioral risk categories.

Figure 20. Proportion of students in demographic and behavioral risk categories. Class 1=optimal class, Class 2=average classroom behavior, Class 3=poor and inconsistent classroom behavior.

Figure 21. Proportion of students in demographic and behavioral risk categories.

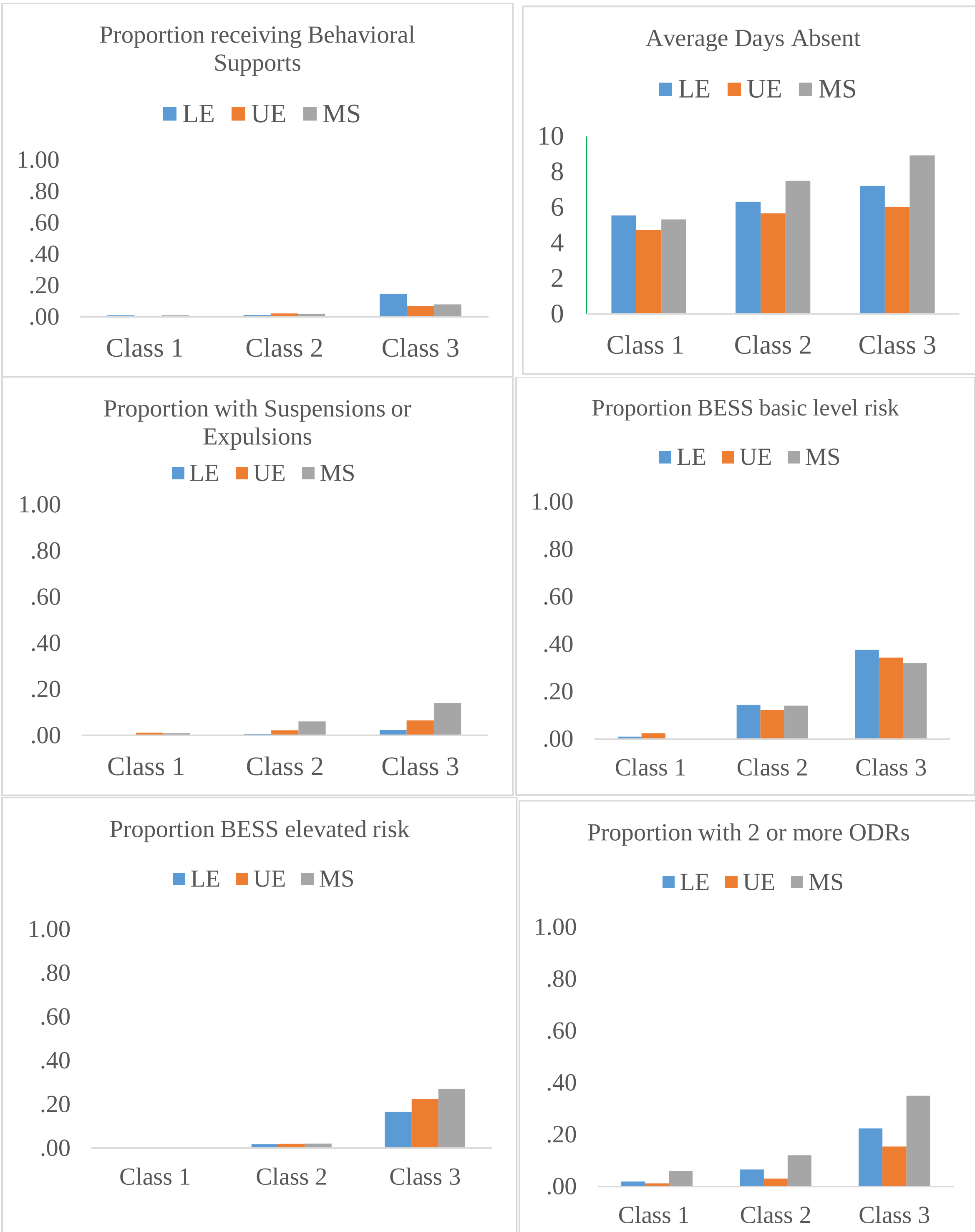


Figure 21. Proportion of students in demographic and behavioral risk categories. Class 1=optimal class, Class 2=average classroom behavior, Class 3=poor and inconsistent classroom behavior.

For each grade group, the variability in the intercepts for Class 1 was small and in most cases not statistically significant. The within teacher variability in the intercept was larger and statistically significant for Class 2, and substantially larger for Class 3. In substantive terms, a portion of the variability in student DBR-SIS scores is due to variation in initial status, e.g. mean levels, and this is more prominent in Classes 2 and 3. In all grade groups, the between teacher variability in the intercept was not statistically significant. Between teacher variability increased over time for all three grade groups, as can be seen in Figure 22.

Figure 22. Between Teacher Variability in the Intercept

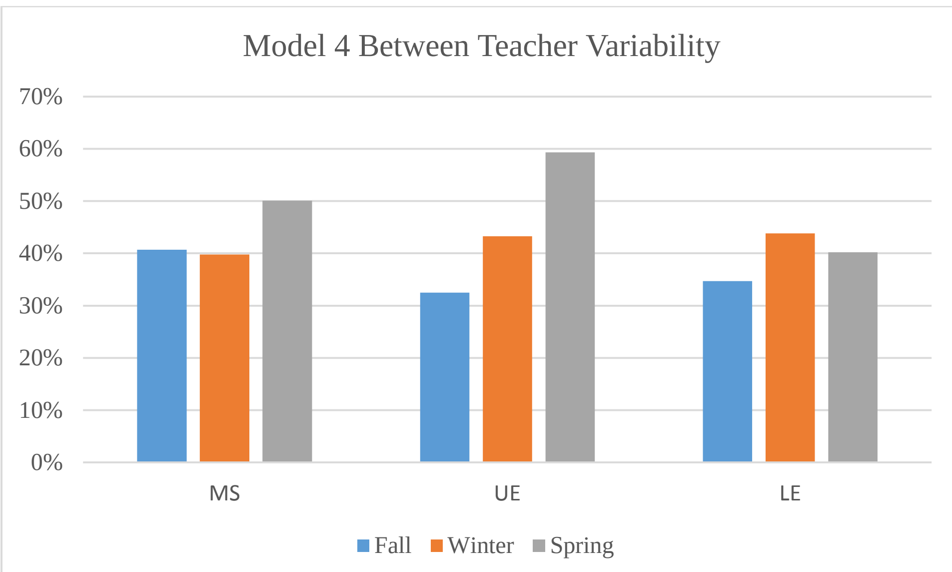


Figure 22. Histogram displays the proportion of intercept variability that is between teachers for Model 4 by grade group.

Summary

The Model 5 3-class parameterization was identified as the best fitting model for all three grade groups. Model 5 represented a latent class structure characterized by homogeneity within class and heterogeneity between classes. The proportions from Model 5 by class and grade level are displayed in Table 29.

Table 29.
Proportion of students by class and grade level in Model 5.

Class	LE	UE	MS
1	16%	22%	17%
2	50%	50%	44%
3	34%	28%	39%

The typology for all three grade groups is characterized by near perfect predicted scores for Class 1 for the entire year, scores lower by 2-3 points for Class 2, and scores lower by 7-8 points for Class 3. A comparison of Classes 1, 2 and 3 by grade group is shown in Figure 23. The intercepts identify a pattern of development across the grades for Classes 2 and 3 whereby there is a slight improvement in the initial status from LE to UE and from UE to MS. For comparative purposes, graphs for the 3-class models are provided in Appendix D for Models 1,2, and 4.

Figure 23. Comparison of Model 5 DBR-SIS predicted trajectories for LE, UE, and MS by Class.

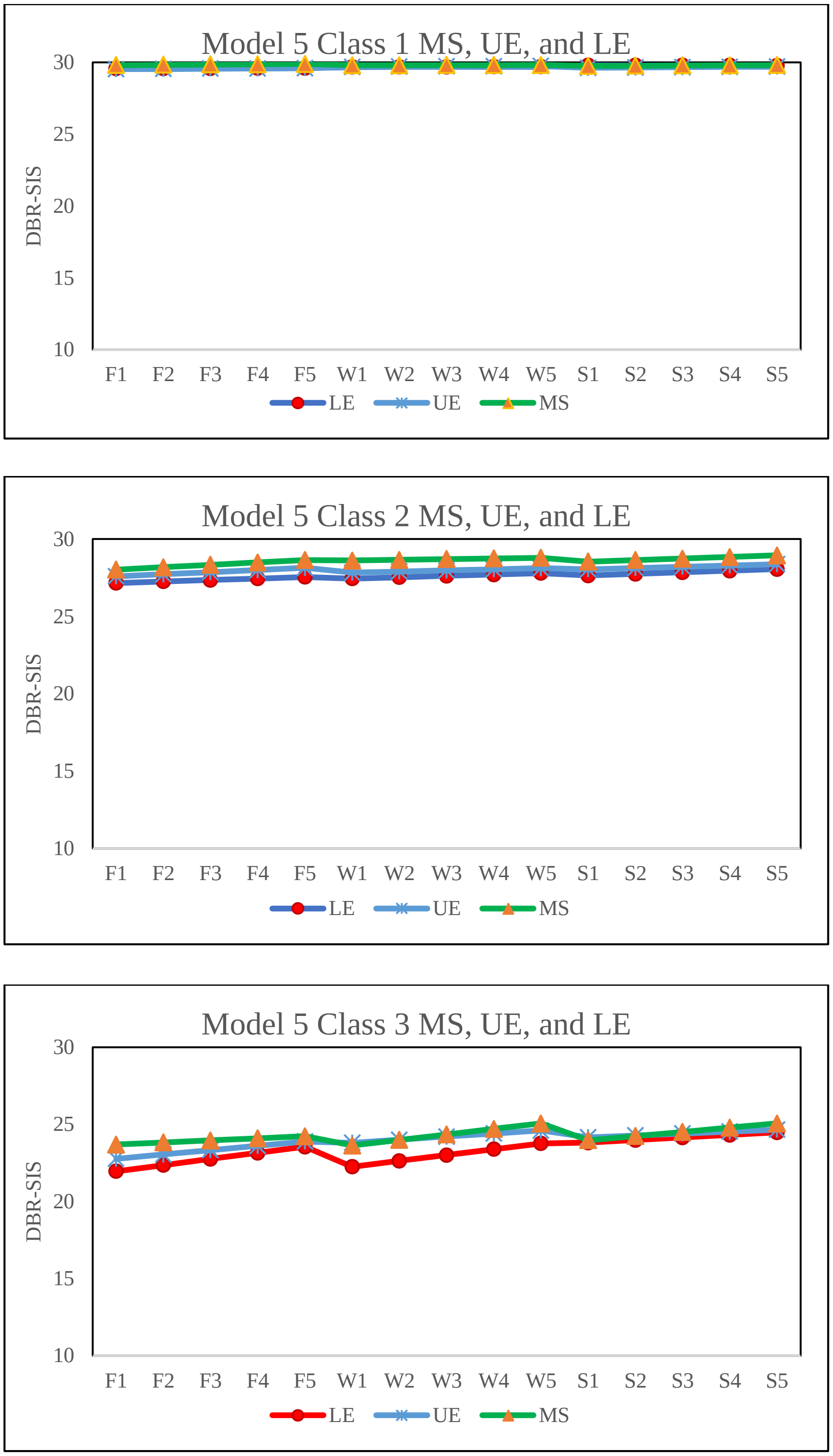


Figure 23. Comparison of DBR-SIS predicted trajectories over fall, winter and spring data collection points for LE, UE, and MS by class.

Chapter 5: Discussion

The goal of this research was to address two distinct but related questions, one methodological and one substantive. From a methodological perspective, this research used real data to investigate how varying decisions in multilevel GM modeling can result in differences in the composition of the latent classes and the growth curve characteristics. Using data from the DBR-SIS representing behavior measures for 1975 students in LE, UE, and MS grades provided a valuable context to investigate how these decisions affected means and variances in intercepts, slopes, random effects, and residuals which aided the development of meaningful interpretations and selection of models aligned with the substantive objective. From a substantive perspective, this research used these results to explore developmental patterns of student classroom behavior over one year and define a typology to describe these patterns.

The pattern of change and characteristics of students within each latent class were identified using a stepwise approach to modeling, taking great care to consider options as indicated by the output and prior research when making decisions. The selected final models were based upon a combination of methodological best practices and substantive theory, with a thorough examination of the mechanism that appeared to be driving the change in class assignment for students across models. The results were replicated across the three grade groups providing evidence in support of the 3-class models.

Research Question 1

When selecting the multilevel GM model that best represents the characteristics of a real data sample of student classroom behavior using the DBR-SIS, in what ways and to what extent

does noninvariance affect decision making, including application of enumeration indices, and resulting latent class characteristics?

The existence of noninvariance in the measures across the hypothesized latent sub-populations had a substantial impact on enumeration indices, decisions, and latent class characteristics. In all three grade groups, the level of fit followed the same pattern. The behavior of the enumeration indices and subsequent decisions are addressed first followed by the effects on the class composition by model.

Enumeration indices. For all grade groups, Model 1 which reflected invariance had the highest AIC, BIC and SABIC and Model 5 reflecting class varying residual variances had the lowest AIC, BIC, and SABIC, indicating Model 5 as the best fitting model. For all models, as the number of classes was increased, the information criteria dropped. This phenomenon is typical and observed in many prior studies (e.g., Morin et al., 2011). For all models, there appeared to be a threshold for class number after which models exhibited convergence problems or non-invertible matrices as more classes were added. In many but not all cases, the 3-class and 4-class models ran without incident, and a non-significant 4-class BLRT indicated that the fourth class did not improve model fit, pointing to the 3-class model as the best model. Morin et al. (2011) used two techniques to identify the preferred class enumeration: elbow plots (Petras & Masyn, 2010) and substantive interpretation. Elbow plots indicated that the information criteria leveled off after 5 classes. Further, Morin et al. (2011) selected a five class model over six or seven because of “at least one very small class including less than 1% of the students ($n <= 10$) and parameter estimates that were hard to interpret.” (p. 625). For the current study, these last techniques were not necessary because only a few models with more than 3 classes ran properly, and when they did, a non-significant BLRT indicated that the 4th class did not improve model fit.

Modeling decisions. The comparisons of Model 1 versus 5, 2 versus 5, and 4 versus 5 provide important insights into modeling decisions. Research in mixture modeling can sometimes reflect the choice to use a LCG model a priori rather than a GM model. The characteristics of the latent classes from the LCG model represent a system that classified students based more upon mean levels, because the variability assumed in the model remains constant across the classes. Using this system resulted in a large Class 1 and smaller proportion of students in classes with lower DBR scores. In contrast, the GM model can represent class noninvariance in random effects and the residual variances. Using this system resulted in a more even distribution across the three classes. Students are sorted based upon not just mean levels but means and variances. From a substantive perspective, GM modeling produces a classification system where students who have high and consistent DBR scores are assigned to one class while students who exhibit similar behavior but with less consistency are assigned to other classes.

The results of Models 4 and 5 were very similar, and the choice of a single model was not entirely clear. Model 4 represented class invariant intercept random effects and class varying residual variances. Model 5 represented both class varying intercept random effects and class varying residual variances. Some students with high average scores in Model 4 but with one or two low scores shifted from Class 2 to Class 3. In substantive terms, this means that students with more variability were shifted to either Class 2 or Class 3 in the change from Model 4 to 5. For Model 5, the average standard deviation of each of the three classes was lower than for the other models, suggesting more homogeneity of students within each class.

Intercept noninvariance. Model 3 for 3 classes resulted in an inadmissible solution from a non-positive definite latent variable matrix in all three grade groups. Other research including Ram and Grimm (2009) suggests simplifying models where there are convergence issues or

negative variances. Steps taken in the literature and also recommended on the Mplus discussion board typically recommends a reparameterization that includes fixing variances to 0. Interestingly, in this study, adding complexity to the model by modeling class noninvariant intercept random effects and class noninvariant residuals led to a model which converged and had admissible and interpretable solutions. An important finding then is a recommendation to consider more complex models, not just simpler models, in some cases when faced with models with convergence or matrix problems. For this research, visual examination of raw trajectories provided some information that variability in the DBR-SIS appeared very different for students who were consistently well behaved from those that were not consistent. Specifically, students who were assigned to Class 1 not only had higher means but almost no variability. Students assigned to Classes 2 and 3 had lower means but also more variability. This provided the needed insight into setting up the model, and model fit confirmed the appropriateness of the specification.

Residual noninvariance. The residual variance measures the variability in the difference between the observed and predicted values for the outcome variable. Typically, high residuals are an indication of poor model fit or a high degree of measurement error in the observed variable. In the current set of models, what is unusual is that the residual variance is different across the three classes and that this phenomena is repeated across the three grade groups. Although differences were not surprising given that the model was specified this way, the magnitude of the differences was surprising. There are a number of ways to interpret this. One is that the measure itself may vary in its reliability depending upon the behavior of the student. Although somewhat obvious, for students with good behavior, the measure is very reliable and also the student has virtually no variability in their behavior. For students with poor classroom

behavior, the measures is less reliable and students have a great deal of variability in their behavior. A second perspective is that the construct is actually quite different for students with optimal behavior versus those with poor classroom behavior. A hypothetical student with an average DBR-SIS of 15 is not likely to function at 50% engagement every day. Rather this student is likely to fluctuate between good days with high engagement and bad days with lower engagement, bringing the average down to 15, but with high variability. We expect these students arrive in the classroom carrying with them the effects from other aspects of their lives that manifest themselves in the classroom. So for these students, a low measure of engagement may equate to a high measure of some other social or emotional problem that has occurred outside the classroom and cannot be measured by the DBR-SIS. In this case, the residual variance picks up on this difference providing a measure of the variability not reflected in the linear model.

