ABSTRACT

Title of dissertation:	EFFECTS OF UNMODELED LATENT CLASSES ON MULTILEVEL GROWTH MIXTURE ESTIMATION IN VALUE-ADDED MODELING

Futoshi Yumoto, Doctor of Philosophy 2011

Dissertation directed by:	Professor Gregory R. Hancock
	Professor Robert J. Mislevy
	Department of Measurement, Statistics and Evaluation

Fairness is necessary to successful evaluation, whether the context is simple and concrete or complex and abstract. Fair evaluation must begin with careful data collection, with clear operationalization of variables whose relationship(s) will represent the outcome(s) of interest. In particular, articulating what it is in the data that needs to be modeled, as well as the relationships of interest, must be specified before conducting any research; these two features will inform both study design and data collection. Heterogeneity is a key characteristic of data that can complicate the data collection design, and especially analysis and interpretation, interfering with or influencing the perception of the relationship(s) that the data will be used to investigate or evaluate. However, planning for, and planning to account for, heterogeneities in data are also critical to the research process, to support valid interpretation of results from any statistical analysis. The multilevel growth mixture model is a new analytic method specifically developed to accommodate heterogeneity so as to minimize the effect of variability on precision in estimation and to reduce bias that may arise in hierarchical

data. This is particularly important in the Value Added Model context – where decisions and evaluations about teaching effectiveness are made, because estimates could be contaminated, biased, or simply less precise when data are modeled inappropriately. This research will investigate the effects of un-accounted for heterogeneity at level 1 on the precision of level-2 estimates in multilevel data utilizing the multilevel growth mixture model and multilevel linear growth model.

EFFECTS OF UNMODELED LATENT CLASSES ON MULTILEVEL GROWTH MIXTURE ESTIMATION IN VALUE-ADDED MODELING

by

Futoshi Yumoto

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
      Professor Gregory R. Hancock, Co-Chair
      Professor Robert J. Mislevy, Co-Chair
      Professor George Macready
      Professor Robert Lissitz
      Professor Paul Hanges

Dedication

This dissertation is dedicated, in part, to my mother Atsuko Yumoto and to my father Kazuaki Yumoto, who wholeheartedly supported me in every endeavor and whose faith and encouragement have continued to push me to learn and grow. I was unable to complete the work and this program while my mother was alive but I hope and believe the final result would have pleased her. This work is also dedicated, in part, to my brother Ryogo Yumoto, who sacrificed his opportunity to move to the United States so that I could do it. My life here has been rich and rewarding, and so I am profoundly grateful to him for making this experience – and essentially, this research, possible. Finally, I want to dedicate this work to my wife, Rieko, and daughters, Sora and Lena, and to Lin and Toto. While the PhD dissertation is a significant intellectual achievement, what it really represents is the end of my attention being divided so completely between research and real life.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Fairness is necessary to successful evaluation. Evaluation can be based on simple measurement of the weights of things for comparison, or based on measuring something as abstract as a single effect within a complex system such as the effect of teachers upon the progress of students' performance. Fairness in evaluation requires that all subjects are measured without a bias (i.e., the evaluation score reflects the true property of subjects) and accurately (i.e., repeated measures yield the same results). A single evaluation process that can produce favorable results to one group and penalize another group when measured identically is unfair. Fairness is essential to *both* simple and complex evaluation. A simple evaluation example is to compare the average weights of students in two classrooms, derived from the weights of individuals measured on a pair of scales. A complex evaluation example, and the focus of this research, is the value-added model (VAM; Sanders & Rivers, 1996), which is an evaluation of teachers based on a statistical estimate of student performance gains that are attributed to the effect of the teacher.

Statistical models – irrespective of their complexity – are always simplifications of the data they represent: when summarizing the weights of students in classrooms, the mean value is a (very simple) model that collapses over the distribution (Miles & Shevlin, 2000), thereby masking features such as whether the distribution is multimodal, whether there are outliers, and so forth. For example, if one classroom has more males, and this was not incorporated into the estimate of the mean (naturally resulting in a more complex summary of the classroom's weight), then one room could appear to have a higher average weight but in fact the reasons for differences between the groups' weights, a model which does not take the sex of students into account produces bias. Similarly,

when applying the VAM approach to estimating teacher effects on student performance, assumptions are made that might affect the estimates or their interpretation/interpretability. The goal of this research is to assess the impact on VAM analyses of the assumption that all students are equally affected by the teacher's effect within a classroom – that is, that improvement of student performance due to the teacher does not vary systematically across students in the classroom. If there is systematic (as opposed to random) variation in the teacher effect within a classroom, and this leads to incorrect estimates of the teacher's overall effect, then this VAM assumption is not supportable, that is, if unaccounted for in the VAM analysis, this heterogeneity (of teacher effect on students) could bias the estimates of teacher effects and result in an unfair evaluation procedure.

In the classroom weights example, the comparison will not be fair if it involves two scales and one scale always shows a higher weight than the other scale for any given student. One scale is *biased* in this example. The other scale might have higher variation in its measurement (i.e., the scale shows wider variation of values for the same subject). This scale has less *precision* or, equivalently, higher *error* in this case. It is relatively easy to control the issues of bias or precision on the scales in this example. The example can be made more complex by adding other factors such as the proportion and males and females in each classroom, as noted above. If the groups being compared differ in a systematic way, they are not strictly comparable. The evaluations of students and teachers pose similar challenges, in terms of isolating the effects of interest while taking all sources of bias into account. The importance of fairness is naturally far greater since decisions and even policies are made based on these estimates. The VAM approach has

many advantages over simpler models, because it permits the inclusion of multiple sources of bias and heterogeneity that could affect VAM-derived estimates.

The "value-added" model is commonly implemented in multi-level modeling frameworks so as to capture the contribution of higher-level effects such as teachers (level 2) and schools (level 3) on the student's achievement and/or improvement (Sanders & Rivers, 1996). As with the proportions of genders in the previous example, there are several factors, such as student ethnicity, socio-economic status, previous performance level, or classroom size to control so as to minimize the systematic bias and errors, deriving from the classroom or school, that can affect the estimation of a particular teacher's "value" added.

A common assumption for this type of teacher evaluation method is that the teacher's effect is constant within a classroom. In other words, all students are assumed to have received the same contribution or benefit from the teacher, so the teacher's effect on students is homogeneous. This assumption may be unreasonable for a student with a minor and undiagnosed disability (e.g., minor learning disability), a student who has no interest in education, or a student who lives in such conditions that study cannot be a priority (e.g., lack of food or safety in life). In fact, this assumption is unrealistic. There are students who are unmotivated, who have different priorities other than focusing on studies, or simply who unable to understand the instruction. These students do not receive any benefit from the teacher no matter how effective or ineffective she or he is. These, and other, unanticipated or unknown sources of heterogeneity can contribute bias and/or imprecision to estimates. As discussed in Chapter 2, a recent study of persistently low performing (PLP) students (Lazarus et al., 2010) strongly indicates that there is a group

of students who do not receive benefit from traditional classroom instruction. The measure of a teacher's effect is likely to be different when there are different proportions of different types students in each class; the presence of a group of class of students within a classroom represents a systematic effect that should be accounted for in the model and estimate. It is not entirely fair for teachers to be responsible for students' improvement if the majority of students are not interested in learning. Fairness in evaluation of the instructor effect on student performance or gains cannot be established in this case unless the effects from such students are either negligible or adequately controlled in the evaluation procedure; simply *assuming* that they are is insufficient. The primary focus of this research is to investigate the impact of ignoring the non-performing classroom group in the last example on the evaluation of teacher's effect on students' gain in test scores, focusing on the bias and precision of the estimated teacher's effect.

This study is a simulation motivated by the situation where teacher effect on student performance must be measured so as to evaluate the teacher's quality, performance, or achievements. In this situation, we assume that there are two types of students in the classroom: fast and slow growers (or students with fast or slow growth profile) in terms of the skills they are being taught, represented by both gains on standardized test scores (slope) and initial achievement level (intercept). Figure 1 shows a graphical representation of students' growth profiles in each of these two groups (the actual slopes for particular students vary around these two lines.) Each classroom may have different proportions of students with each growth profile, and that is represented in this simulation in order to determine whether unknown, or unmodeled, heterogeneity in student type (based on proportion of students with each growth profile) – which is

inconsistent with the VAM assumption that all students get the same effect from a given teacher – affects VAM-based estimates. An additional challenge for VAM-derived estimates is that some teachers may actually facilitate the transition of students from the slow growth group to the fast growth group. This would have a substantial impact on students but may not be reflected within the current value added evaluation context, at least in the short term. This aspect of the VAM approach is beyond the scope of this study, but represents additional aspects teacher effects that should be evaluated for the fairest estimates of their quality, performance or achievement.
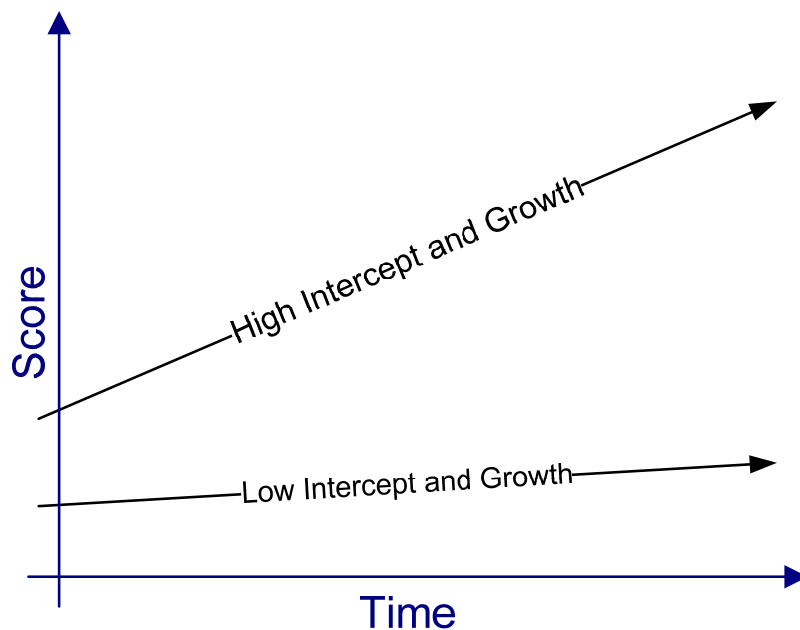
Figure 1. Graphical representation of growth profile

To illustrate the study design, the simulation features are consistent with, for example, a school with six 8th grade classrooms: classrooms A, B, C, D, E and F, each with 40 students. Since there are six classrooms, in this example (but not in the simulation design which are 30, 60, or 90), cluster number (CN) is six. Figures 2 through 4 illustrate such a school, with classes A to F and cluster size (CS) of 40.

There are three levels of classroom, high achieving classes A and B, moderately achieving classes C and D, and low achieving classes E and F. These three levels are shown as the cluster type 1, 2, or 3 in Figure 2. Each cluster type has an equal number of classes, two classes (of CN=6 in this example) per cluster, as shown in Figure 2.

Cluster-Type

	1	2	3
Cluster Size (CS)	1/3 of CN (Classes E & F)	1/3 of CN (Classes C & D)	1/3 of CN (Classes A & B)

Figure 2. Graphical representation of simulation sample: Cluster size by cluster type

With this example, imagine that the proportion of students in the two growth groups is different among three types of classrooms, resulting in the overall achievement level of that classroom. High achieving classes A and B have all 40 students in fast growth group. Moderately achieving classes C and D have 75% (30/40) students in the fast growth group and 25% (10/40) students in the slow growth group. In low achieving classes E and F, each growth group makes up 50%, or 20 students. Figure 3 shows the different proportion of students in each growth profile, or the mixture proportion (MP), in each cluster type (cluster type 1 is the low achieving because it is a mixture containing mostly low growth students, cluster type 2 is the moderately achieving because it containing lower proportion of low growth students, and cluster type 3 is the high achieving group because they are mixtures with highest proportion of high growth students).

Cluster Type

	1	2	3
Cluster Size (CS)	Low Growth	Low Growth	
	High Growth	High Growth	High Growth

Figure 3. Graphical representation of simulation sample: mixture proportion

We are positing for the sake of this study that only students in the fast growth group receive any benefit of instruction from teachers –that is, the teacher's effect is zero for students in low intercept/growth group. Figure 4 shows the teacher's effect based on the growth profile and mixture proportion. The teacher's effect is the same as the cluster effect (CE) in this study

Cluster Type

	1	2	3
Cluster Size (CS)	No Cluster Effect	No Cluster Effect	
	Cluster Effect	Cluster Effect	Cluster Effect

Figure 4. Graphical representation of simulation sample: Cluster effect

In this scenario, it is very difficult for teachers with low achieving classes to obtain a high value added score as compared to teachers with high achieving class – even if they have add identical value compared with teachers in high achieving classes – because the expected average teacher effect is attenuated by the group of students who are not responsive to any instruction. In other words, teachers are penalized, in terms of the estimation of their effectiveness, by the kinds of students they have in the classroom under the assumption that all students receive the same benefit from the instruction.

7

Fairness in teacher evaluation cannot be established without accounting for the growth profile (type) of students.

This study systematically investigates the bias in estimating teacher effects when student type is unmodeled, that is, under VAM assumption that the students receive a homogeneous (randomly, not systematically, varying) effect from the teacher, by manipulating conditions identified in the example above, including cluster number (CN) which was fixed at 6 in Figures 2-4 but which varies as described below and more extensively in Chapters 2-4. The growth profiles are consistent throughout the study: high mean score or intercept and high growth rate or slope for the fast growth group, and the low mean score and low growth rate for the slow growth group. The actual parameters are explained in Chapter 3. This study manipulates conditions used in the 8th grade school example above, including the class size, number of classes in a school, proportions of students in each growth group, and teachers with different effects. The study tried to identify conditions, and/or interactions among conditions, which are plausible or empirically established, and which have the greatest potential to cause unfairness in evaluation.