Mplus defaults. Another important finding addressed the use of Mplus defaults for the residual variance parameterization. The Mplus defaults provide for unique residuals for each point in time for LCG models or GM models. As additional parameters were added to the model, the model results were inadmissible due to a non-positive definite latent variable matrix. In addition, Model 2 reflected the Mplus defaults for intercept random effects for mixture modeling. Surprisingly, the results from Model 2 were erratic for all three grade groups and difficult to interpret. It is important to fully understand what is intended in using the Mplus defaults and ensure that these defaults appropriately reflect the nature of the data and research question. The flexibility of Mplus in converting to a simulation study provided an opportunity to test a theory that the problems with the latent variable matrix were due to cluster size, not model parameterization. Simulations demonstrated that increasing the number of clusters repaired the

problem, suggesting that the model was over-parameterized. In response to this, the residual variances were constrained in the LE and MS models to be invariant within the data collection periods of fall, winter and spring, and invariant across the year for UE. Models 4 and 5 went a step further and provided for noninvariance in the residual variances between classes.

Model convergence. When models did not converge or had inadmissible solutions, a number of different approaches were followed while remaining systematic in these procedures. For example, in the MS sample, the 4-class model for Model 1 did not converge due to the model estimation reaching a saddle point. Once the random starts were increased using the Starts command, this problem was resolved. In contrast, the Model 2 4-class model for MS also reached a saddle point, but increasing the random starts did not resolve this problem. As mentioned above, Model 3 results were inadmissible due to a non-positive definite latent variable matrix. Techniques such as increasing random starts, iterations, or changing the convergence criteria did not resolve the problem.

Model convergence and inadmissible solutions are problems that occur frequently in GM modeling, both in applied and simulation studies (Liu & Hancock, 2014). In the current study, model convergence was a consistent problem when increasing the number of classes from 3 to 4. Local solutions were also a problem in some cases, but increasing the number of random starts resolved most of these situations. The approaches taken here were systematic and consistent, supporting the modeling decisions. It is important to read the output carefully to ensure the model ran correctly. Mplus is a powerful software tool with many techniques embedded in the program that seems to fix itself, resulting in output and solutions along with non-invertible matrix warnings. Users must make sure they understand and can accept the warnings for the research question at hand.

Comparison of latent class composition. Figure 16 displays a comparison of the latent class composition for all models and all grade groups. Model 1, which represents a classification mechanism by mean level, with class invariant residual variances and no random effects, places more students in Class 1 for all grade groups. Model 2 provides for random effects, but not class varying random effects. This model classifies even more students in Class 1. Model 3 did not have admissible solutions so is not shown here. Model 4 provided for class varying residuals and Model 5 provide for class varying residuals and intercept random effects. In these models, students from the Model 1 Class 1 were classified into Class 1 or Class 2 based upon variability. In each sample, Model 5 classifies the majority of students into Class 2, fewer students into Class 1, and more students into Class 3.

Figure 16. Proportion of students assigned to each class by model.

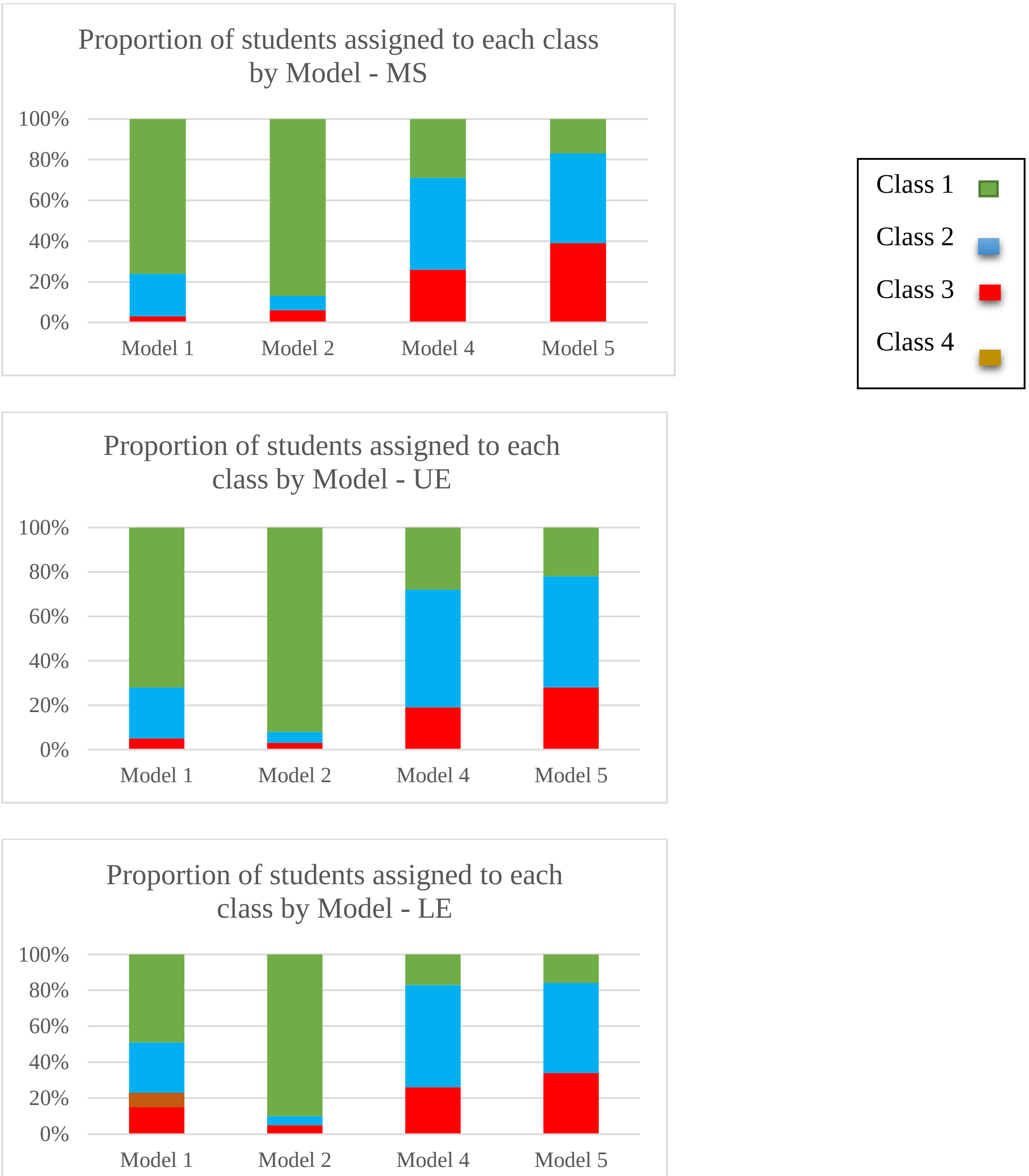


Figure 16. Proportion of students assigned to each class by model.MS=middle school, UE=upper elementary, LE=lower elementary. Red=problem behavior (Class 3 or 4), Blue=Class 2, average behavior, Green=Class 1, optimal behavior.

Latent class reassignment across models. The analysis of reassignment of students to different classes provided important insights on the nature of the different models. For most models, students who shifted from Class 1 to Class 2 were not representative of students with behavioral problems. Rather, they represented students with higher variability in the DBR-SIS. On the other hand, students who were reclassified from Class 1 to Class 3 and from Class 2 to Class 3 were more likely to be male, black, Hispanic, have elevated BESS scores and were at greater risk based upon ODRs. These results provided strong evidence for Model 5 and for the use of the DBR-SIS as a screener because the measure was able to differentiate students who are assessed with or who are at risk for behavioral problems using the BESS.

Research Question 2

To what extent does the trajectory class membership selected from research question 1 predict distal outcomes of student behavior measured using alternatives measures such the BESS risk classification, office disciplinary referrals (ODRs), and suspensions, providing validity evidence for the latent trajectory classes?

A crucial part of evaluating the latent classes in the Model 5 requires examination of criterion measures and other evidence to support the validity of the classes in identification of student classroom behavior trends. Outcome measures and demographic characteristics which have been shown in prior research to be associated with student behavior considered at risk provide the needed criterion measures by which to judge the reasonability of the latent classes. For example, prior research suggests classes with less engaged, more disruptive and disrespectful students would have a greater proportion of students who are male, minority, or disabled. In addition to these demographic variables, classroom behavior problems are also associated with higher BESS T-scores, where scores over 60 indicate that the student would be identified as at

risk for behavioral and emotional problems. In addition, more ODRs, more suspensions and more expulsions are clear indicators that the student behavior is problematic enough that it is causing disruptions to their school attendance. Using descriptive statistics and predicted proportions, this study found for all grade groups, Class 1 had a lower proportion of students in the higher risk demographic categories. In nearly all cases, this increased for Class 2 and was highest for Class 3. This indicates that Model 5 using the DBR-SIS sorted students into low, medium and higher classroom behavior risk classification successfully. In addition, the per-student average number of days absent is lowest for the students in Class 1 and highest for the students in Class 3 a pattern that exists for all three grade groups.

Using the auxiliary (e) option in Mplus provided additional support for Model 5 showing that Class 1 was significantly different from Class 3 in all grade groups and all risk categories. The Class 1 predicted proportions in each category were also lower than Classes 2 or 3. In many cases, the three classes were all significantly different. The only outlier in this pattern are Hispanic students in UE. However, this outlier is inconclusive because the sample size for Hispanic students was small. In summary, these results provide stronger evidence to support the validity of the classification system for identification of patterns of behavior which are connected to other criterion measures.

Research Question 3

What are the characteristics of the students comprising latent classes of classroom behavior trajectories that emerge from the use of GM models, and in what ways do the latent classes vary in intercepts, slope, and variances?

Description of behavior typologies. A summary of the findings from this study provides a description of the three latent trajectory classes from Model 5. For each grade group, Class 1 is

the smallest class, with students characterized by consistently good behavior at or near perfect scores for the DBR-SIS measures. Class 2 consists of approximately half of the students displaying classroom behavior that is conducive to learning approximately 90% of the time with a slight improvement in behavior across the year. Class 3 consists of approximately 1/3 of the students and represents students who may be at risk for behavioral problems in the classroom. On average, they exhibit behavior that is conducive to classroom learning only about 80% of the time and experience a slight improvement in their behavior across the year. The most salient feature of students in Class 3 is that their classroom behavior is highly inconsistent from day to day.

The proportions by class and grade level were similar to the typologies reported in DiStefano and Kamphaus (2006) based upon a cluster analysis of the BASC TRS-C measure on a sample of children ages 6-11. In that study, the researchers identified three distinct groupings. An optimal group comprised 36% of the students, a group defined as having typical adjustment comprised 42%, and 22% were identified as functionally impaired. The current study contributes a new finding in that the lowest class does not reflect only the most severe behavior cases; rather, this finding supports the policy of screening students to identify more than just the functionally impaired for possible support and intervention. This research suggests that students whose behavior is more erratic should be given a second look as this may be an indication of need for some type of support or intervention.

Implications

The use of real data from this classroom behavior study provided an opportunity to study the effect of systematically relaxing invariance assumptions on model characteristics, class enumeration, class composition and growth parameter estimates, while using real contextual

distal variables such as the number of ODRs to validate latent class profiles. These results have a number of methodological and substantive implications, of which the methodological implications are presented first.

Methodological Implications

1. Consider the nature of the data when identifying candidate model specifications for growth mixture modeling. If within subject variability is high for some and low for others, then more complex noninvariance models should be investigated. Consider a variety of parameterizations for residuals including over time noninvariance and between class noninvariance.
2. Start the modeling process with a LCG model, and successively increase the number of classes. Follow this with increasingly complex models to reflex increasing levels of noninvariance to identify suitable candidate models. The recommendations in Ram and Grimm (2009) provided an effective technique for examining a variety of GM models. Starting with a LCG model with one class, successively increasing the number of classes, systematically relaxing the invariance assumptions and examining model fit and characteristics provided an effective technique for identification of a suitable model. Remarkably, the learnings gained from completing this stepwise process added a great deal to this research. For example, if the goal is to accurately identify the students with the most severe problem behavior, the LCG model worked well. However, identification of the most severe problem behaviors is not a challenging task for most teachers, so from an applied perspective, the LCG model was inferior. If the goal is to provide a typology based upon student classroom behavior patterns and screen a greater proportion of

students who are exhibiting behavior not conducive to learning, the GM models were superior.

3. Compare class characteristics across models to inform model selection and strengthen validity evidence. This study illustrates the substantial reclassification of subjects leading to different substantive interpretations resulting from using different invariance assumption in growth mixture modeling building upon prior research in Morin et al. (2011) using GM modeling stepwise approach from Ram and Grimm (2009). A comparison of the methodological results from Morin to this study illuminates the importance of looking at subject reclassification. Similar to Morin, this study found that the model fit, shape of the trajectories, proportion of subjects assigned to each class, and the substantive interpretation varied widely from the most restrictive LCG model to the least restrictive GM model. Different from Morin, this study also provided a comparison of descriptive statistics by class across models, distribution of subjects by demographic variables by class across models, and comparison of model predicted versus sample trajectories to inform the model selection decision. These comparisons shed the light needed to recognize that students who have average to problematic classroom behavior differ not just by mean levels but they differ substantially in their variability. These findings illustrate that a student with an average DBR-SIS of 15 may not be at 50% engagement all the time, but rather varies between high and low levels of engagement on a day to day basis. These differences in the measures represent dramatically different substantive interpretations and implications for interventions in practice.

4. When a model with many constraints including invariance constraints does not run properly, do not stop and return to a model with more invariance or where parameters are

constrained to zero. Modeling challenges may mean that a more complex model with more variability reflected in the parameterization is needed. In this study, the Model 3 parameterization reflected class varying intercept random effects, but class invariant residual variances. These results suggest that between class heterogeneity was substantial enough to render the class invariant residual variances ineffective to the extent that the models produced non-positive definite matrices and no solutions. Rather than stopping here, this research investigated more complex models which turned out to be more appropriate for the data and research questions.

Substantive Implications

1. The identified models reflected a variety of student classroom behavior patterns. The final selected models classify students based upon both their means and variability in their behavior measures. This represents the first time that reported research has identified a student behavior typology that considers variability in student behavior.
2. A developmental typology of student classroom behavior change over time is best described as varying in level and consistency. Minimal linear change occurred across a one year time span.
3. The distribution by latent class developed in this research creates a typology that compares favorably with prior research, but the DBR-SIS measures are simpler to use and less expensive than the BASC TRS-C.
4. Several prior studies identified a small proportion of students in the class with severe behavior problems. If identification of these most severe cases is desired, the LCG model results work well, but if the desire is to screen in such a way to identify more students who need behavioral supports, the GM modeling approach is superior. This is the case for

the DBR-SIS because the variability is minimal among students with good behavior and variability is high among students with behavior that is problematic.

5. Researchers and practitioners who use single case design to evaluate response to intervention (RTI) may want to rethink their protocols when it includes a requirement for student behavior to stabilize before administering an intervention. Whether a single case design is used in research or practice, high variability can confound the protocol. Standards for research and practice suggest waiting for the pattern to stabilize, but for some students, this may never be observed. This research shows that for some students, their behavior is characterized by its inconsistency, inconsistency that interferes with their ability to learn and participate in school. Research in single case design would benefit from greater emphasis on interventions that reduce variability (Horner, Swaminathan, Sugai, & Smolkowski, 2012; Riley-Tilman & Burns, 2009). Standards for practice may need to reflect protocols for students who exhibit a high amount of variability. In addition, standards for research using effect sizes in single case research should identify measures of variability to quantify effectiveness of interventions. Related to this, future research into behavior interventions needs to explore interventions that help students to be more consistently engaged, respectful and not disruptive in school.
6. The results from this study provide a new approach to identifying optimal cut scores to identify students at risk of failure in school due to behavioral challenges. A conventional approach uses receiver operating curves (ROC) (Fawcett, 2006; Johnson et al, in press). ROC analyses identify cut scores for a measure based upon how it compares to a criterion measure. Johnson used a ROC analyses to identify cut scores for the DBR-SIS by comparing scores to a true measure of risk, the BESS scores. GM modeling does not

require the use of a criterion measure to identify classifications of risk; rather, it uses patterns in the outcome measure to develop classifications. Criterion measures were not used to identify classes in the current study; the BESS and ODRs provide evidence of the validity of the classifications in identification of risk level. Further, ROC studies do not have as much flexibility to reflect variability in a measure as GM modeling. ROC studies consider variability through the use of confidence intervals which use a pooled variance. In cases such as the DBR-SIS where the within student variance is high, the pooled variance results reduces the information considered in the analysis and could result in reduced accuracy.

Limitations and Recommendation for Future Research

The methodological results provide an important example of how differences in invariance assumptions can affect models and interpretations, but unlike a simulation study, findings with regard to model selection, model fit, and parameters may not be generalizable. Even though current research is robust and ongoing regarding the efficacy of techniques to identify best models and extract the correct number of classes using different information criteria and likelihood based tests (Asparouhov & Muthén, 2013; Liu & Hancock, 2014; Nylund, Asparouhov, & Muthén, 2007), it is not known how accurate these criteria were using the DBR-SIS data and the current sample. In addition, the results of class reassignment apply uniquely to the current research using a measure that has different levels of variance across the distribution. The distribution of the variance is not typically known for a measure prior to embarking on a study, but this is clearly something that needs to be considered prior to making decisions on model selection.

This research contributes to the literature both methodologically and substantively, and it has supported the discovery of many areas for future methodological research for GM modeling. A number of simulation studies are suggested by the work completed here. Future simulation research should consider the impact of restricting error variances to homoscedastic across classes or across time when using growth mixture modeling over subpopulations when in fact they vary for both single level and multilevel samples. The current study identified the 3-class model as the superior model for all three grade groups. Future simulation studies should investigate whether 3 class models are identified when more than 3 subpopulations are generated (Enders & Tofighi, 2009). The current study used a single piecewise linear growth model to represent change. Future simulation studies should investigate the impact of nonlinear growth on the effectiveness of information criteria in class identification and class enumeration. Further, in the current study, Model 3 which represented class varying intercept random effects but class invariant residuals was not convergent. Future research should investigate whether this is always the case when residuals are class noninvariant. The current study contained measures that were highly variable, but on average did not change much across time. These slopes were represented by fixed effects in the models. For all the simulation studies mentioned above, and for applied studies, future research should investigate the impact of class varying slope random effects on class enumeration. The models used in this study did not use any Level 2 (teacher level) covariates. This decision greatly simplified the models, but also limited the scope of the research. Future multilevel GM modeling simulation studies should include Level 2 covariates to study the impact on enumeration, model fit, and model selection. Finally, the model specification did not allow intercepts to vary across class at Level 2, the between teacher level. Future research should focus on examining the latent factor approach to estimating class varying random effects.

The current research also had limitations with respect to classroom behavior research. Typologies based upon diagnostic rating scales such as the BASC TRS-C provide more information on the nature of student social, emotional and behavioral challenges than what we are able to determine using the DBR-SIS. In addition, the current research used a composite of the AE, RS, and DB single item scales. Future research should consider whether latent class trajectories and mixtures vary when running models using the individual scales. Future research could study the same population using GM modeling but with the BESS score or with both the BESS and the DBR-SIS. Future research could also consider a comparison of the use of receiver operating curves to GM models to identify cut points to screen students who may need additional behavioral supports. GM modeling may provide additional insights because it can take into consideration both class varying mean levels and class varying differences in variability.

The results from this study apply to many areas of substantive interest in behavior research and practice, single case design research and practice, and methodological areas in growth mixture modeling. While some other studies have touched upon the topic of noninvariance in growth mixture modeling, this study has accomplished the task of using this sophisticated modeling technique to add to our knowledge of student classroom behavior. Even though some may suggest that growth mixture modeling is exploratory, completion of this study has revealed characteristics of the modeling process and the population of students that otherwise would not have been discovered. This study may take us one step further in helping students with behavior problems learn to moderate their behavior and achieve success in school.

References

American Educational Research Association Annual Meeting (2015). Retrieved from http://www.aera.net/EventsMeetings/AnnualMeeting/PreviousAnnualMeetings/2015AnnualMeeting/tabid/15930/Default.aspx.

Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Department of Psychiatry, University of Vermont.

Akaike, H. (1977). On the entropy maximization principle. In P. R. Krishnaiah (Ed.) *Applications of Statistics: Proceedings of the Symposium Held at Wright State University, Dayton, Ohio, 27-41*.

Asparouhov, T., & Muthén, B. (2013). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. *Mplus Web Notes, No. 15*.

Asparouhov, T., & Muthén, B. (2014a). Structural equation models and mixture models with continuous non-normal skewed distributions. *Mplus Web Notes: No. 19.*

Asparouhov, T., & Muthén, B. (2014b). Using Mplus individual residual plots for diagnostics and model evaluation in SEM. *Mplus Web Notes: No. 20.*

Bailey, K. D. (1994). *Typologies and taxonomies: an introduction to classification techniques* (Vol. 102). Thousand Oaks: Sage.

Bauer, D. J. (2008). Observation on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42(*4), 757-786.

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

Bollen, K. A., Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, N.J: Wiley-Interscience.

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2003). The normative development of child and adolescent problem behavior. *Journal of Abnormal Psychology, 112,* 179–192.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* Guilford Publications.

Brown, C. H., Wang, W., Kellam, S. G., Muthén, B. O., Petras, H., Toyinbo, P., ... & Group, M. (2008). Methods for testing theory and evaluating impact in randomized field trials: Intent-to-treat analyses for integrating the perspectives of person, place, and time. *Drug and alcohol dependence, 95*, S74-S104.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin,* 105(3), 456-466.

Chafouleas, S. (2011). Direct behavior rating: a review of the issues and research in its development. *Education and Treatment of Children, 34,* 575-591.

Chafouleas, S. M., Sanetti, L. M., Kilgus, S. P., & Maggin, D. M. (2012). Evaluating sensitivity to behavioral change using direct behavior rating single-item scales. *Exceptional Children, 78*(4), 491-505.

Chafouleas, S. M., Volpe, R. J., Gresham, F. M., & Cook, C. R. (2010). School-based behavioral assessment within problem-solving models: Current status and future directions. *School Psychology Review,* 39(3), 343-349.

Chan, D., Ramey, S., Ramey, C., & Schmitt, N. (2010). Modeling intraindividual changes in children’s social skills at home and at school: A multivariate latent growth approach to

understanding between-settings differences in children's social skill development. *Multivariate Behavioral Research*, *35*(3), 365-396.

Chen, Q., Kwok, O., Luo, W., & Willson, V. L. (2010). The impact of ignoring a level of nesting structure in multilevel growth mixture models: A monte carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(4), 570-589.

Christ, T. J., Nelson, P. M., Van Norman, E. R., Chafouleas, S. M., & Riley-Tillman, T. C. (2014). Direct behavior rating: An evaluation of time-series interpretations as consequential validity. *School Psychology Quarterly*, *29*(2), 157-170.

Clark, S. L., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis.* Retrieved from https://www.statmodel.com/download/relatinglca.pdf.

Crocker, L. & Algina, J. (2008). *Introduction to Classical and Modern Test Theory.* Matson, OH: Cengage Learning.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*. Series B (methodological), 1-38.

Dever, B. V., Dowdy, E., Raines, T. C., & Carnazzo, K. (2015). Stability and change of behavioral and emotional screening scores. *Psychology in the Schools*. Advance online publication.

Distefano, C. & Kamphaus, R. W. (2006). Investigating subtypes of child development: A comparison of cluster analysis and latent class cluster analysis in typology creation. *Educational and Psychological Measurement, 66(*3), 778-794.

DiStefano, C., & Kamphaus, R. W. (2008). Latent growth curve modeling of child behavior in elementary school. *Research in the Schools*, *15*(1), 27-37.

DiStefano, C., Kamphaus, R. W., Horne, A. M., & Winsor, A. P. (2003). Behavioral adjustment in the U.S. elementary school: Cross-validation of a person-oriented typology of risk. *Journal of Psychoeducational Assessment, 21*(4), 338-357.

DiStefano, C. A., Kamphaus, R. W., & Mîndrila, D. L. (2010). A typology of teacher-rated child behavior: Revisiting subgroups over 10 years later. *School Psychology Quarterly, 25*(3), 152-163.

Dowdy, E., Nylund-Gibson, K., Felix, E. D., Morovati, D., Carnazzo, K. W., & Dever, B. V. (2014). Long-term stability of screening for behavioral and emotional risk. *Educational and Psychological Measurement, 74*(3), 453-472.

Elliott, S. & Gresham, F. (2007). *Social Skills Improvement System Performance Guides: Technical Supplement*. Minneapolis, MN: NCS Pearson, Inc.

Enders, C. K. (2010). Applied missing data analysis. Guilford Publications.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 430-457.

Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(1), 75-95.

Epstein, M., Atkins, M., Cullinan, D., Kutash, K., & Weaver, K. (2008). *Reducing behavior problems in the elementary school classroom.* National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/behavior_pg_092308.pdf.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874.

Finch, W. H. & French, B. F. (2014). Multilevel latent class analysis: Parametric and nonparametric models. *The Journal of Experimental Education, 82*(3), 307-333.

Gandomi, A & Haider, M. (2015). Beyond the hype: Big data concepts, methods and analytics. *International Journal of Information Management*, 35, 137-144.

Georges, A., Brooks-Gunn, J., & Malone, L. M. (2012). Links between young Children's behavior and achievement: The role of social class and classroom composition. *American Behavioral Scientist*, *56*(7), 961-990.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*(2), 117-135.

Grimm, K. J., Ram, N., & Estabrook, R. (2010). Nonlinear Structured Growth Mixture Models in M plus and OpenMx. *Multivariate behavioral research*, *45*(6), 887-909.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Measurement methods for the social sciences series, Vol. 2).

Hedeker, D., & Mermelstein, R. J. (2007). Mixed-effects regression models with heterogeneous variance: Analyzing ecological momentary assessment (EMA) data of smoking. In Little, T. D., Bovaird, J. A., & Card, N. A. (Eds.). *Modeling contextual effects in longitudinal studies*, 183-206, Routledge.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*(1), 36-53.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin*, *111*(1), 127-155.

Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations for the systematic analysis and use of single-case research. *Education and Treatment of Children, 35*(2), 269-290.

Huberty, C. J., DiStefano, C., & Kamphaus, R. W. (1997). Behavioral clustering of school children. *Multivariate Behavioral Research, 32*(2), 105-134.

Johnson, A. H., Miller, F. G., Chafouleas, S. M., Welsh, M. E., Riley-Tillman, T. C., & Fabiano, G. A. (in press). Evaluating the technical adequacy of DBR-SIS in tri-annual behavioral screening: A multisite investigation. *Journal of School Psychology.*

Johnson, A. H., Miller, F., Chafouleas, S. M. & Kooken, J. (2011). Direct behavior rating: Updates and current research directions. *The Connecticut School Psychologist.*

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Kamphaus, R. W., Huberty, C., DiStefano, C., & Petoskey, M. (1997). A typology of teacher-rated child behavior for a national U.S. sample. *Journal of Abnormal Child Psychology, 25*(6), 453-463.

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 Behavioral and Emotional Screening System.* Circle Pines, MN: Pearson.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* Oxford University Press.

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: Externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology, 28(2),* 161-179.

Kellam, S. G., Wang, W., Mackenzie, A. C., Brown, C. H., Ompad, D. C., Or, F., ... & Windham, A. (2014). The impact of the Good Behavior Game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Prevention Science*, 15(1), 6-18.

Kim, S. (2014). Determining the number of latent classes in single- and multiphase growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 263-279.

Kim, S., Orpinas, P., Martin, R., Horne, A. M., Sullivan, T. N., & Hall, D. B. (2010). A typology of behavioral adjustment in ethnically diverse middle school students. *Journal of Psychoeducational Assessment*, *28*(6), 524-535.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Kooken, J. (2014). *Model Selection for Data that Lack Normality.* (Unpublished doctoral comprehensive paper). University of Connecticut, Storrs, CT.

Kooken, J., Welsh, M. E., McCoach, D. B., Miller, F. G., Chafouleas, S. M., Riley-Tilman, T. C., and Fabiano, G. (2015). *Test Order in Teacher-Rated Behavior Assessments: Is Counterbalancing Necessary?* Manuscript under review.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). *Single-case intervention research design standards. Remedial and Special Education*, *34(1),* 26-38.

Kreuter, F. & Muthén, B. (2008). Analyzing criminal trajectory profiles: Bridging multilevel and group-based approaches using growth mixture modeling. *Journal of Quantitative Criminology*, *24*, 1-31.

Kuntsche, E., Otten, R., & Labhart, F. (2015). Identifying risky drinking patterns over the course of Saturday evenings: An event-level study. *Psychology of Addictive Behavior*, 29(3), 744-752.

Lane, K. L., Carter, E. W., Pierson, M. R., & Glaeser, B. C. (2006). Academic, social, and behavioral characteristics of high school students with emotional disturbances or learning disabilities. *Journal of Emotional and Behavioral Disorders, 14*(2), 108-117.

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

Li, X. & Hedeker, D. (2012). A three-level mixed-effects location scale model with an application to ecological momentary assessment data. *Statistics in Medicine*, 31, 3192-3210.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *American Statistical Association*, 83(404), 1198-1202.

Liu, M. & Hancock, G. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement*, 74(4), 557-584.\

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778.

Marsh, H. W. & Hau, K-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology, 32*, 151-170.

May, S., Ard, W. I., Rodd, A., W., Horner, R. H. Glasgow, A., Sugai, G. & Sprague, J. (2003). *School-Wide Information System.* University of Oregon, Eugene: Educational and Community Supports.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*(4), 471-490.

McCoach, D. (2010). Hierarchical linear modeling. In Hancock, G. (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 123-140). New York: Routledge.

McIntosh, K., Frank, J. L., & Spaulding, S. A. (2010). Establishing research-based trajectories of office discipline referrals for individual students. *School Psychology Review*, *39*(3), 380-394.

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

Miller, F. G., Cohen, D., Chafouleas, S. M. Riley-Tilman, T. C., Welsh, M. E., & Fabiano, G. A. A comparison of measures to screen for social, emotional and behavioral risk. *School Psychology Quarterly, 30* (2), 184-196.

Morin, A. J. S., Maïano, C., Nagengast, B., Marsh, H. W., Morizot, J., & Janosz, M. (2011). General growth mixture analysis of adolescents' developmental trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(4), 613-648.

Morin, A. J. S., Maïano, C., Marsh, H. W., Nagengast, B., & Janosz, M. (2013). School life and adolescents' self-esteem trajectories. *Child Development*, *84* (6), 1967-1988.

Muthén, B. O. (2001). Second generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In A. Sayer & L. Collins (Eds.), *New methods for the analysis of change* (pp.291-322). Washington, DC: American Psychological Association.

Muthén, B. O. (2004). Latent variable analysis. In D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, (pp. 345-368). Thousand Oaks: Sage Publication, Inc.

Muthén, B. O., & Asparouhov, T. (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 172*(3), 639-657.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*(6), 882-891.

Muthén, L. K., & Muthén, B. O. (1998-2014). *Mplus User's Guide. Seventh Edition.* Los Angeles: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM Algorithm. *Biometrics*, *55*(2), 463-469.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, *4*(2), 139-157.

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development*, *70*(5), 1181-1196.

Netemeyer, R. G., Bearden, W. O., Sharma, S. (2003). *Scaling Procedures: Issues and Applications.* Thousand Oaks: Sage Publications.

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 et seq. (West 2003).

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*(4), 535-569.

Owen, J., Adelson, J., Budge, S., Wampold, B., Kopta, M., Minami, T., & Miller, S. (2015). Trajectories of change in psychotherapy. *Journal of Clinical Psychology*, *71*(9), 817-827.

Palardy, G. J. & Vermunt, J. K. (2010). Multilevel growth mixture models for classifying groups. *Journal of Educational and Behavioral Statistics, 35*(5), 532-565.

Pas, E. T., Bradshaw, C. P. & Mitchell, M. M. (2011). Examining the validity of office discipline referrals as an indicator of student behavior problems. *Psychology in the Schools,* 48(6), 541-555.

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

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85-112.

Peugh, J., & Fan, X. (2015). Enumeration index performance in generalized growth mixture models: A Monte Carlo test of Muthén's (2003) hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(1), 115-131.

Podesta, J., Pritzker, P., Moniz, E., Holdren, J., & Zients, J. (2014). *Big data: seizing opportunities, preserving values.* Executive Office of the President, The White House Washington Study.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*(6), 565-576.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks: Sage Publications.

Reynolds, C.R. and Kamphaus, R.W. (2006) *BASC-2: Behavior Assessment System for Children, Second Edition.* Upper Saddle River, NJ: Pearson Education, Inc.

Reynolds, M. R., Sander, J. B., & Irvin, M. J. (2010). Latent curve modeling of internalizing behaviors and interpersonal skills through elementary school. *School Psychology Quarterly, 25(*4), 189-201.

Riley-Tillman, T. Chris, and Burns, Matthew K. (2009). *Evaluating Educational Interventions: Single-Case Design for Measuring Response to Intervention*. New York, NY, USA: Guilford Press, 2009. ProQuest ebrary. Web. 4 October 2015.

Schaeffer, C. M., Petras, H., Ialongo, N., Masyn, K. E., Hubbard, S., Poduska, J., & Kellam, S. (2006). A comparison of girls' and boys' aggressive-disruptive behavior trajectories across elementary school: prediction to young adult antisocial outcomes. *Journal of Consulting and Clinical Psychology, 74*(3), 500-510.

Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. S. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology, 39*(2), 189-200.

Simmons, D. C., Kim, M., Kwok, O. M., Coyne, M. D., Simmons, L. E., Oslund, E., ... & Rawlinson, D. A. (2013). Examining the effects of linking student performance and progression in a Tier 2 kindergarten reading intervention. *Journal of Learning Disabilities, 48*(3) 255-270.

SPSS (2009). *PASW Statistics for Windows,* Version 18.0. Chicago: SPSS Inc.

Stapleton, L. M. (2006). Using multilevel structural equation modeling techniques with complex sample data. *Structural equation modeling: A second course,* 345-383. United States of America: Information Age Publishing Inc.

Stapleton, L. (2013). Multilevel structural equation modeling with complex sample data. *Structural equation modeling: A second course* (2nd ed., pp. 521–562). United States of America: Information Age Publishing Inc.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods, 3,* 4-70.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486-492.

Vermunt, J. (2010). Longitudinal research using mixture models. In Oud, J. H. L. & Satorra, A. (Eds.), *Longitudinal Research with Latent Variables,* (pp. 119-152), Berlin Heidelberg: Springer.

Welsh, M. E., Miller, F. G., Kooken, J., Chafouleas, S. M., & McCoach, D. B. (in press). The Kindergarten Transition: Behavior trajectories and school adjustment in the first year of school. *Journal of Research in Childhood Education.*

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*(2), 357-364.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation,* 12(3), 1-26.

Appendix A

Final Model Fit Indices

Table A1.