There are four simulation conditions to manipulate as illustrated in Figures 2 through 4.

- Cluster number (CN) is the number of clusters (e.g., classes) in the sample (e.g., school).

- Cluster size (CS) is the number of students in a cluster (e.g., 40 students in a classroom).

- Mixture proportion (MP) is the proportion of students in each growth profile within a cluster (e.g., 100%, 75%, or 50% of students in fast growth group in three levels of classes).

- Cluster effect (CE) is a cluster's true effect (i.e.. an individual teacher's effect) on the students in a cluster (e.g., classroom).

This study systematically manipulated these four simulation conditions to investigate the bias and precision in an estimated cluster effect (or teacher's effect) when estimated with or without accounting for the heterogeneity (i.e., mixture proportion) in data. The goal of the study is to identify if there are substantial, systematic biases in the cluster effect estimates in any simulation conditions that would make fair evaluation difficult, if not impossible. The simulation conditions of this study are still much simpler than the real world; as noted earlier, statistical models – irrespective of their complexity– are always simplifications of the data (or the systems) they represent – and so are simulation studies. However, the study was designed to determine if the VAM approach and specifically, the assumption of a homogeneous, or randomly varying, teacher effect for all students is reasonable or not. The simulation study is described completely in Chapter 3.

1.1 Heterogeneity and estimation

Heterogeneity of the student population or distribution is a key characteristic of data that can complicate evaluation design and especially, analysis and interpretation. Heterogeneity can either be random or systematic. Random variability is what makes estimation necessary, otherwise there would be a single value to summarize any effect or system. *Systematic* variability is the heterogeneity that makes estimation complex, because as noted earlier it leads to bias and imprecision in estimation if it is not included

9

in the analytic model. In the context of VAM-derived teacher effects, the systematic variability, or heterogeneity among the students belonging to high and low growth groups as described in Figures 2-4, represents a clear violation of the assumption that student growth patterns are homogeneous (i.e., vary only randomly) within a classroom. Therefore, if student growth patterns vary systematically (i.e., are heterogeneous) within classrooms, the variability attributed to the unmodeled heterogeneity inflates variability higher up in the model (e.g., at level-2, classroom/teacher), causing mis-estimation and even mis-interpretation of the teacher's effect. When it is assumed that students are homogeneous, or that the teacher's effect on the students varies only randomly across students, then any actual heterogeneity represents unknown or uncontrolled sources of variability in the system – violating the VAM assumption.

It is crucial to identify the relationship(s) that any dataset will be used to investigate, and what it is in the data that needs to be modeled, before conducting any research; specification of these two features will inform both study design and data collection, shaping the research design and/or hypothesis. However, planning for, and planning to account for, heterogeneities in data are also critical to the research process, to support valid interpretation of results from any statistical analysis. This study sought to investigate the effects of un-accounted for heterogeneity at student level (i.e., level 1) on the precision of teachers' effect (i.e., level-2) estimates in multilevel data.

The importance of modeling variability explicitly is reflected in both empirical studies (e.g., Clogg & Goodman, 1985; Goodman, 1974; Henry & Muthén, 2010; Jo, 2002; Kreuter & Muthén, 2008; Lambert, 1992; Lazarsfeld & Henry, 1968; Muthén & Shedden, 1999; Nagin, 1999; Nagin & Land, 1993; Samuelsen, 2005) and

methodological developments (e.g., Asparouhov & Muthén, 2008; Bartholomew & Knott, 1999; Bollen & Curran, 2006; Goodman, 1974; Lazarsfeld, 1950; McLachlan & Peel, 2000; Muthén, 2001; Nagin, 1993; Quandt, 1958; Quandt & Ramsey, 1972; Raudenbush & Bryk, 2002; Skrondal & Rabe-Hesketh, 2004; Titterington, Smith & Makov, 1985; Verbeke & Lesaffre, 1996; Vermunt & Van Dijk, 2001). Methodological work has benefitted from and expanded to accommodate and model, the influence of both observed (manifest) and unobserved (latent) variables on the estimation and interpretation objectives of multivariate statistical analysis (Loehlin, 1998). Software like M*Plus* (Muthén & Muthén, Ver 6.1, 2010) and Latent Gold (Vermunt & Magidson, Ver 4.5, 2010) has been both developing, and supporting, the capacity of investigators to consider and analyze manifest and observed contributors to heterogeneity in their data (e.g., Feldman, Masyn & Conger, 2009; Henry & Muthén, 2010; Jo, 2002; Kreuter, Yan, & Tourangeau, 2008; Marsh et al., 2009; Preacher, Zyphur, & Zhang, 2010; Schaeffer et al., 2006; Van Horn et al., 2009). As is explicated in Chapter 2, estimates from statistical models can vary depending on whether manifest and/or latent variables are modeled (see, e.g., Hancock & Lawrence, 2006; Muthén & Asparouhov, 2009) – and particularly whether these are modeled appropriately or not (e.g., Chen et al., 2010; Palardy & Vermunt, 2010). Since the estimates can vary in relation to these features, so, too can the inferences based on those estimates.

1.2 Multivariate analytic methodological innovations for heterogeneity and precision

Multivariate methods have been developing and evolving with increasing, and increasingly sophisticated, mechanisms for modeling both the heterogeneity in data and the actual relationships under study. A recent development is the multi-level model,

which has become increasingly widespread in educational research. In 1972, Lindley and Smith published the first multi-level model – in *Journal of the Royal Statistical Society* (Lindley & Smith, 1972), which was developed to accommodate heterogeneity in the individual that detracted from the precise estimation of the group-level parameter of interest; random effects can include variation in group level parameters (e.g., group level mean/intercept), or degree of individual mean deviation from the overall group-level mean. The terms *multilevel* or *hierarchical* describe the situation where sets of observations are treated as levels, hierarchically nested within other sets or levels, such as a students nested within schools (Nezlek & Zyzniewski, 1998). Verbeke and Molenberghs (2000) describe the specification of random effects as the second of a two-stage modeling method ("general linear mixed modeling", Ch. 3); considering their "stages" as <u>levels</u> corresponds to a multi-level model. The random effect represents an additional level of analysis, so that regression coefficients become random variables; with observations nested within, for example, individuals (for whom a single constant regression coefficient would be estimated). The multilevel model is generally used to account for the interdependence of individuals within the same group and model the effects of both individual-level and group-level variation (i.e., heterogeneity) on an outcome simultaneously (Pollack, 1998).

Burstein, Linn, and Capell (1978) utilized multi-level data analysis to accommodate the presence of heterogeneity in regression estimators across classrooms within a single sample*.* Other investigators have focused on the bias introduced into estimation when within-group correlations are, or are not, explicitly accounted for within modeling (see Kreft & de Leeuw, 1998). The treatment of data as explicitly hierarchical,

with observations at one level (e.g., at the individual level) nested within other levels (e.g., the group or class level) depends critically on how the levels and hierarchy are described and defined (see Kreft et al., 1995), and this is true with random effects, mixed effects, or multi-level models. Verbeke and Molenberghs (2000) considered observations, repeated over time, to be nested within individuals. Explicit modeling of the hierarchies in data is sometimes called hierarchical linear modeling (Raudenbush & Bryk, 2002). To maintain generality, we refer to this type of model as a multi-level model (MLM).

Considering scores nested within students nested within schools makes the estimated student level (level 1) effects of pre- on post-test scores more precise and less biased (Raudenbush & Bryk, 2002). That is, by planning for and accommodating the heterogeneity arising from specific features in the data, the effect of variability on the estimates can be minimized. For example, if students within a classroom are more homogeneous (random variation is lower) than the overall student population, accounting for clustering of data (e.g., modeling students as if they are nested within a classroom) *reduces* overall error. In this example, accommodating the lower level of variability within this classroom improves precision and reduces bias in estimates based on this classroom by reducing error. Muthén and Asparouhov (2009) investigated the source of heterogeneity in multilevel data by treating a mathematics achievement score as level 1 (student level), and student scores as nested within a school (level 2). Their two-level regression (actually, a mixture regression) model showed heterogeneous residual variance varying across level-1 *covariates* (p. 649), indicating the presence of additional, unaccounted-for variability at level 1 – going beyond the multi-level modeling approach to accommodating variability. In fact, Muthén and Asparouhov (2009) found that the

effects of level-1 covariates were different for estimates of both level-1 and level-2 effects on the dependent variable when their additional variability was modeled at level 1; thus, the explicit inclusion of covariates at level 1 had a significant impact on their results and interpretation. This finding, further described in Chapter 2, underscores the importance of the careful investigation of the sources of heterogeneity in multilevel analysis –beyond the simple nesting effects of student and school in the "conventional" two-level model. Muthén and Asparouhov (2009) concluded that the conventional two-level model, with effects estimated separately for student and for school, could not effectively eliminate the effects of level 1 heterogeneity on the estimation of level 2 effects. This implies that simply using a multi-level modeling approach may be insufficient for valid modeling and interpretation of results, because variability at level 1, if unaccounted for, could affect estimates of level 2 effects. Thus, estimating student scores via a multilevel regression will be less precise *and* will be biased, if this conventional two-level model is used but unmodeled student-level covariates are actually contributing to the heterogeneity underlying within-school correlations among student scores. This is described more fully in Chapter 2.

1.3 Consideration of latent covariates and the general mixture model

The covariate used to account for the variability at level 1 in Muthén and Asparouhov's (2009) analysis represents a latent class. As noted, Muthén and Asparouhov (2009) found that the conventional multilevel model was insufficient to yield precise estimates of level 2 effects; their solution was to utilize latent covariates, inferred from the data, because none of the manifest covariates had any explanatory power. In fact, the latent covariate used to account for the variability at level 1 in Muthén and

Asparouhov's (2009) analysis represents a latent class. Magidson and Vermunt (2004) describe a latent class as some factor causing "…some of the parameters of a postulated statistical model <to> differ across unobserved subgroups," (p. 175) where categories of subgroups of this unobserved or latent categorical variable make up the levels of the latent class (LC). Therefore, a LC is a subgroup indicator, like a covariate, but it is latent and must be *inferred from data*. An example of a latent class model, first noted by Lazarsfeld (1950), includes the classification of applicants into subgroups (e.g., acceptance and rejection groups for uniformed services recruits), that were estimated from a set of dichotomous responses on a questionnaire (see also Lazarsfeld & Henry, 1968). That is, the classification of applicants was not based on any observed data, but based on their dichotomous responses, the latent (unobserved) classes into which the applicants were sorted were inferred.

An example of the development and growing support of the capacity of investigators to consider and analyze both manifest and observed contributors to heterogeneity is a family of methods called "mixture models" (Muthén, 2002). Verbeke and Molenberghs (2000) refer to a "mixture" as a regression that includes both random and fixed effects. However, in the more general context (as described in Muthén, 2002), mixture models are a type of statistical method used to conduct an analysis while simultaneously examining if there is more than one sub-population (e.g., at least two subgroups with different distributions) in data (e.g., Muthén, 2002; Magidson & Vermunt, 2004).

Mixture models (in this more general sense) have been applied in research domains as diverse as organization (Lazarsfeld, 1950), education (Dayton, 1991), and

medicine (Rindskopf & Rindskopf, 1990). In each case, some analytic method (e.g., linear regression) is the objective, but subpopulations in the data may warrant different regression features.

The most general mixture model can be defined as analysis that includes the search for latent subpopulations while simultaneously estimating statistical models including several causal effects, which is beyond straightforward multiple regression. For example, multilevel, structural equation, growth, and the combination of these types of modeling approaches fall under "general mixture models" (see Bartholomew, 1987; Muthén, 1989; Muthén, 2001; Skrondal & Rabe-Hesketh, 2004; Vermunt & Magidson, 2002). Latent class analysis and finite mixture modeling (McLachlan & Peel, 2000) are technically subsumed within mixture modeling, as they are very specific types of mixture models. This most general formulation of mixture models, which we refer to as "the general mixture model approach" comprises models ranging from simple estimation of a latent class or a finite mixture, through less complex models with simultaneous latent class or finite mixture evaluation, to highly complex modeling such as latent growth plus latent class/finite mixture combinations.

The general mixture model approach has the potential to completely reshape how educational research is done. For example, with general mixture models one can both identify differential patterns of growth in a group of students while simultaneously identifying the subgroups within the sample for which targeted interventions (e.g., different types of instruction) can be tailored. In educational research, manifest variables such as socio-economic status (SES, low/high) are often important covariates, but these should not be confused with latent class (LC) variables. Less general models (e.g., latent

class or finite mixture models) cannot serve these purposes because the primary focus of the less general models is to identify the latent class from the set of observed categorical or continuous variables, instead of permitting the identification of such classes from estimates derived from other simultaneous analyses (see, e.g., Goodman, 1974; Hagenaars & McCutcheon, 2002; Muthén, 2000; Nagin, 1999).

It is important to note that the latent class analysis is a valid and useful analytic method in research where the identification of latent classes is the primary focus. In the present context, however, the latent classes represent a complicating feature of the estimation (of teacher effect), introduced with the intention of reducing bias, and are not an end in themselves.