Fit Indices for Models 1-5: MS

Model	Classes	AIC	BIC	SABIC	Entropy	LMR p-value	BLRT p-value	Best LL replicated	Convergence or matrix problems
1	1	41,469	41,509	41,480	-	-	-	Yes	No
	2	37,682	37,752	37,701	0.97	0.10	<.001	Yes	No
	3	36,606	36,706	36,633	0.97	0.17	<.001	Yes	No
	4	36,096	36,228	36,132	0.93	0.81	1.00	Yes	Yes[1]
2	2	34,562	34,684	34,595	0.95	0.23	<.001		
	3	34,403	34,556	34,445	0.95	0.72	<.001	Yes	No
	4	34,266	34,450	34,317	0.95	-	-	No	Yes[2]
3	2	33,649	33,798	33,690	0.98	-		Yes	Yes[3]
	3	33,350	33,556	33,407	0.79	-	-	No	Yes[3]
4	2	30,968	31,104	31,006	0.94	0.03 [5]	<.001 [5]	Yes	No
	3	29,548	29,728	29,597	0.93	[6]	[6]	No	No
	4	-	-	-	-	-	-	No	Yes[4]
5	2	30,152	30,315	30,197	0.96	<.001	<.001	Yes	No
	3	28,489	28,721	28,553	0.94	0.25	<.001	Yes	No
	4	-	-	-	-	-	-	No	Yes[4]

Note. Slopes are estimated with fixed effects for all models.

AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; SABIC = Sample-size adjusted BIC.

LMR = Lo, Mendell and Rubin's Likelihood Ratio Test; BLRT= Bootstrap Likelihood Ratio Test; LL = Log Likelihood.

Model 1= intercepts and slopes variances constrained to 0 for all classes at within and between latent class level; residuals invariant between classes.

Model 2 = intercepts vary within latent class at level 1 and level 2; intercept variances do not vary between classes; residuals invariant between classes.

Model 3 = intercepts vary within and between latent class at level 1 and within latent class at level 2; residuals invariant between classes.

Model 4 = intercepts vary within latent class at level 1 and level 2; intercept variances do not vary between classes; residuals noninvariant between classes.

Model 5: intercepts vary within and between latent class at level 1 and level 2; intercept variances vary between latent classes at level 1; residuals noninvariant between latent classes.

[1]Model reached a saddle point, but this was resolved with additional random starts.

[2]Model reached a saddle point, not resolved.

[3]Latent covariance (Psi) matrix not positive definite.

[4]The estimated covariance matrix in one class could not be inverted.

[5] An error occurred in the LMR and BLRT step in the estimation of one less class, but this was resolved with additional random starts.

[6] An error occurred in the LMR and BLRT step in the estimation of one less class not resolved with additional random starts.

Table A2.

Fit Indices for Models 1-5: UE-1

Model	Classes	AIC	BIC	SABIC	Entropy	LMR p-value	BLRT p-value	Best LL replicated	Convergence or matrix problems
1	1	51,920	51,961	51,932	-	-	-	Yes	No
	2	46,751	46,824	46,773	0.99	0.33	<.001	Yes	No
	3	44,599	44,704	44,632	0.97	0.03	<.001	Yes	No
	4	43,763	43,900	43,804	0.93	0.76	1.00	Yes	No
2	2	40,779	40,908	40,819	0.99	0.48	<.001	Yes	No
	3	40,493	40,654	40,543	0.98	0.30	<.001	Yes	No
	4	40,402	40,595	40,462	0.98	-	-	Yes	Yes[1]
3	2	39,901	40,057	39,949	0.79	-	-	Yes	Yes[2]
	3	-	-	-	-	-	-	No	Yes[2]
4	2	37,118	37,260	37,162	0.91	0.02	<.001	Yes	No
	3	-	-	-	-	-	-	No	Yes[2]
	4	-	-	-	-	-	-	No	Yes[2]
5	2	36,300	36,470	36,353	0.93	<.001	<.001	Yes	No
	3	-	-	-	-	-	-	No	Yes[2]

Note. Slopes are estimated with fixed effects for all models.

AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; SABIC = Sample-size adjusted BIC.

LMR = Lo, Mendell and Rubin's Likelihood Ratio Test; BLRT= Bootstrap Likelihood Ratio Test; LL = Log Likelihood.

Model 1= intercepts and slopes variances constrained to 0 for all classes at within and between latent class level; residuals invariant between classes.

Model 2 = intercepts vary within latent class at level 1 and level 2; intercept variances do not vary between classes; residuals invariant between classes.

Model 3 = intercepts vary within and between latent class at level 1 and within latent class at level 2; residuals invariant between classes.

Model 4 = intercepts vary within latent class at level 1 and level 2; intercept variances do not vary between classes; residuals noninvariant between classes.

Model 5 = intercepts vary within and between latent class at level 1 and level 2; intercept variances vary between latent classes at level 1;
residuals noninvariant between latent classes.

[1] Non positive definite first order derivative matrix.

[2] Covariance matrix in one class could not be inverted.

Table A3.

*Fit Indices for Models 1-5: UE-2**

Model	Classes	AIC	BIC	SABIC	Entropy	LMR p-value	BLRT p-value	Best LL replicated	Convergence or matrix problems
1	1	51,920	51,961	51,932	-	-	-	Yes	No
	2	46,751	46,824	46,773	0.99	0.33	<.001	Yes	No
	3	44,669	44,766	44,699	0.97	0.03	<.001	Yes	No
	4	43,763	43,900	43,804	0.93	0.76	1.00	Yes	No
2	2	40,928	41,047	40,965	0.94	0.48	<.001	Yes	No
	3	40,755	40,906	40,801	0.97	0.61	<.001	Yes	No
	4	40,519	40,703	40,576	0.96	0.79	<.001	Yes	No
	5	40,466	40,684	40,532	0.97	-	-	No	Yes$_1$
3	2	40,168	40,301	40,209	0.85	-	-	Yes	Yes$_2$
	3	-	-	-	-	-	-	No	Yes$_2$
4	2	37,224	37,348	37,262	0.91	0.002	<.001	Yes	No
	3	35,882	36,042	35,932	0.92	0.20	<.001	Yes	Yes$_3$
	4	-	-	-	-	-	-	No	Yes$_3$
5	2	36,666	36,803	36,708	0.93	<.001	<.001	Yes	No
	3	35,053	35,241	35,111	0.93	0.11	<.001	Yes	No
	4	-	-	-	-	-	-	Yes	Yes$_3$

Note. Slopes are estimated with fixed effects for all models.

AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; SABIC = Sample-size adjusted BIC.

LMR = Lo, Mendell and Rubin's Likelihood Ratio Test; BLRT= Bootstrap Likelihood Ratio Test; LL = Log Likelihood.

Model 1= intercepts and slopes variances constrained to 0 for all classes at within and between latent class level; residuals invariant between classes.

Model 2 = intercepts vary within latent class at level 1 and level 2; intercept variances do not vary between classes; residuals invariant between classes.

Model 3 = intercepts vary within and between latent class at level 1 and within latent class at level 2; residuals invariant between classes.

Model 4 = intercepts vary within latent class at level 1 and level 2; intercept variances do not vary between classes; residuals noninvariant between classes.

Model 5 = intercepts vary within and between latent class at level 1 and level 2; intercept variances vary between latent classes at level 1;

residuals noninvariant between latent classes.

[1] The first derivative matrix is not positive definite and could not be inverted.

[2] Latent covariance (Psi) matrix not positive definite.

[3] The estimated covariance matrix in once class could not be inverted.

*Residual variances were constrained as time invariant.

Table B2

Fit Indices for Models 1-5: Lower elementary

Model	Classes	AIC	BIC	SABIC	Entropy	LMR p-value	BLRT p-value	LL replicated	Convergence or matrix problems
1	1	49,016	49,056	49,028	-	-	-	-	No
	2	44,154	44,227	44,175	0.98	0.10	<.001	Yes	No
	3	42,599	42,702	42,630	0.94	0.19	<.001	Yes	No
	4	41,997	42,131	42,036	0.93	0.69	<.001	Yes	No
	5	41,650	41,816	41,699	0.93	0.76	1.00	Yes	No
2	2	40,086	40,211	40,122	0.95	0.49	<.001	Yes	No
	3	39,957	40,114	40,003	0.96	0.52	<.001	Yes	No
	4	39,864	40,052	29,919	0.96	-	-	No	Yes[1]
3	2					-	-		
	3	39,220	39,431	39,282	0.81	-	-	Yes	Yes[2]
4	2	36,657	37,796	36,697	0.91	0.02	<.001	Yes	No
	3	35,293	35,477	35,347	0.93	0.31	<.001	Yes	No
	4	-	-	-	-	-	-	No	Yes[3]
5	2	36,172	36,338	36,220	0.92	0.01	<.001	Yes	No
	3	34,519	34,756	34,588	0.94	0.55	<.001	Yes	Yes[3,4]
	4	-	-	-	-	-	-	No	Yes[3]

Note. Slopes are estimated with fixed effects for all models.

AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; SABIC = Sample-size adjusted BIC.

LMR = Lo, Mendell and Rubin's Likelihood Ratio Test; BLRT= Bootstrap Likelihood Ratio Test; LL = Log Likelihood.

Model 1= intercepts and slopes variances constrained to 0 for all classes at within and between latent class level; residuals invariant between classes.

Model 2 = intercepts vary within latent class at level 1 and level 2; intercept variances do not vary between classes; residuals invariant between classes.

Model 3 = intercepts vary within and between latent class at level 1 and within latent class at level 2; residuals invariant between classes.

Model 4 = intercepts vary within latent class at level 1 and level 2; intercept variances do not vary between classes; residuals noninvariant between classes.

Model 5 = intercepts vary within and between latent class at level 1 and level 2; intercept variances vary between latent classes at level 1;

residuals noninvariant between latent classes.

[1] Non positive definite first order derivative matrix.

[2] Latent covariance (Psi) matrix not positive definite.

[3] The estimated covariance matrix in once class could not be inverted.

[4] Models would not converge so covariances were constrained to equal across classes.

Appendix B

Parameter Estimates

Table B1

Results from Final Unconditional 3-class Model 5—MS

Parameter Within Teacher	C1 Optimal Estimate (SE)	C2 Average Estimate (SE)	C3 Lower Estimate (SE)
Variances			
IW1	0.02 (.01)	1.00 (0.23)**	16.52 (3.93)**
SW1	0.00	0.00	0.00
IW2	0.01 (.01)	0.58 (0.25)**	14.92 (3.39)**
SW2	0.00	0.00	0.00
IW3	0.02 (.02)	0.42 (0.12)**	14.07 (2.91)**
SW3	0.00	0.00	0.00
Residual Variances			
DBR1	0.06 (0.02)**	2.12 (0.31)**	10.32 (1.30) **
DBR2	0.06 (0.02)**	2.12 (0.31)**	10.32 (1.30) **
DBR3	0.06 (0.02)**	2.12 (0.31)**	10.32 (1.30) **
DBR4	0.06 (0.02)**	2.12 (0.31)**	10.32 (1.30) **
DBR5	0.06 (0.02)**	2.12 (0.31)**	10.32 (1.30) **
DBR6	0.09 (0.02)**	0.97 (0.17)**	8.28 (1.14) **
DBR7	0.09 (0.02)**	0.97 (0.17)**	8.28 (1.14) **
DBR8	0.09 (0.02)**	0.97 (0.17)**	8.28 (1.14) **
DBR9	0.09 (0.02)**	0.97 (0.17)**	8.28 (1.14) **
DBR10	0.09 (0.02)**	0.97 (0.17)**	8.28 (1.14) **
DBR11	0.12 (0.02)**	1.05 (0.23)**	9.42 (1.14) **
DBR12	0.12 (0.02)**	1.05 (0.23)**	9.42 (1.14) **
DBR13	0.12 (0.02)**	1.05 (0.23)**	9.42 (1.14) **
DBR14	0.12 (0.02)**	1.05 (0.23)**	9.42 (1.14) **
DBR15	0.12 (0.02)**	1.05 (0.23)**	9.42 (1.14) **
Covariance (IW1,IW2)	0.00 (0.01)	0.42 (0.15)**	9.11 (2.94)**
Covariance (IW2,IW3)	0.00 (0.01)	0.27 (0.10)**	8.07 (2.51) **
Covariance (IW1,IW3)	0.00 (0.01)	0.28 (0.12)*	8.19 (2.80) **
Between Teacher			
Means			
IB1	29.78 (0.07)**	28.04 (0.21)**	23.71 (.061) **
SB1	0.02 (0.10)*	0.16 (0.06)**	0.14 (0.08)
IB2	29.73 (0.08)**	28.63 (0.15)**	23.66 (0.62) **
SB2	0.01 (0.01)	0.05 (0.03)	.36 (0.11) **

IB3	29.63 (0.09)**	28.54 (0.14)**	24.03 (0.55) **
SB3	0.02 (0.01)	0.10 (0.03)**	.27 (0.09) **
Variances			
IB1	0.05 (.03)	0.05 (.03)	0.05 (.03)
SB1	0.00	0.00	0.00
IB2	0.06 (.04)	0.06 (.04)	0.06 (.04)
SB2	0.00	0.00	0.00
IB3	0.15 (.08)	0.15 (.08)	0.15 (.08)
SB3	0.00	0.00	0.00
Covariance (IB1, IB2)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
Covariance (IB2, IB3)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)
Covariance (IB1, IB3)	0.01 (0.03)	0.01 (0.03)	0.01 (0.03)

**p < .01
*p < .05

Table B2
Results from Final Unconditional 3-class Model 5 UE

Parameter Within Teacher	C1 Optimal Estimate (SE)	C2 Average Estimate (SE)	C3 Maladaptive Estimate (SE)
Variances			
IW1	0.19 (0.10)	1.36 (0.28) **	14.96 (2.57) **
SW1	0.00	0.00	0.00
IW2	0.10 (0.09)	1.02 (0.17)**	9.22 (2.06) **
SW2	0.00	0.00	0.00
IW3	0.22 (0.13)	0.93 (0.17)**	12.61 (3.70)**
SW3	0.00	0.00	0.00
Residual Variances			
DBR1	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR2	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR3	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR4	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR5	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR6	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR7	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR8	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR9	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR10	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR11	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR12	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR13	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR14	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
DBR15	0.15 (0.03)**	1.33 (0.17)**	6.95 (0.74)**
Covariance (IW1,IW2)	0.10 (0.20)	0.10 (0.20)	0.10 (0.20)
Covariance (IW2,IW3)	0.12 (0.10)	0.12 (0.10)	0.12 (0.10)
Covariance (IW1,IW3)	0.14 (0.09)	0.14 (0.09)	0.14 (0.09)
Between Teacher			
Means			
IB1	29.37 (0.13)**	27.58 (0.20)**	22.93 (0.67)**
SB1	0.02 (0.02)	0.14 (0.04) **	0.28 (0.08)**
IB2	29.47 (0.17)**	27.79 (0.21)**	24.08 (0.53)**
SB2	0.01 (0.01)	0.07 (0.03)*	0.20 (0.07)**
IB3	29.31 (0.15)**	27.98 (0.20)**	24.53 (0.64)**
SB3	0.02 (0.01)	0.08 (0.03)**	0.13 (0.07)**
Variances			

IB1	0.37 (0.24)	0.37 (0.24)	0.37 (0.24)
SB1	0.00	0.00	0.00
IB2	0.65 (0.66)	0.65 (0.66)	0.65 (0.66)
SB2	0.00	0.00	0.00
IB3	1.25 (0.74)	1.25 (0.74)	1.25 (0.74)
SB3	0.00	0.00	0.00
Covariance (IB1, IB2)	0.43 (0.40)	0.43 (0.40)	0.43 (0.40)
Covariance (IB2, IB3)	0.90 (0.70)	0.90 (0.70)	0.90 (0.70)
Covariance (IB1, IB3)	0.61 (0.42)	0.61 (0.42)	0.61 (0.42)

**$p < .01$

*$p < .05$

Table B3
Results from Final Unconditional 3-class Model 5 LE

Parameter Within Teacher	C1 Optimal Estimate (SE)	C2 Average Estimate (SE)	C3 Maladaptive Estimate (SE)
Variances			
IW1	0.06 (0.03)	2.80 (0.63)**	15.07 (2.39)**
SW1	0.00	0.00	0.00
IW2	0.04 (0.01)**	2.30 (0.63)**	11.34 (1.96)**
SW2	0.00	0.00	0.00
IW3	0.02 (0.02)	1.74 (1.32)**	11.28 (1.89)**
SW3	0.00	0.00	0.00
Residual Variances			
DBR1	0.23 (0.04)**	1.68 (0.21)**	9.30 (1.55)**
DBR2	0.23 (0.04)**	1.68 (0.21)**	9.30 (1.55)**
DBR3	0.23 (0.04)**	1.68 (0.21)**	9.30 (1.55)**
DBR4	0.23 (0.04)**	1.68 (0.21)**	9.30 (1.55)**
DBR5	0.23 (0.04)**	1.68 (0.21)**	9.30 (1.55)**
DBR6	0.10 (0.03)**	1.44 (0.23) **	9.70 (1.79)**
DBR7	0.10 (0.03)**	1.44 (0.23) **	9.70 (1.79)**
DBR8	0.10 (0.03)**	1.44 (0.23) **	9.70 (1.79)**
DBR9	0.10 (0.03)**	1.44 (0.23) **	9.70 (1.79)**
DBR10	0.10 (0.03)**	1.44 (0.23) **	9.70 (1.79)**
DBR11	0.04 (0.02)	1.08 (0.17) **	6.48 (1.49)**
DBR12	0.04 (0.02)	1.08 (0.17) **	6.48 (1.49)**
DBR13	0.04 (0.02)	1.08 (0.17) **	6.48 (1.49)**
DBR14	0.04 (0.02)	1.08 (0.17) **	6.48 (1.49)**
DBR15	0.04 (0.02)	1.08 (0.17) **	6.48 (1.49)**
Covariance (IW1,IW2)	0.02 (0.02)	1.93 (0.50)**	6.47 (1.74)**
Covariance (IW2,IW3)	0.00 (0.01)	1.36 (0.78)	7.36 (1.56)**
Covariance (IW1,IW3)	0.01 (0.01)	1.35 (0.71)	7.16 (1.56)**
Between Teacher			
Means			
IB1	29.47 (.20)**	27.16 (0.40)**	22.00 (0.60)**
SB1	.01 (.02)	0.10 (0.04)*	0.40 (0.10)**
IB2	29.49 (0.26)**	27.45 (0.44)**	22.32 (0.56)**
SB2	0.01 (0.01)	0.09 (0.04)*	0.38 (0.12)**
IB3	29.65 (.27)	27.64 (0.54)**	23.89 (0.37)**
SB3	0.00 (0.01)	0.10 (0.03)**	0.17 (0.07)*