Exemplifying this potential, Muthén and Asparouhov (2009) used a multilevel mixture model, instead of the conventional multilevel model, where subgroups of students were identified within the latent variable "student type" with levels "fast learner" or "slow learner". This student level (level 1) latent class variable (LC) accounted for the heterogeneity in level 1 residual variance that was unaccounted for by observed covariates or the conventional multilevel model; the mixture model that included this LC also identified effects which were estimated at the school level (level 2), ultimately changing the estimated effects of covariates at both student and school levels, and leading to different interpretations of parameter estimates than was supported by the conventional two-level model. They also tested for the presence of a LC at the school level and found that, although such a level 2 LC could be identified, it had a very limited impact upon the estimation or interpretation of other parameters. Muthén and Asparouhov's (2009) example showed the importance of thorough investigation of heterogeneity in variance at

each level and in particular, that the conventional multi-level model will not always suffice to limit bias and optimize precision of estimates.

1.4 The general mixture model and latent growth curves in growth mixture models

Just as hierarchies in data led to multivariate methodological developments such as the multi-level model, individual effects in intercepts and slopes of repeated measures datasets led to the development of the latent growth curve model (or growth/growth curve model, e.g., see Preacher, Wichman, MacCallum, & Briggs, 2008). The purpose of growth models is to model change over time with particular emphasis on the variability in starting points (i.e., intercepts) and/or change over time (i.e., growth/slope). A latent growth curve mixture model or growth mixture model (GMM) is an extension of latent growth curve model (LGM). The idea behind GMM is to permit further examination – and estimation – of the heterogeneity of growth trajectories that may be explained by latent classes. For example, there may be groups of students with distinctive growth trajectories that cannot be explained well by one set of slopes, intercepts, and their correlations. As noted earlier, accounting for heterogeneities in data is critical to support valid interpretation of results from statistical analysis. The inclusion of slopes and intercepts (growth curve modeling) as multiple levels (multi-level modeling), plus identification of important covariates such as student type (mixture modeling) are united in the estimation underlying the growth mixture model.

The multilevel extension of GMM was recently introduced and has been applied in education (e.g., Muthén & Asparouhov, 2009; Palardy & Vermunt, 2010). The Muthén and Asparouhov (2009) example outlined above can be generalized to other educational outcomes like the evaluation of teacher effectiveness – which would typically be

estimated using a value added model (VAM; McCaffrey et al., 2003; Sanders & Rivers, 1996). As will be explained in Chapter 2, the VAM is actually a special case of the GMM; suggesting that growth/growth mixture modeling is a natural tool for estimating the development of student capabilities over time – as well as other effects (e.g., teacher and school) that could be – and may need to be shown to be – contributing to students' growth.

1.5 Multilevel growth mixture modeling supporting precise estimation and inference

The heterogeneity in data that obscures, or diminishes the precision of estimates of, parameters or relationships of interest can either be ignored (leading to imprecise/biased and possibly incorrect estimates) or modeled explicitly (also possibly leading to incorrect estimates if the modeling is not appropriate). Analytical developments have included finite mixture/latent class, multilevel, growth curve, mixture, and multilevel growth mixture modeling approaches, as described above. Each of these developments addresses previously-unaccounted for heterogeneity in data and precision in estimates. Similarly, the goals of this research address the potential impact of unaccounted-for heterogeneity at level 1 (e.g., student level) in the level-2 estimates (e.g., teacher), testing whether the inclusion of this type of heterogeneity merits further consideration for a group-level statistical evaluation procedure, including teacher or school evaluation with VAM.

This research investigated model feature effects on the precision of individual parameter level estimates at level 2 of a multi level growth mixture model. The goals of this study were to: (1) investigate the bias and precision of level-2 parameter estimates in

the multilevel model affected by incorrectly modeled level 1 effects; (2) the effectiveness of information criteria to identify the true number of latent classes in MLGMM.

These were accomplished via a simulation study, described more fully in Chapter 3. Representing an educational study context, this simulation focused on the precision/bias and variability of estimation of level 2 (i.e., teacher-level) effects on student performance (level 1), that is,, within a VAM framework. The research also investigated the issue of latent class identification/misidentification which has the potential to cause serious estimation bias at more than one level (Chen et al., 2010).

The dissertation is organized as follows: The different models, their comparisons, contrasts and implications for data and assumptions are described more fully in Chapter 2. Chapter 3 presents the methodology that was used to complete this study. Chapter 4 presents the results from a pilot study supporting the proposed simulation methodology. Results of the study are presented fully in Chapter 5 followed by the discussion of the research in Chapter 6.

Chapter 2: Literature Review

Heterogeneity is a characteristic of data that can complicate research design, analysis and interpretation. As outlined in Chapter 1, the multilevel growth mixture model (MLGMM; Asparouhov & Muthén, 2008; Muthén, 2004) is a new analytic method specifically developed to accommodate heterogeneity so as to minimize the effect of variability on precision in estimation and to reduce bias that may arise in hierarchical data. This is particularly important in the VAM context – where decisions and evaluations about teaching effectiveness are made, because estimates could be contaminated, biased, or simply less precise when modeled inappropriately. Therefore, the research questions for this proposed work are:

1) Are the level-2 parameter estimates in the multilevel model affected (in terms of bias and precision) by incorrectly modeled level 1 effects?

2) What information criteria can be used to identify the true number of latent classes in MLGMM?

To answer these research questions, the following objectives were set for the simulation study:

1) to contrast the estimation (precision and bias) of level 2 effects from a two-level growth model with heterogeneity at level 1 <u>un-modeled</u> (incorrectly specified) and a two-level growth mixture model with heterogeneity at level 1 <u>modeled</u> (correctly specified) in order to investigate the systematic biases associated with inappropriate model specifications in MLGMM and VAM;

2) to investigate the effectiveness of information criteria to identify true models;

3) to examine the accuracy of model identification in MLGMM.

This chapter outlines the motivation and algebraic foundations for the research questions and study design.

2.1 Educational effects estimation with growth curve mixture models

Growth curve, mixture, and multi-level models are all very important in educational research (e.g., Boscardin et al., 2008; Muthén et al., 2003) and their combination, the growth curve mixture model, has also been promising (Muthén, Asparouhov, 2009; Palardy & Vermunt, 2010). Growth mixture models have also been applied in different fields including preventive intervention (e.g., Muthén et al., 2002), criminology (e.g., Kreuter & Muthén, 2008; Schaeffer et al., 2006), epidemiology (e.g., Croudace et al., 2003), and substance abuse (e.g., Boscardin et al., 2008). A common theme in this body of research is to identify latent classes from growth trajectories that are both substantively and statistically distinct. Thus, the results become more informative in that specific strategies (e.g., interventions) can be formulated for each level of the so-identified latent class. For example Muthén et al., (2002) report that the effect of drug treatments were found not to differ statistically for the experimental group as compared to the placebo group in a placebo-controlled clinical trial; this failure to achieve significance was later determined to have been due to non-compliance (i.e., study subjects did not follow directions or take drugs as prescribed). Without the mixture approach, this analysis – a conventional multilevel model – would have led to the conclusion that the drug under study did not work better than placebo. However, when this non-compliance was included as a latent class (was/ was not compliant, inferred from data) variable in their model, then the drug effects (relative to placebo) were identified (and estimated) in the compliant subgroup of the active arm, and additional design

features were uncovered for future clinical trials (i.e., to specifically encourage compliance in the trial participants). Another benefit of the mixture analysis in this example was the improved precision of the estimated drug effect, derived from the compliant group versus the placebo group.

Similarly, in an educational context, students may be classified into meaningful groups based on differences in growth trajectories over time. Identifying patterns of growth specific to different groups has the potential to inform the development of strategies supporting effective instruction for each group of students, whether academic (e.g., alternative instruction), behavioral/psychological (e.g., behavioral intervention), or social (e.g., individual counseling).

In the context of VAM to estimate teacher effectiveness, some group of students (i.e., in one subgroup of the student-level latent class variable) might not receive any contribution from teachers, similar to the situation for the non-compliant group in the clinical trial example above; this subgroup of students may bias the estimates of teacher effect, whereas the teacher effect might be more precisely estimated among students who <u>do</u> benefit from the teacher (similar to the compliers in the clinical trial example). This study investigates a situation similar to this example, where there is a differential *level-2 effect* on level-1 depending on the class (i.e., latent class) to which a subject in level-1 belongs.

2.2 The problem: Estimation with growth curve modeling

As described in Muthén and Asparouhov (2009), researchers could reach a different conclusion if the level-1 latent class variable was ignored when it actually has a significant impact on estimation of level-2 effects on the dependent variable. This

sophisticated approach to the analysis of educational data is compelling – as evidenced by Muthén and Asparouhov (2009)'s example; but the analytic complexity entails other challenges to be considered in order to support valid inferences. Namely, Palardy and Vermunt (2010) raised two important issues to consider in multilevel growth mixture models. First, one must be extremely cautious with the use of covariates to identify latent classes, since covariates can change the distribution of random effects from which latent classes are identified. Secondly, the choice to include random effects at a higher level or not – in addition to those at the lower level – could affect interpretation and results, because "*the latent class and random effect compete to explain the same variability in the growth trajectory*" (p. 555, emphasis added)*.* This second point of Palardy and Vermunt (2010) is consistent with Muthén and Asparouhov's (2009) findings. Thus, there are many modeling "tricks" that could be brought to bear when heterogeneity is complex and/or unknown; but as noted above, analytic complexities bring their own challenges. Thus, the method has great potential, but this must be balanced against these two particular challenges. Therefore to estimate the impact of ignoring these challenges this simulation study evaluated the precision of level-2 estimates by introducing the level-1 heterogeneity in the form of differential growth trajectories among level-1 subjects within a series of GMMs fit to simulated datasets (described in Ch 3). Including heterogeneity in the growth trajectories will show whether the identification of latent trajectories in MLGMM is really important challenge. Estimation with growth curve modeling improves precision of estimates and accommodates the realistic condition of the nesting of observations within a classroom, but without examining the possibility of heterogeneity among the growth curves, the true potential of the method is unknown.

As described in Chapter 1, linear growth or growth curve modeling is a statistical method that was relatively recently developed. It has become increasingly important in educational research in the past decade or so. For example, the U.S. Department of Education initiated the pilot study, Growth Model Pilot, in 2005 to address the potential/perceived unfairness in Adequate Yearly Progress (AYP) evaluation (No Child Left Behind Act of 2001, sec 6161). The introduction of governmental initiatives such as Growth Model Pilot study (US Department of Education, 2005) and Race to the Top (US Department of Education, 2010), which place a strong emphasis on estimating (and thereby facilitating improvement in) the effectiveness of teachers, reinforces the need for unbiased and precise estimation of performance over time for both students (scores nested within students) and teachers (students nested within teachers). Thus, the methods introduced in Chapter 1 will become increasingly important in educational policy and decision making. As noted earlier, Palardy and Vermunt (2010) concluded that covariates may *arbitrarily* separate variability based on manifest classes in covariates such as ethnicity or socio-economic status; Muthén and Asparouhov (2009) noted that failures to accommodate covariates in growth mixture models can also adversely impact the precision and bias in growth model estimates. Therefore, although growth curve (and related) modeling methods exist and have been used in educational contexts including teacher evaluation, for these methods to be both useful and used appropriately, the impacts of latent classes within the growth curve modeling framework should be better understood, as was done in this study.

As was suggested in Chapter 1, the modeling methods explored in this study are closely related. In fact, growth mixture modeling (GMM; Muthén 2001; Muthén 2004) is

a mixture extension of the latent growth curve or latent growth model (LGM), and MLGMM is a multilevel extension of GMM. Therefore, MLGMM is described in this section to provide background to the simulation structure and to show how GMM and LGM are special cases of MLGMM. Figure 5 shows the hierarchy of the family of latent growth model. VAM is a special case of the multi-level growth mixture model (MLGMM); it is equivalent to a MLGMM where there is only one class, i.e., there is no mixture because everyone is assumed to be in the same class. When MLGMM is used instead of VAM, because it does include latent class estimation, it does not assume that all students are in the same class.

Latent Growth Model Hierarchy

Most
General

Multilevel Growth Mixture Model
(MLGMM)

Growth Mixture Model
(GMM)

Multilevel Latent Growth Model
(MLGMM)

Mixture

Multilevel

Most
Specific

Latent Growth Model
(LGM)
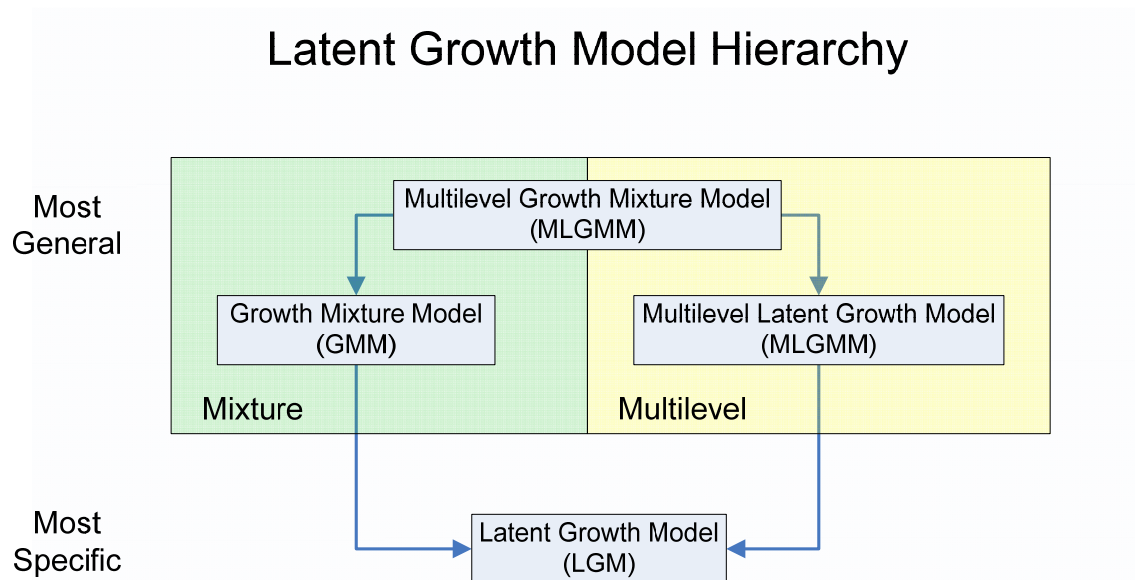
Figure 5. Hierarchy of family of latent growth models

2.2.1   Algebraic representation of Multilevel Growth Mixture Model (MLGMM)

The formulation of MLGMM has two parts: the within-group (i.e., level-1) and between-group (i.e., level-2) models. This formulation includes both within-group level and between-group level latent class variables. This study focused on the estimates from the between-level slope, but the entire formulation is presented below for context. A

cluster represents the unit of the between-group or level-2 identifier in this paper and is used interchangeably with a group.