Variances			
IB1	0.33 (0.38)	0.33 (0.38)	0.33 (0.38)
SB1	0.00	0.00	0.00
IB2	0.45 (0.64)	0.45 (0.64)	0.45 (0.64)
SB2	0.00	0.00	0.00
IB3	0.38 (0.64)	0.38 (0.64)	0.38 (0.64)
SB3	0.00	0.00	0.00
Covariance (IB1, IB2)	0.30 (0.46)	0.30 (0.46)	0.30 (0.46)
Covariance (IB2, IB3)	0.36 (0.64)	0.36 (0.64)	0.36 (0.64)
Covariance (IB1, IB3)	0.26 (0.46)	0.26 (0.46)	0.26 (0.46)

**p < .01
*p < .05

Appendix C

Comparisons for Model 2

Table C1.
Comparison of Latent Class Assignment Model 2 versus Model 5, MS

		Model 5 Latent Class				
MS		1	2	3	Model 2 Sum	Model 2 Proportion
Model 2 Latent Class	1	100	262	159	521	88%
	2	0	0	38	38	6%
	3	0	0	32	32	5%
	Model 5 Sum	100	262	229	591	
	Model 5 Proportion	17%	44%	39%		100%

Table C2.
Demographic distribution by change group—MS—Model 2 versus Model 5

Category	All MS	Always Class 1	Always Class 3	Class 1,2	Class 1,3	Class 2,3
Male	.54	.32	.81	.51	.61	.79
White	.82	.95	.72	.85	.74	.66
Black	.13	.01	.25	.08	.21	.32
Retained	.06	.01	.03	.06	.09	.08
Special Education	.18	.03	.13	.19	.23	.29
EBD	.02	.00	.00	.01	.04	.03
Behavior	.04	.01	.09	.02	.04	.21
Days Absent	7.62	5.32	9.68	7.50	8.30	10.90
Suspension or Expulsion	.09	.01	.22	.06	.11	.21
Brisk	.19	.00	.38	.14	.31	.34
Erisk	.11	.00	.47	.02	.16	.53
Hispanic	.09	.06	.19	.07	.13	.08
ODRRISK	.20	.06	.63	.12	.25	.55
DBR mean	27.07	29.80	20.95	28.50	25.48	21.46
DBR SD	3.22	.20	3.86	.85	2.31	3.67
BESS T score mean	51.70	40.30	64.00	49.19	56.99	66.55
BESS T score SD	10.71	3.99	9.75	7.61	9.02	8.83
n	591	100.00	32.00	262.00	159.00	38.00

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS,
ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 1 and Class b in Model 5.
Only nonempty change classes are reported.

Table C3.
Comparison of Latent Class Assignment Model 2 versus Model 5, UE

		Model 5 Latent Class				
UE		1	2	3	Model 2 Sum	Model 2 Proportion
Model 2 Latent Class	1	165	359	148	672	93%
	2	0	0	30	30	4%
	3	0	0	23	23	3%
	Model 5 Sum	165	359	201	725	
	Model 5 Proportion	23%	50%	28%		100%

Table C4.
Demographic distribution by change group—UE—Model 2 versus Model 5

Category	All UE	Always Class 1	Always Class 3	Class 1,2	Class 1,3	Class 2,3
Male	.51	.38	.91	.47	.66	.63
White	.81	.86	.57	.84	.76	.63
Black	.13	.08	.30	.12	.16	.30
Retained	.04	.01	.26	.03	.05	.10
Special Education	.11	.05	.26	.11	.15	.20
EBD	.00	.01	.00	.00	.01	.00
Behavior	.03	.01	.22	.02	.04	.10
Days Absent	5.55	4.72	8.06	5.67	5.46	7.59
Suspension or Expulsion	.03	.01	.04	.02	.07	.07
Brisk	.16	.02	.13	.12	.35	.47
Erisk	.07	.00	.70	.02	.14	.30
Hispanic	.05	.05	.13	.03	.07	.07
ODRRISK	.06	.01	.26	.03	.14	.17
DBR Mean	27.22	29.65	17.79	28.00	25.31	21.28
DBR SD	3.05	.36	4.50	1.12	2.14	3.21
BESS T score Mean	49.66	41.32	68.43	48.63	56.02	62.04
BESS T score SD	9.83	5.41	9.71	7.53	8.39	8.04
n	725	165	23	359	148	30

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS,
ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 1 and Class b in Model 5.
Only nonempty change classes are reported.

Table C5.
Comparison of Latent Class Assignment Model 2 versus Model 5, LE

		Model 5 Latent Class				
LE		1	2	3	Model 2 Sum	Model 2 Proportion
Model 2 Latent Class	1	105	334	158	597	91%
	2	0	0	28	28	4%
	3	0	0	32	32	5%
	Model 5 Sum	105	334	218	657	
	Model 5 Proportion	16%	51%	33%		100%

Table C6.
Demographic distribution by change group—LE—Model 2 versus Model 5

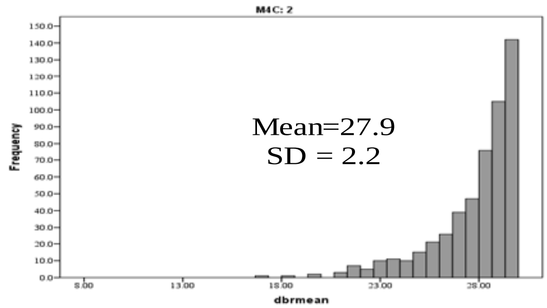
Category	All LE	Always Class 1	Always Class 3	Class 1,2	Class 1,3	Class 2,3
Male	.52	.30	.66	.54	.56	.79
White	.82	.86	.66	.86	.75	.86
Black	.10	.05	.19	.08	.15	.07
Retained	.06	.01	.16	.05	.10	.14
Special Education	.11	.06	.19	.10	.12	.29
EBD	.01	.00	.03	.01	.02	.00
Behavior	.06	.01	.16	.01	.12	.29
Days Absent	6.49	5.55	6.58	6.32	6.98	9.19
Suspension or Expulsion	.01	.00	.00	.01	.01	.14
Brisk	.20	.01	.44	.14	.34	.50
Erisk	.06	.00	.41	.02	.09	.32
Hispanic	.09	.06	.09	.09	.10	.04
ODRRISK	.11	.02	.44	.07	.16	.32
DBR mean	26.43	29.69	20.02	27.59	24.09	20.84
DBR SD	3.38	.46	3.58	1.51	2.74	4.37
BESS T score mean	50.41	40.16	65.57	48.81	55.27	63.06
BESS T score SD	9.88	5.45	8.48	7.54	8.30	7.53
n	657	105	32	334	158	28

Note: Retained=student was retained at least one grade, EBD=Emotional or behavioral disability, Behavior=student is receiving behavior supports, Days absent=number of days absent through final data collection point, Suspension or expulsion=student received at least one, Brisk=basic risk status on the BESS, Erisk=elevated risk status on the BESS,
ODRRISK=two or more office disciplinary referrals.
Class a,b= student was classified in Class a in Model 1 and Class b in Model 5.
Only nonempty change classes are reported.

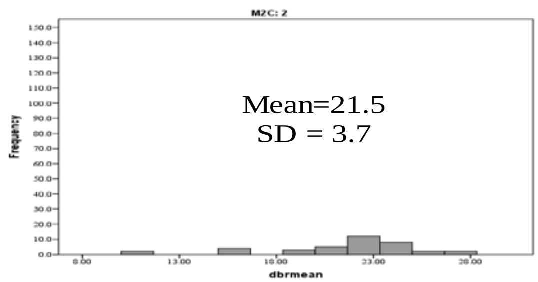
Figure C1. MS – DBR-SIS student level means for Model 2 by class

Model 2

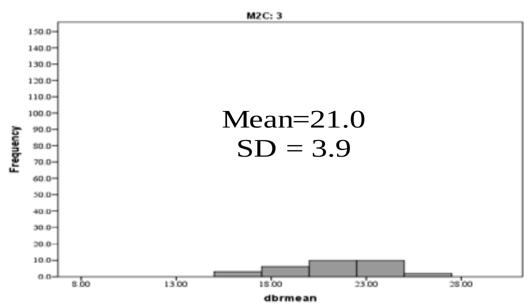
Class 1



Class 2



Class 3



Class Proportions	
Class 1	87%
Class 2	7%
Class 3	6%
Skewness	
Class 1	-1.636
Class 2	-1.195
Class 3	-1.748
Kurtosis	
Class 1	2.861
Class 2	1.585
Class 3	5.233
BESS T scores (SD)	
Class 1	49.9 (9.5)
Class 2	66.5 (8.8)
Class 3	64 (9.8)

Figure C1 Histogram displays the distribution of the within student mean DBR-SIS score by model and class for MS. The proportion of students assigned to each class, skewness, kurtosis, and class average BESS T scores are displayed at the bottom.

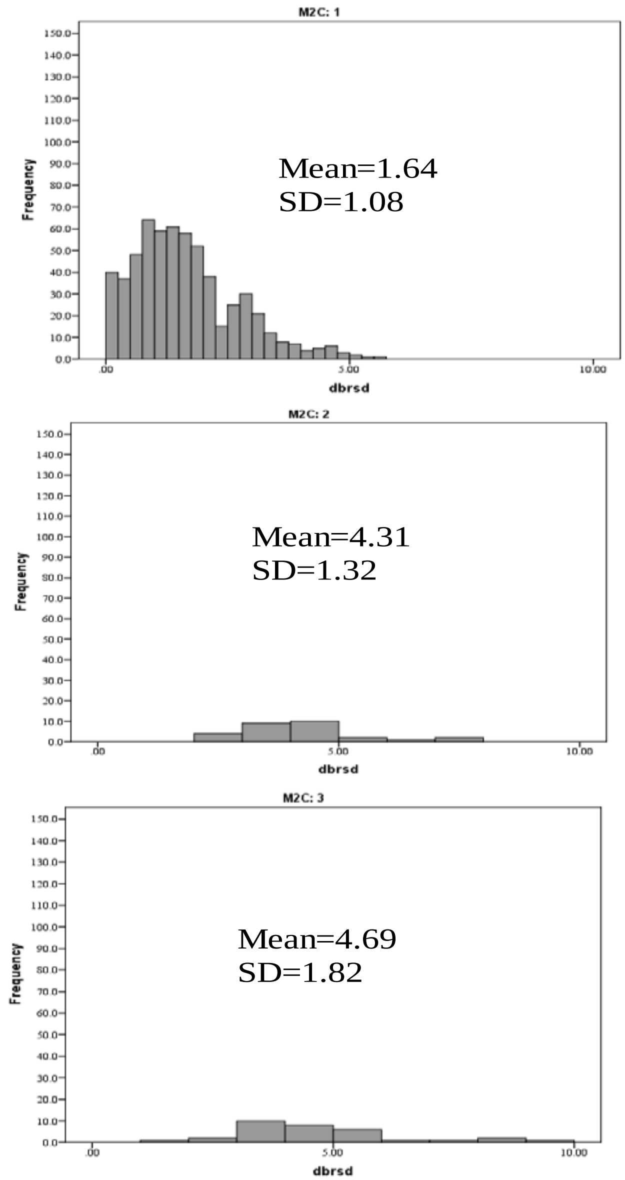
Figure C2. MS – DBR-SIS student level standard deviations for Model 2 by class

Model 2

Class 1

Class 2

Class 3

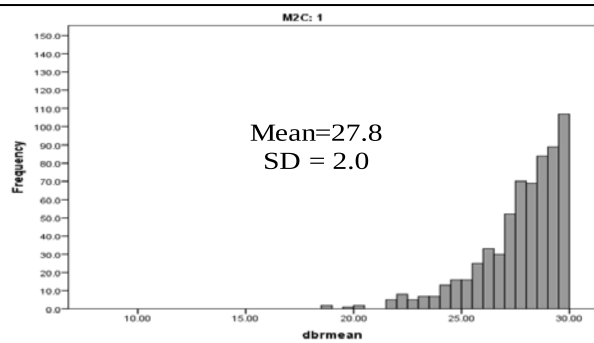


Class Proportions	
Class 1	91%
Class 2	4%
Class 3	5%

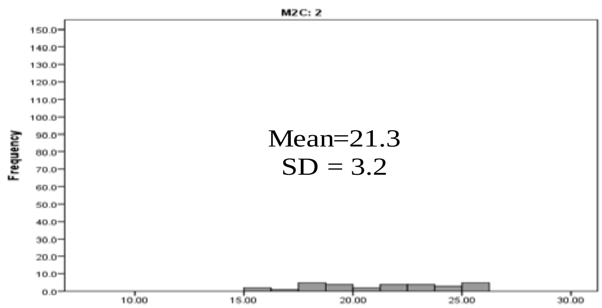
Figure C2. Histogram displays the distribution of the within student DBR-SIS standard deviation by model and class for MS. The proportion of students assigned to each class are displayed at the bottom.

Figure C3. UE – DBR-SIS student level means for Model 2 by class

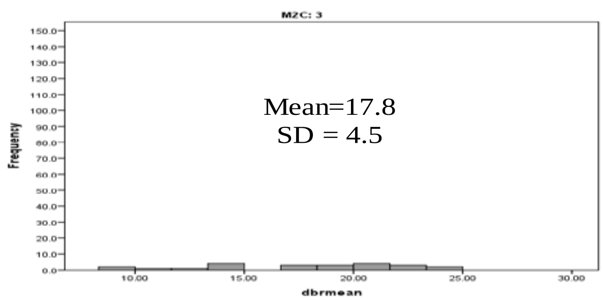
Class 1



Class 2



Class 3



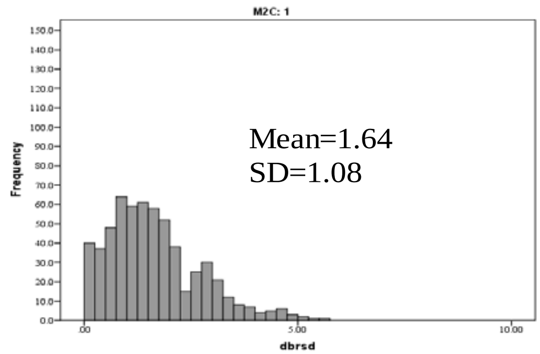
Class Proportions	
Class 1	92%
Class 2	5%
Class 3	3%
Skewness	
Class 1	-1.42
Class 2	-0.07
Class 3	-0.41
Kurtosis	
Class 1	2.25
Class 2	-1.05
Class 3	-0.82
BESS T scores (SD)	
Class 1	48.5 (8.8)
Class 2	62.0 (8.0)
Class 3	68.4 (9.7)

Figure C3 Histogram displays the distribution of the within student mean DBR-SIS score by model and class for UE. The proportion of students assigned to each class, skewness, kurtosis, and class average BESS T scores are displayed at the bottom.

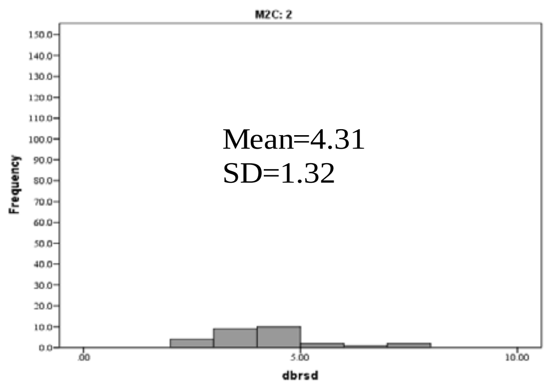
Figure C4. UE– DBR-SIS student level standard deviations for Model 2 by class

Model 2

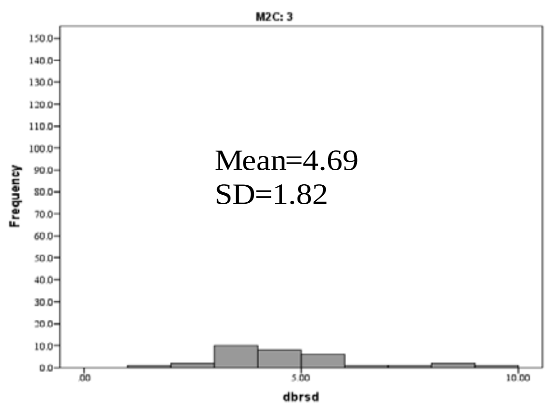
Class 1



Class 2



Class 3



Class Proportions	
Class 1	91%
Class 2	4%
Class 3	5%

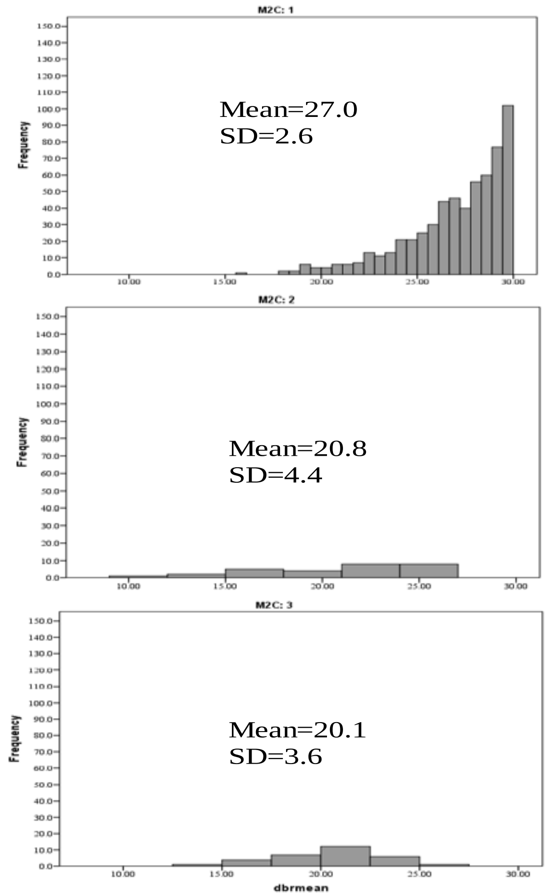
Figure C4. Histogram displays the distribution of the within student DBR-SIS standard deviation by model and class for UE. The proportion of students assigned to each class are displayed at the bottom.