Level-1

- Within-group level measurement model

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}a_{1tij} + e_{tij}, e_{tij} \sim N(0, \sigma^2) \tag{1}$$

Where π_{0ij} is an intercept for individual i in cluster j, π_{1ij} is a slope for individual i in cluster j, a_{1tij} is covariates at time t for individual i in cluster j, and e_{tij} is an error at time t for individual i in cluster j.

- Within-group level structural model for the intercepts and slopes (Level-1)
  - o Intercepts

$$\pi_{0ij} = \sum_{k=1}^{K} \beta_{00k}c_{kij} + \sum_{m=1}^{M} \beta_{0jk}X_{mij} + r_{0ij} \tag{2}$$

  - o Slopes

$$\pi_{1ij} = \sum_{k=1}^{K} \beta_{10k}c_{kij} + \sum_{m=1}^{M} \beta_{1jk}X_{mij} + r_{1ij} \tag{3}$$

$$r_{ij} \sim N(\mathbf{0}, \mathbf{\Theta_r}) \tag{4}$$

- Model for subjects' latent class memberships, give their covariates

$$\text{logit}[P(c_{kij} = 1)] = \delta_{0k} + \sum_{m=1}^{M} \delta_{mk}X_{mij} \tag{5}$$

Level-2

- Between-group level model
  - o Intercepts

$$\beta_{0ij} = \sum_{l=1}^{L} \gamma_{000l}d_{lj} + \sum_{n=1}^{N} \gamma_{00n}W_{nj} + u_{0j} \tag{6}$$

  - o Slopes

$$\beta_{10j} = \sum_{l=1}^{L} \gamma_{100l}d_{lj} + \sum_{n=1}^{N} \gamma_{10n}W_{nj} + u_{1j} \tag{7}$$

27

$$u_{ij} \sim N(\mathbf{0}, \mathbf{\Theta_u}) \tag{8}$$

- Model for Between-group for the latent class variable and class membership.

$$\text{logit}[P(d_{lj} = 1)] = \upsilon_{0k} + \sum_{n=1}^{N} \upsilon_{nk} W_{nj} \tag{9}$$

where

t: time point

i: individual

j: group/cluster

a_{1tij}: individual level, time related variable

X_{ij}: within-group level covariate

W_j: between-group/cluster level covariate

Equations 1 through 5 show the within-group level models and Equations 6 through 9 show the between-group level models. In equation 1, Y_{tij} is the observed individual outcome at time/occasion t for individual i within a group/cluster j (e.g., school), π_{0ij} is the expected value of Y for this individual when t=0, π_{1ij} is the expected slope/growth on the outcome for this individual, a_{1tij} measures the time/occasions for this individual and e_{tij} is the residual/error associated with this model for this individual. It is possible to include more time/occasion variables to model other growth effects (e.g., quadratic effect) in addition to the linear growth effect shown here. Equations 2 through 4 show the within-group level model or the repeated measure for intercepts and slopes and Equation 5 shows the model for subjects' latent class memberships, given their covariates. Within-class intercepts and slopes are expressed with three factors, *m*

covariates X_{mij}, k latent classes c_{kij}, and random effects u_{0ij}. c_{kij} is equal to one when an individual i in cluster j belongs to the latent class k and otherwise zero where $k = 1, 2, 3,….,K$ and K is the total number of within-group latent classes β_{0jk} and β_{1jk} are the mean intercept and slope value for within-group class k. Equation 5 represents a multinomial logistic regression to describe the likelihood of membership in each of the latent class variable's levels, associated with predictors where $k=1$ is the reference class level.

Between-group level equations 6 through 9 are almost identical to within-level equations from 2 to 5. Within-group heterogeneity in intercepts and slopes are regressed on three factors: between-group covariates, W_{nj}, between-group latent class variable, d_{lj}, and random effects (u_{0j} and u_{1j}) where d is the between-group latent class variable with l levels, and L is the total number of between-group latent classes ($l = 1, 2, 3,….,L$). d_{lj} is one when a cluster j belongs to the latent class l and otherwise zero. γ_{0l} and γ_{1l} are the mean intercept and slope value for between-group latent class variable level l. Equation 9 represents a multinomial logistic regression to describe the likelihood of class membership associated with predictors where $k=1$ is the reference class level. The errors/residuals in each of the within-level measurement model, within-level structural/repeated measure models, and between-group models, are all assumed to be normal, independent across levels (e.g., between level-1 and level-2), and uncorrelated with the covariates. There are three levels of equations for MLGMM (i.e., two levels for within-group and one level for between-group). The term *cluster* is defined as the grouping unit at level-2.

Figure 6 is a graphical representation of unconditional MLGMM based on the

Muthén and Muthén (1993-2010) representation for the M*Plus* software.
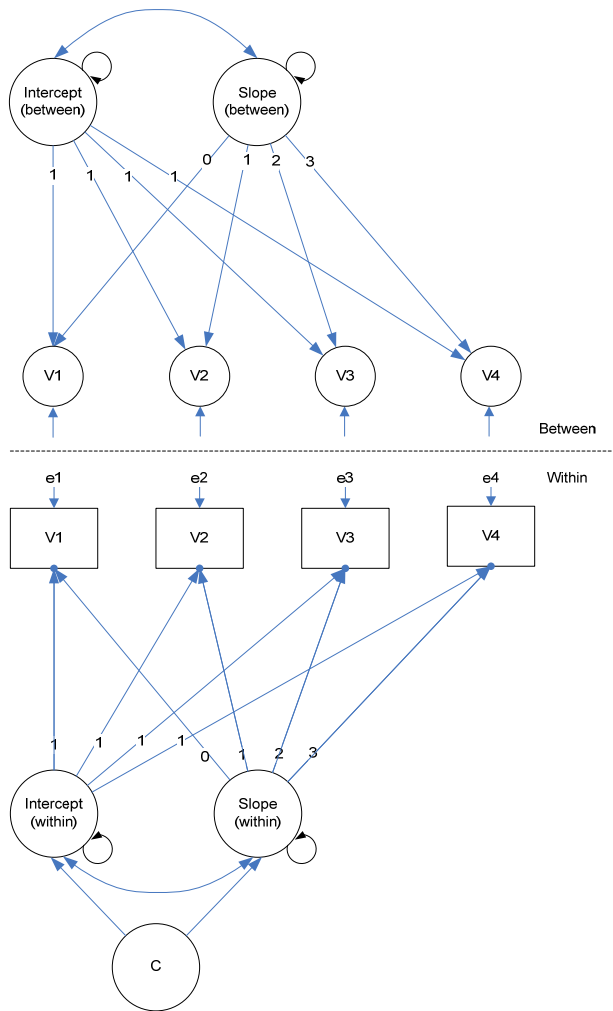
Figure 6. Graphical representation of unconditional MLGMM