Figure C5. LE – DBR-SIS student level means for Model 2 by class

Class 1

Class 2

Class 3



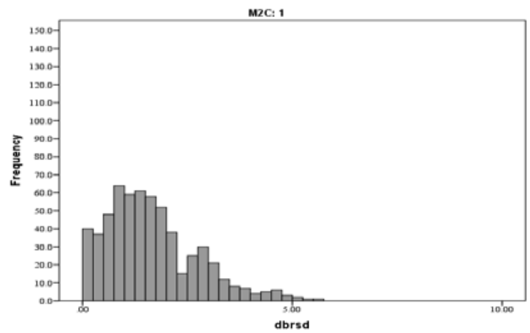
Class Proportions	
Class 1	91%
Class 2	4%
Class 3	5%
Skewness	
Class 1	-1.21
Class 2	-0.94
Class 3	-1.51
Kurtosis	
Class 1	1.18
Class 2	0.16
Class 3	4.08
BESS T scores (SD)	
Class 1	49.0 (8.9)
Class 2	63.1 (7.5)
Class 3	65.6 (8.5)

Figure C5. Histogram displays the distribution of the within student mean DBR-SIS score by model and class for LE. The proportion of students assigned to each class, skewness, kurtosis, and class average BESS T scores are displayed at the bottom.

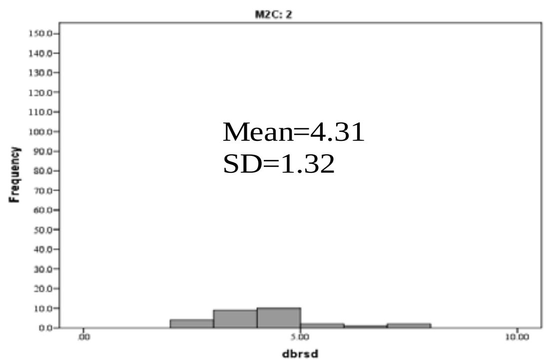
Figure C6. LE– DBR-SIS student level standard deviations for Model 2 by class

Model 2

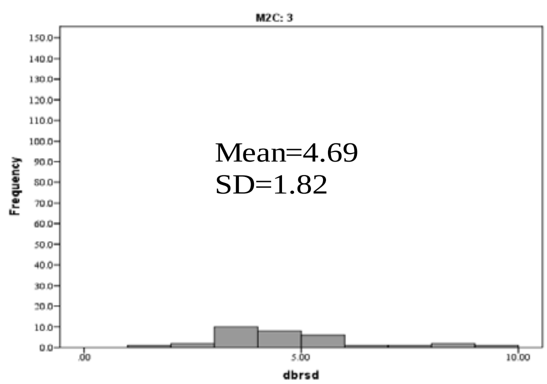
Class 1



Class 2



Class 3



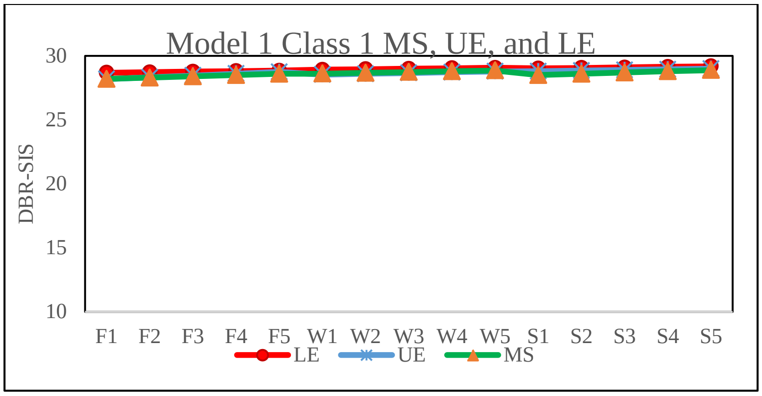
Class Proportions	
Class 1	91%
Class 2	4%
Class 3	5%

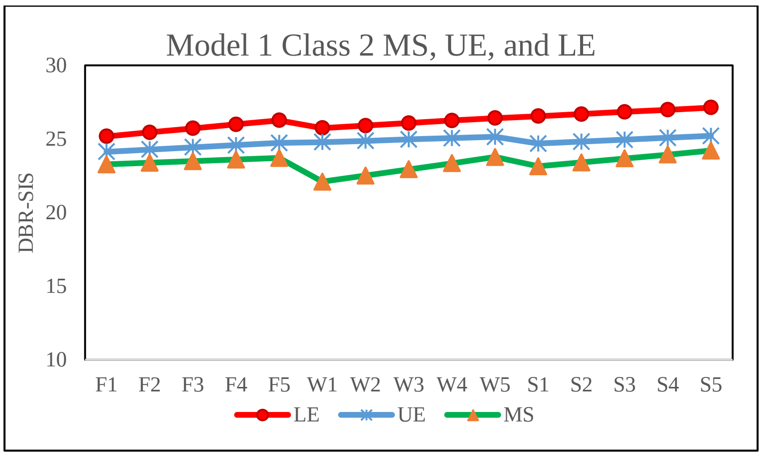
Figure C6. Histogram displays the distribution of the within student DBR-SIS standard deviation by model and class for LE. The proportion of students assigned to each class are displayed at the bottom.

Appendix D

Growth Curves by Model and Class

Figure D1. Model 1 Growth Curves by Class and Grade Group





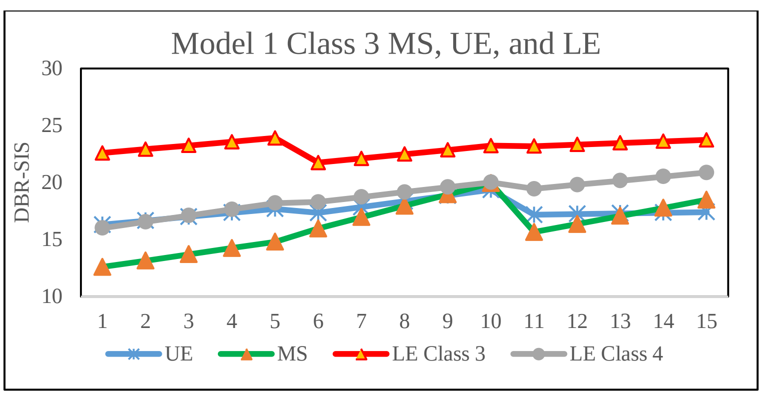
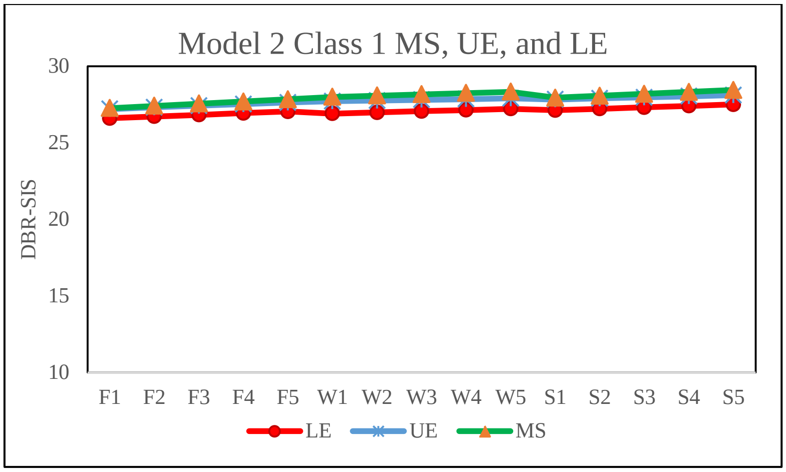
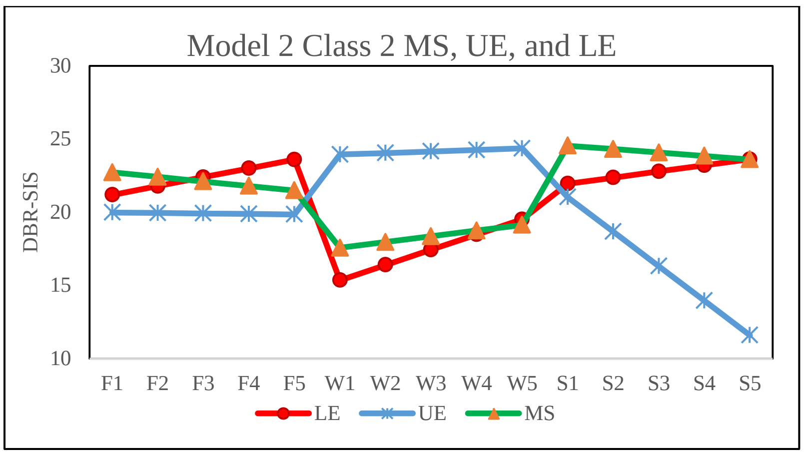


Figure D1. Model 1 DBR-SIS growth curves for LE, UE, and MS for Class 1, 2 and 3. The Class 3 growth curve also displays the Class 4 curve for LE.

Figure D2. Model 2 Growth Curves by Class and Grade Group





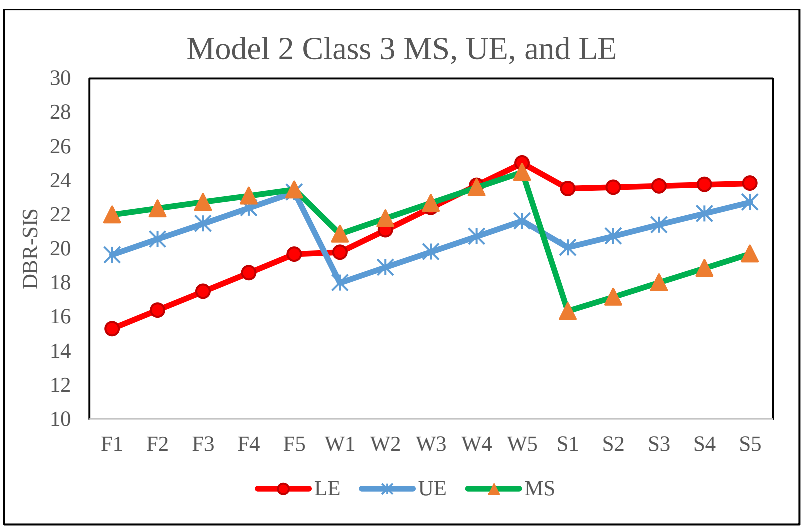
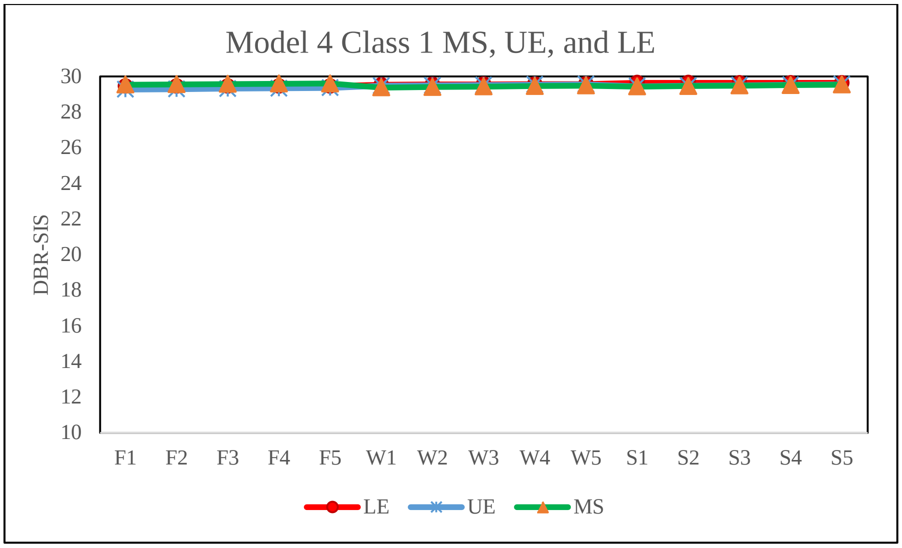
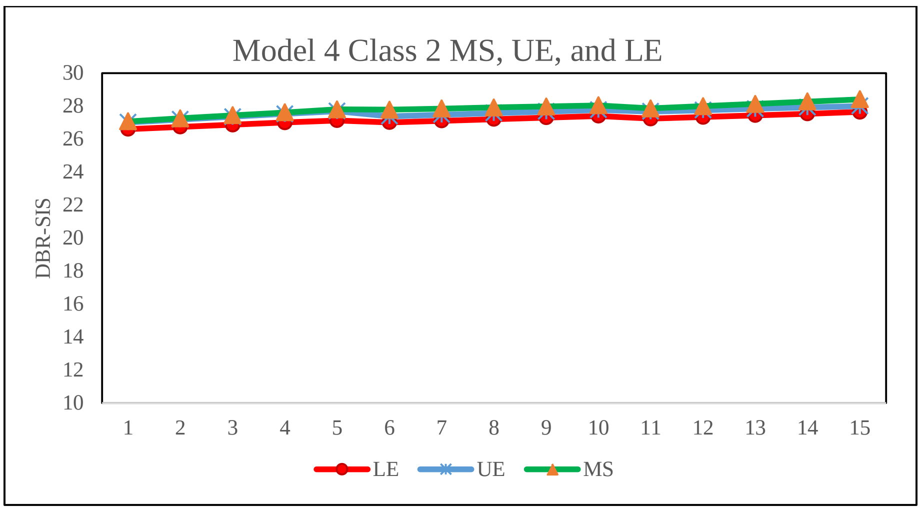


Figure D2. Model 2 DBR-SIS growth curves for LE, UE, and MS for Class 1, 2 and 3.

Figure D3. Model 4 Growth Curves by Class and Grade Group





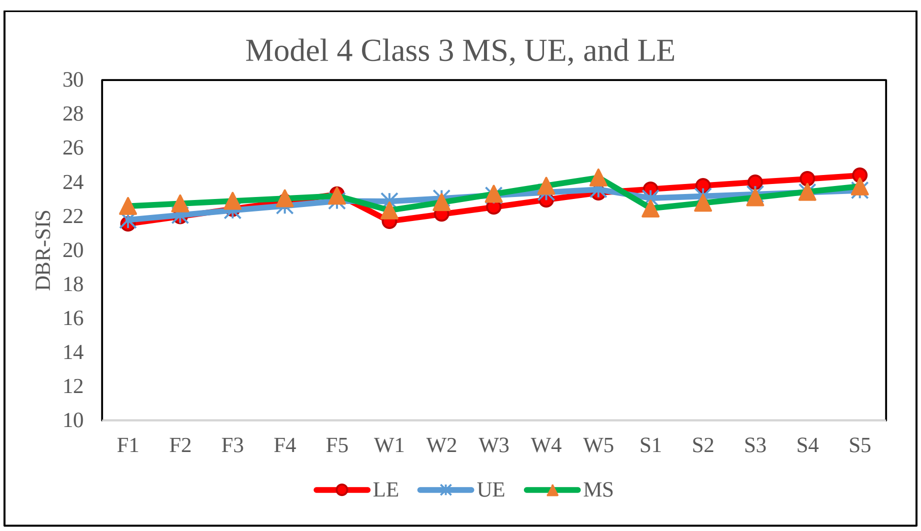


Figure D3. Model 4 DBR-SIS growth curves for LE, UE, and MS for Class 1, 2 and 3.

Appendix E
Model 5 Frequency Distributions for AE, RS, and DB
MS, UE, and LE

Figure E1.
Model 5 Mean frequency distributions for AE, RS, and DB – MS

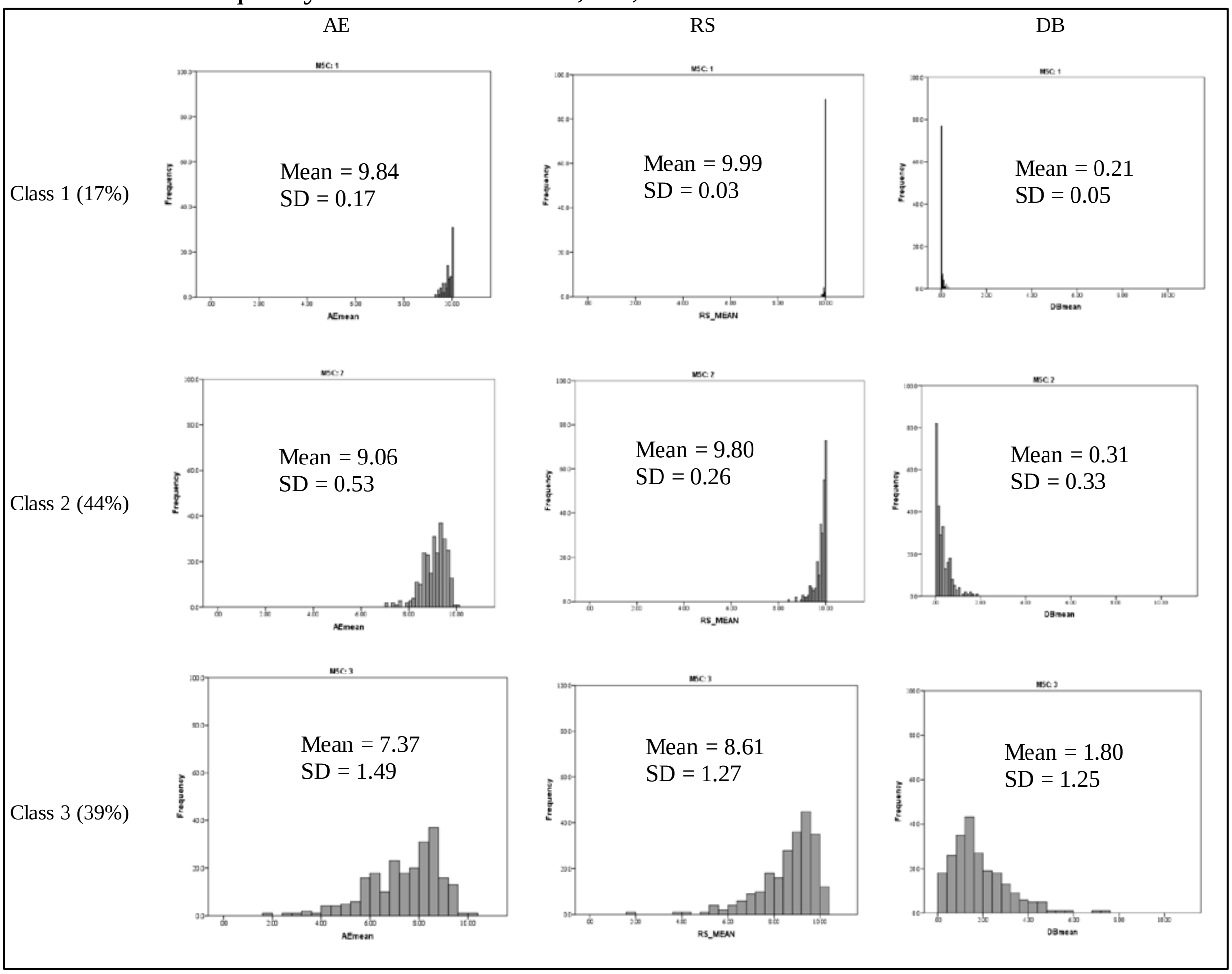


*Figure E1.*Histogram displays the distribution of the within student mean AE, RS, and DB scores by Class for MS Model 5.

Figure E2.
Model 5 Standard deviation frequency distributions for AE, RS, and DB – MS

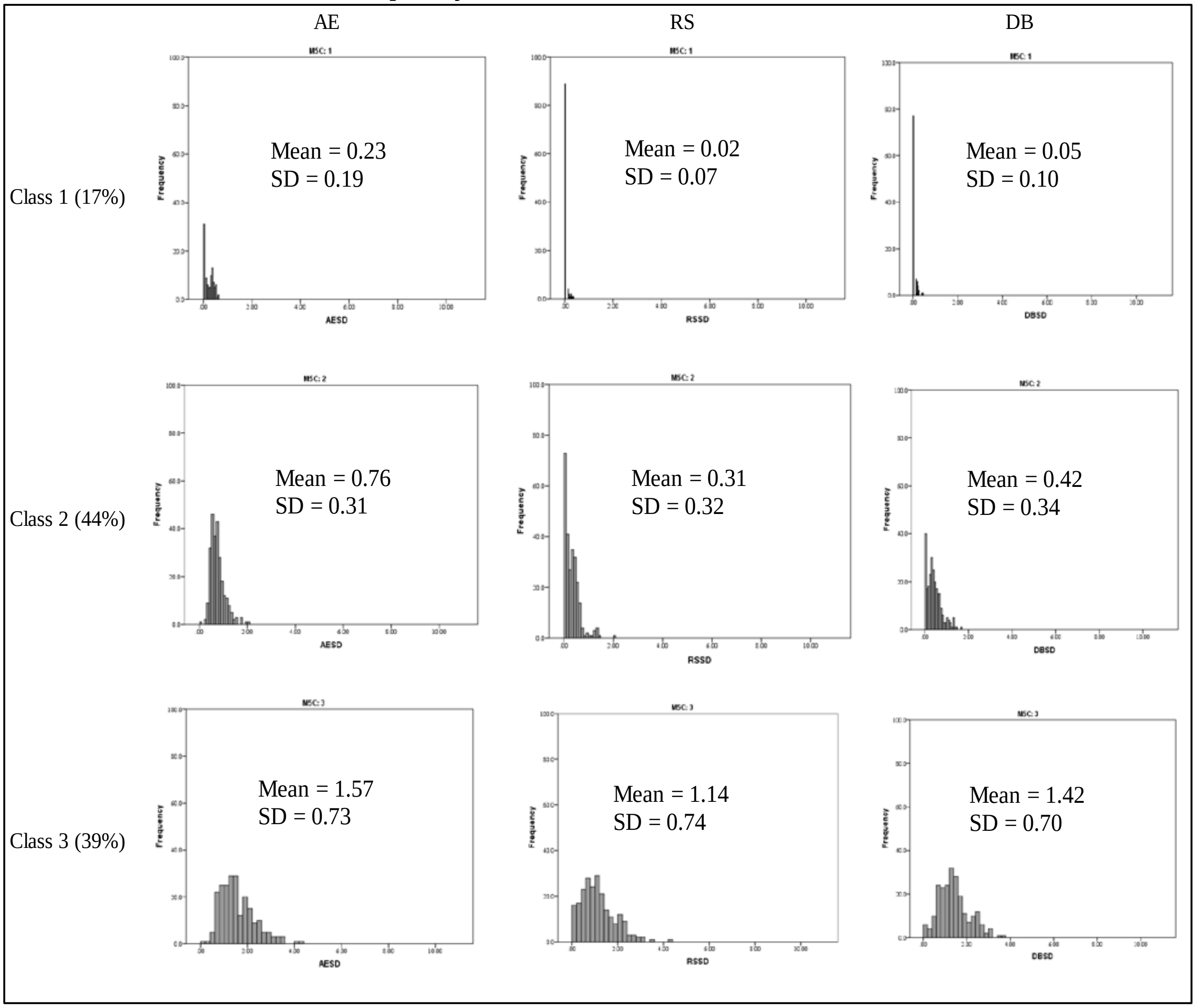


*Figure E2.*Histogram displays the distribution of the within student standard deviations for AE, RS, and DB score by Class for MS.

Figure E3.
Model 5 Mean frequency distributions for AE, RS, and DB – UE

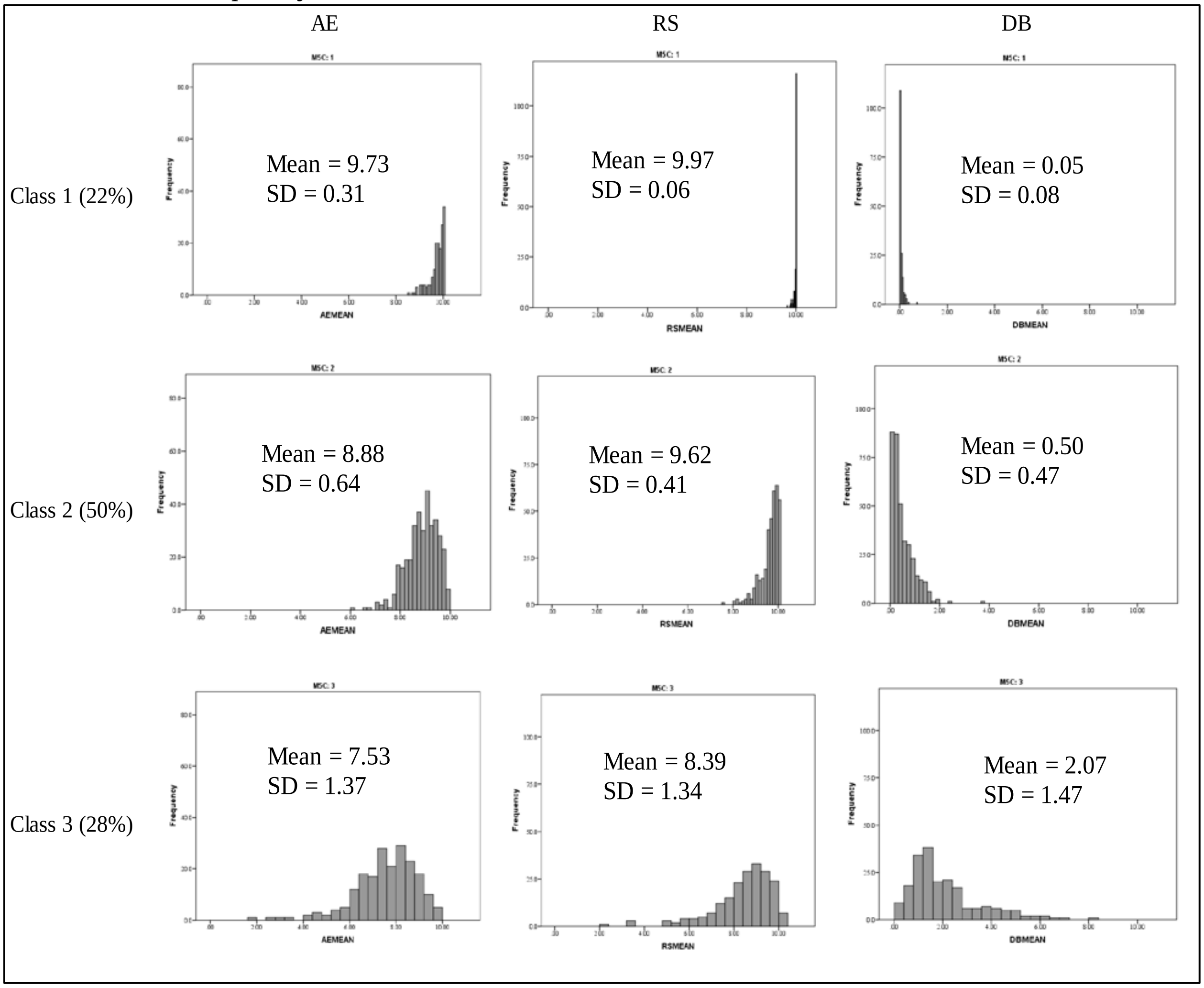


*Figure E3.*Histogram displays the distribution of the within student means for AE, RS, and DB score by Class for UE.

Figure E4.
Model 5 Standard deviation frequency distributions for AE, RS, and DB – UE

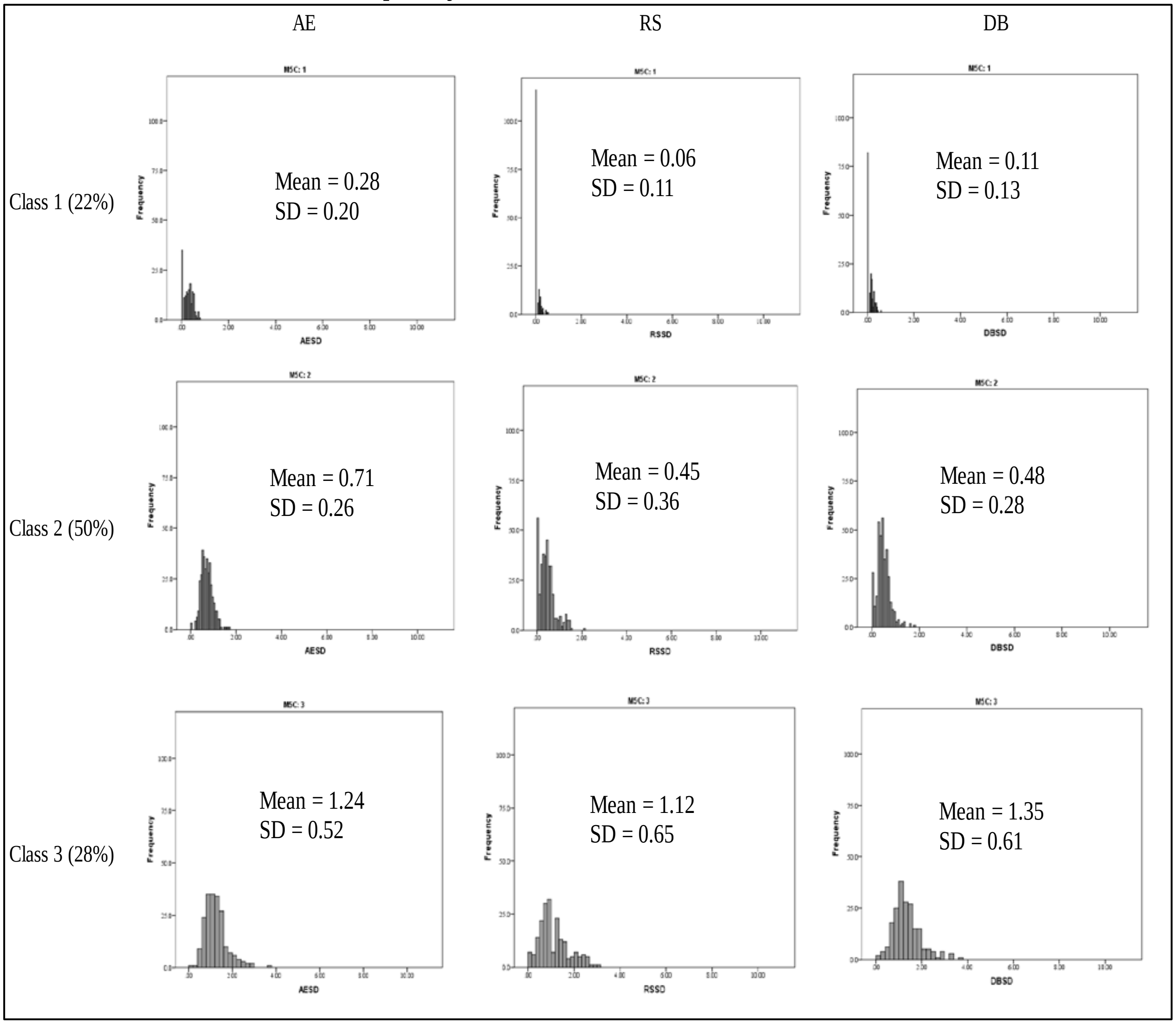


*Figure E4.*Histogram displays the distribution of the within student standard deviations for AE, RS, and DB score by Class for UE.

Figure E5.
Model 5 Mean frequency distributions for AE, RS, and DB – LE

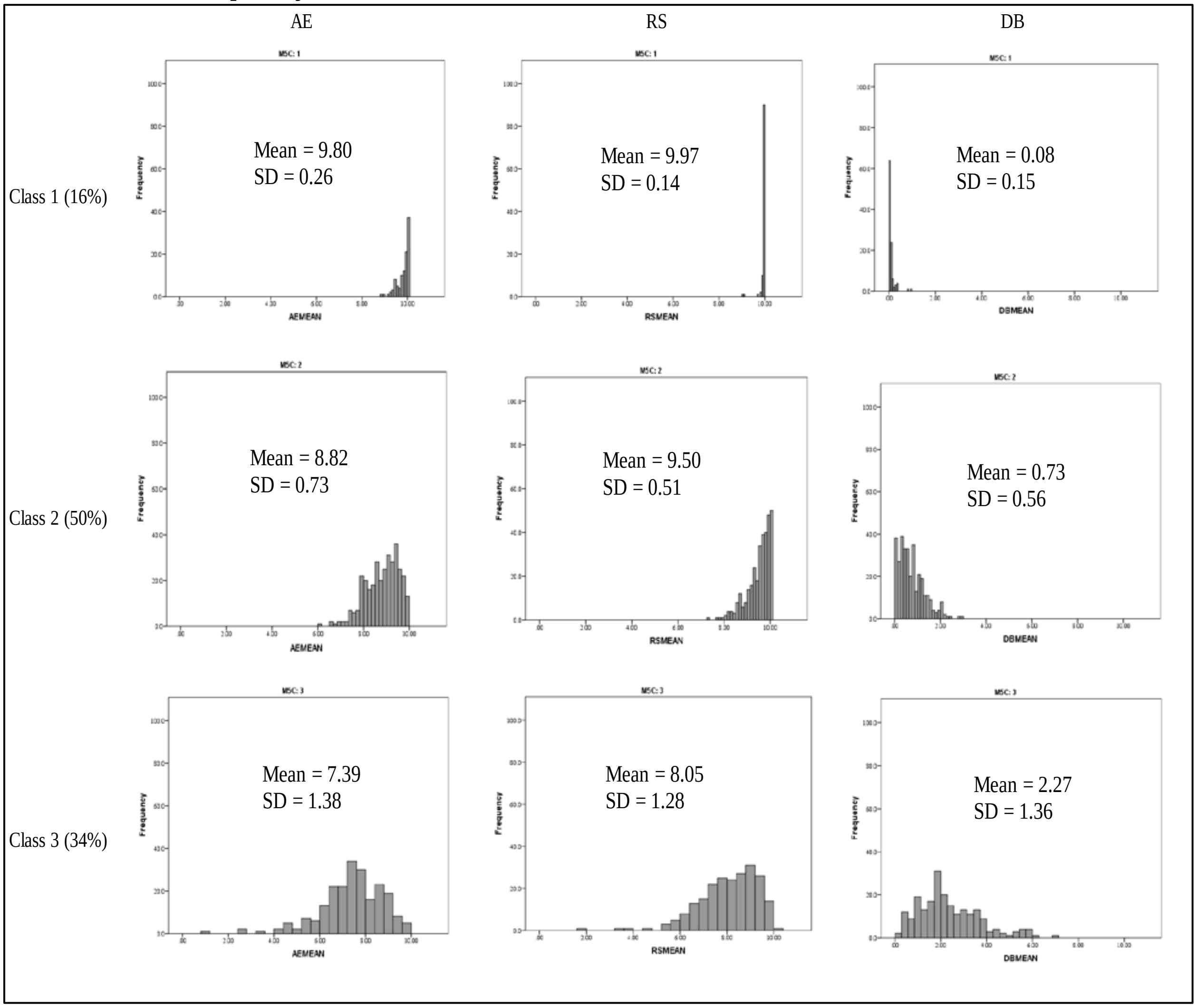


*Figure E5.*Histogram displays the distribution of the within student means for AE, RS, and DB score by Class for LE.

Figure E6.
Model 5 Standard deviation frequency distributions for AE, RS, and DB – LE

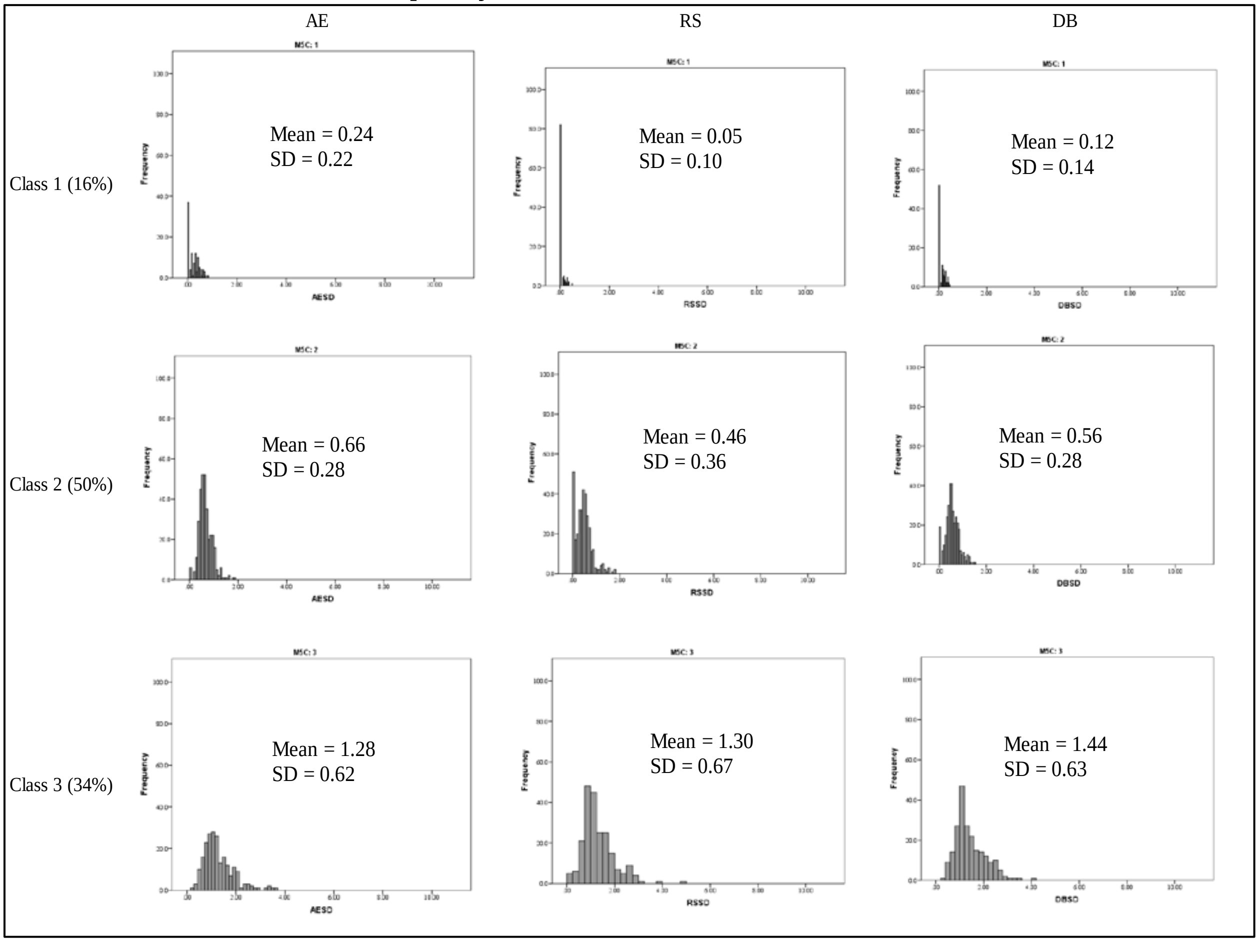


*Figure E6.*Histogram displays the distribution of the within student standard deviations for AE, RS, and DB score by Class for LE.

Appendix F

Abbreviations

Abbreviation	**Description**
AE	Academic Engagement
AIC	Akaike Information Criteria
BASC TRS-C	Behavior Assessment System for Children Teacher Rating Scales for Children
BESS	Behavioral and Emotional Screening System
BIC	Bayesian Information Criteria
BLRT	Bootstrap Likelihood Ratio Test
Brisk	Basic level risk on the BESS
DB	Disruptive Behavior
DBR-SIS	Direct Behavior Rating Single Item Scales
Erisk	Elevated level risk on the BESS
FIML	Full Information Maximum Likelihood Estimator
GM Model	Growth Mixture Model
LCG Model	Latent Class Growth Model
LE	Lower Elementary
LGCM	Latent Growth Curve Model
LMR	Lo, Mendell, and Rubin Likelihood Ration Test
MS	Middle School
ODR	Office Disciplinary Referral
RS	Respectful Behavior
RTI	Response to Intervention
SABIC	Sample Size Adjusted Bayesian Information Criteria
UE	Upper Elementary

Appendix G

Mplus Syntax

```
Title: DBR Multilevel Model 1 LCGM
Data: File is DBR.dat;
Variable:
Names are id sc grade st DBR1-DBR15
BESST1 BESST2 BESST3 timebess1 timebess2 timebess3
M W B R SPED EBD BEH Abs Sus Exp ODR Spedss EBDSS BEHSS risk SErisk
Brisk erisk;
Usev dbr1-dbr15 st; ! risk ;
idvariable=id;
useobservations (grade eq 4 or grade eq 5);
auxiliary = risk (E) ODR (E) SPED (E) M (E) W (E) B (E);
classes = c(3);
cluster=st;
missing are all(-999);
Analysis: type= twolevel mixture;

!Algorithm=integration;
!integration=10;
!mconvergence=0.01;
!miterations=5000;
processors = 8;
!starts=40 8;
!Stseed=107446;
optseed=608496;
```

```
Model:
%within%
%overall%
Iw1 Sw1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Iw2 SW2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Iw3 SW3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;
iw1@0; iw2@0; iw3@0;
sw1@0; sw2@0; sw3@0;
   dbr1 (1);
   dbr2 (1);
   dbr3 (1);
   dbr4  (1);
   dbr5 (1);
   dbr6 (1);
   dbr7 (1);
   dbr8 (1);
   dbr9 (1);
   dbr10 (1);
   dbr11 (1);
   dbr12 (1);
   dbr13 (1);
   dbr14 (1);
   dbr15 (1);


%between%
%overall%
```

```
Ib1 Sb1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Ib2 Sb2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Ib3 Sb3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;
 ! DBR7@0
  ib1@0; ib2@0; ib3@0;
  sb1@0; sb2@0; sb3@0;




!savedata:
!file is model1class3fws.dat;
!save=cprob;
output:
!sampstat standardized ;
 tech14;
plot:
series=dbr1-dbr15 (*);
type=plot3;
```

Title: DBR Multilevel Model 2

```
Data: File is DBR.dat;
Variable:
Names are id sc grade st DBR1-DBR15
BESST1 BESST2 BESST3 timebess1 timebess2 timebess3
M W B R SPED EBD BEH Abs Sus Exp ODR Spedss EBDSS BEHSS risk SErisk
Brisk erisk;
Usev dbr1-dbr15 st; ! risk ;
idvariable=id;
useobservations (grade eq 7 or grade eq 8);
!auxiliary = risk (E) ODR (E) SPED (E) M (E) W (E) B (E);
classes = c(3);
cluster=st;
missing are all(-999);
Analysis: type= twolevel mixture;
!Algorithm=integration;
!integration=10;
!mconvergence=0.01;
!miterations=5000;
processors = 8;
starts=40 8;
Stseed=939021;
!stseed=415931 ;
Model:
%within%
%overall%
Iw1 Sw1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Iw2 SW2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
```

```
Iw3 SW3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;

sw1@0; sw2@0; sw3@0;
dbr1 (1);
  dbr2 (1);
  dbr3 (1);
  dbr4  (1);
  dbr5 (1);
  dbr6 (2);
  dbr7 (2);
  dbr8 (2);
  dbr9 (2);
  dbr10 (2);
  dbr11 (3);
  dbr12 (3);
  dbr13 (3);
  dbr14 (3);
  dbr15 (3);

%between%
%overall%
Ib1 Sb1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Ib2 Sb2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Ib3 Sb3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;
  sb1@0; sb2@0; sb3@0;
```

```
!Savedata:
!file is model2class.dat;
!save=cprob;

plot:
series=dbr1-dbr15 (*);
type=plot3;
```

```
Title: DBR Multilevel Model 3
Data: File is DBR.dat;
Variable:
Names are id sc grade st DBR1-DBR15
BESST1 BESST2 BESST3 timebess1 timebess2 timebess3
M W B R SPED EBD BEH Abs Sus Exp ODR Spedss EBDSS BEHSS risk SErisk
Brisk erisk;
Usev dbr1-dbr15 st; ! risk ;
idvariable=id;
useobservations (grade eq 7 or grade eq 8);
!auxiliary = risk (E) ODR (E) SPED (E) M (E) W (E) B (E);
classes = c(3);
cluster=st;
missing are all(-999);
Analysis: type= twolevel mixture;

!Algorithm=integration;
!integration=10;
!mconvergence=0.01;
!miterations=5000;
processors = 8;
!starts=100 20;
!Stseed=256077;
!stseed=415931 ;

Model:
%within%
%overall%
```

```
Iw1 Sw1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Iw2 SW2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Iw3 SW3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;

!sw1@0; sw2@0; sw3@0;
dbr1 (1);
  dbr2 (1);
  dbr3 (1);
  dbr4  (1);
  dbr5 (1);
  dbr6 (2);
  dbr7 (2);
  dbr8 (2);
  dbr9 (2);
  dbr10 (2);
  dbr11 (3);
  dbr12 (3);
  dbr13 (3);
  dbr14 (3);
  dbr15 (3);

%c#1%

iw1; iw2; iw3;

%c#2%
iw1; iw2; iw3;
```

```
%c#3%
iw1; iw2; iw3;

%between%
%overall%
Ib1 Sb1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Ib2 Sb2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Ib3 Sb3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;
 ! DBR7@0

  !sb1@0; sb2@0; sb3@0;

%c#1%

ib1; ib2; ib3;

%c#2%
ib1; ib2; ib3;

%c#3%
ib1; ib2; ib3;

Savedata:
file is model3c3.dat;
save=cprob;
plot:
series=dbr1-dbr15 (*);
type=plot3;
```

Title: DBR Multilevel Model 4

```
Data: File is DBR.dat;
Variable:
Names are id sc grade st DBR1-DBR15
BESST1 BESST2 BESST3 timebess1 timebess2 timebess3
M W B R SPED EBD BEH Abs Sus Exp ODR Spedss EBDSS BEHSS risk SErisk
Brisk erisk;
Usev dbr1-dbr15 st; ! risk ;
idvariable=id;
useobservations (grade eq 7 or grade eq 8);
!auxiliary = risk (E) ODR (E) SPED (E) M (E) W (E) B (E);
classes = c(3);
cluster=st;
missing are all(-999);
Analysis: type= twolevel mixture;

!Algorithm=integration;
!integration=10;
!mconvergence=0.01;
!miterations=5000;
processors = 8;
!starts=40 8;
Stseed=94312;
!stseed=415931 ;

Model:
%within%
%overall%
```

```
Iw1 Sw1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Iw2 SW2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Iw3 SW3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;
iw1; iw2; iw3;
sw1@0; sw2@0; sw3@0;
iw1 with iw2;
iw1 with iw3;
iw2 with iw3;
%c#1%
   dbr1 (1);
   dbr2 (1);
   dbr3 (1);
   dbr4  (1);
   dbr5 (1);
   dbr6 (2);
   dbr7 (2);
   dbr8 (2);
   dbr9 (2);
   dbr10 (2);
   dbr11 (3);
   dbr12 (3);
   dbr13 (3);
   dbr14 (3);
   dbr15 (3);

  %c#2%
    dbr1 (4);
   dbr2 (4);
```

```
 dbr3 (4);
 dbr4  (4);
 dbr5 (4);
 dbr6 (5);
 dbr7 (5);
 dbr8 (5);
 dbr9 (5);
 dbr10 (5);
 dbr11 (6);
 dbr12 (6);
 dbr13 (6);
 dbr14 (6);
 dbr15 (6);

%c#3%
  dbr1 (7);
 dbr2 (7);
 dbr3 (7);
 dbr4  (7);
 dbr5 (7);
 dbr6 (8);
 dbr7 (8);
 dbr8 (8);
 dbr9 (8);
 dbr10 (8);
 dbr11 (9);
 dbr12 (9);
 dbr13 (9);
```

```
  dbr14 (9);
  dbr15 (9);

%between%
%overall%
Ib1 Sb1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Ib2 Sb2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Ib3 Sb3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;
 ! DBR7@0
sb1@0; sb2@0; sb3@0;
ib1; ib2; ib3;

ib1 with ib2;
ib1 with ib3;
ib2 with ib3;

!Savedata:
!file is model5class3v2.dat;
!save=cprob;

plot:
series=dbr1-dbr15 (*);
type=plot3;
```

Title: DBR Multilevel Model 5

```
Data: File is DBR.dat;
Variable:
Names are id sc grade st DBR1-DBR15
BESST1 BESST2 BESST3 timebess1 timebess2 timebess3
M W B R SPED EBD BEH Abs Sus Exp ODR Spedss EBDSS BEHSS risk SErisk
Brisk erisk;
Usev dbr1-dbr15 st ;
idvariable=id;
useobservations (grade eq 7 or grade eq 8);
!auxiliary = risk (E) ODR (E) SPED (E) M (E) W (E) B (E);
classes = c(3);
cluster=st;
missing are all(-999);
Analysis: type= twolevel mixture;

!Algorithm=integration;
!integration=10;
!mconvergence=0.01;
!miterations=5000;
processors = 8;
!starts=40 8;
!Stseed=107446;
stseed=415931 ;
Model:
%within%
%overall%
Iw1 Sw1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
```

```
Iw2 SW2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Iw3 SW3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;
iw1; iw2; iw3;
sw1@0; sw2@0; sw3@0;
%c#1%
  dbr1 (1);
  dbr2 (1);
  dbr3 (1);
  dbr4  (1);
  dbr5 (1);
  dbr6 (2);
  dbr7 (2);
  dbr8 (2);
  dbr9 (2);
  dbr10 (2);
  dbr11 (3);
  dbr12 (3);
  dbr13 (3);
  dbr14 (3);
  dbr15 (3);
  iw1;
  iw2;
  iw3;
  iw1 with iw2 (10);
  iw1 with iw3 (11);
  iw2 with iw3 (12);

 %c#2%
```

```
 dbr1 (4);
dbr2 (4);
dbr3 (4);
dbr4  (4);
dbr5 (4);
dbr6 (5);

dbr7 (5);
dbr8 (5);
dbr9 (5);
dbr10 (5);
dbr11 (6);
dbr12 (6);
dbr13 (6);
dbr14 (6);
dbr15 (6);
  Iw1;
iw2;
iw3;

  iw1 with iw2 (13);
iw1 with iw3 (14);
iw2 with iw3 (15);
%c#3%
  dbr1 (7);
dbr2 (7);
dbr3 (7);
```

```
   dbr4  (7);
   dbr5 (7);
   dbr6 (8);
   dbr7 (8);
   dbr8 (8);
   dbr9 (8);
   dbr10 (8);
   dbr11 (9);
   dbr12 (9);
   dbr13 (9);
   dbr14 (9);
   dbr15 (9);
     Iw1;
   iw2;
   iw3;
  iw1 with iw2 (16);
   iw1 with iw3 (17);
   iw2 with iw3 (18);
%between%
%overall%
Ib1 Sb1| dbr1@0 DBR2@1 DBR3@2 DBR4@3 DBR5@4;
Ib2 Sb2| DBR6@0 DBR7@1 DBR8@2 DBR9@3 DBR10@4;
Ib3 Sb3| DBR11@0 DBR12@1 DBR13@2 DBR14@3 DBR15@4;
 ! DBR7@0

  sb1@0; sb2@0; sb3@0;
```

```
 ib1 with ib2;
    ib1 with ib3 ;
    ib2 with ib3 ;

Savedata:
file is model5class3.dat;
save=cprob;

plot:
series=dbr1-dbr15 (*);
type=plot3;
```