2.2.2 Algebraic representation of Growth Mixture Model

The growth mixture model (GMM) is a special case of MLGMM where no

between-group models are included. Formulation of GMM models is achieved by

dropping the group/cluster notation *j* from MLGMM Equations 1 through 4 above,

resulting in the following specifications:

- Individual level measurement model

$$Y_{ti} = \pi_{0i} + \pi_{1i} a_{1ti} + e_{ti}, e_{ti} \sim N(0, \sigma^2) \tag{10}$$

     o Individual level structural model for the:

     o Intercepts

$$\pi_{0i} = \sum_{k=1}^{K} \beta_{00k} c_{ki} + \sum_{m=1}^{M} \beta_{01k} X_{mi} + r_{0i} \tag{11}$$

     o Slopes

$$\pi_{1i} = \sum_{k=1}^{K} \beta_{10k} c_{ki} + \sum_{m=1}^{M} \beta_{11k} X_{mi} + r_{1i} \tag{12}$$

$$, r_i \sim N(\mathbf{0}, \mathbf{\Theta_r}) \tag{13}$$

   • Model for the latent class variables

$$\text{logit}[P(c_{ki} = 1)] = \delta_{0k} + \sum_{m=1}^{M} \delta_{mk} X_{mi} \tag{14}$$
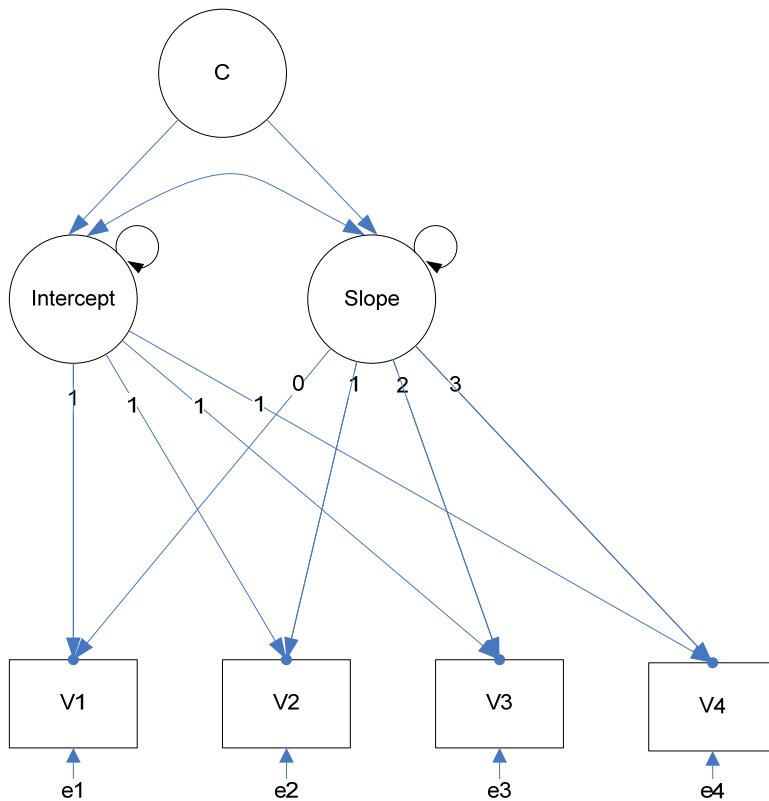
Figure 7. Graphical representation of linear GMM

It is clear from Equations 11 through 14 that the cluster, j, does not appear in any model, reflecting the assumption that individual growth factors are sufficient to estimate the effects of interest in the data. Figure 7 is a graphic representation of GMM, which is the same as the within-subject part of Figure 6 showing MLGMM.

2.2.3 Algebraic representation of multilevel latent growth model

A multilevel latent growth model (MLLGM) is a non-mixture case (i.e., without a latent class variable) of MLGMM. Therefore the formulation of MLLGM takes the MLGMM formulations and excludes both within-level latent class variables, c_{kij}, and the between-level latent class variable, c_{kij}, from Equations 1 through 4 and 6 through 8, so they become:

- Within-group level measurement model

$$Y_{tij} = \pi_{0ij} + \pi_{1ij} a_{1tij} + e_{tij}, e_{tij} \sim N(0, \sigma^2) \text{ (15)}$$

- Within-group level structural model for the intercepts and slopes
  - Intercepts

$$\pi_{0ij} = \beta_{00j} + \sum_{m=1}^{M} \beta_{0mj} X_{mi} + r_{0ij} \tag{16}$$

  - Slopes

$$\pi_{1ij} = \beta_{10j} + \sum_{m=1}^{M} \beta_{1mj} X_{mi} + r_{1ij} \tag{17}$$

$$r_{ij} \sim N(\mathbf{0}, \mathbf{\Theta_r}) \tag{18}$$

- Between-group level model
  - Intercepts

$$\beta_{0ij} = \gamma_{000} + \sum_{n=1}^{N} \gamma_{00n} W_{nj} + u_{0j} \tag{19}$$

  - Slopes

$$\beta_{10j} = \gamma_{100} + \sum_{n=1}^{N} \gamma_{10n} W_{nj} + u_{1j} \tag{20}$$

$$u_{ij} \sim N(\mathbf{0}, \mathbf{\Theta_u}) \tag{21}$$

The formulation of MLLGM is identical to a MLGMM where the latent class variable has just one level (and everyone falls into this single level). The graphical representation of MLLGM (not shown) is also very similar to MLGMM, obtained by simply excluding the latent class variables C and D from Figure 6 and any connections from/to these latent class variables (in Equations 16 through 20).

2.2.4 Algebraic representation of latent growth model

The simplest form of MLGMM is the latent growth model (LGM) where there are neither latent class variables nor group/cluster information included in the model. LGM is expressed with the following four equations:

- Within-group level measurement model

$$Y_{ti} = \pi_{0i} + \pi_{1i} a_{1ti} + e_{ti}, e_{ti} \sim N(0, \sigma^2) \tag{22}$$

- Within-group level structural model for the intercepts and slopes
  - Intercepts

$$\pi_{0i} = \beta_{00} + \sum_{m=1}^{M} \beta_{0m} X_{mi} + r_{0i} \tag{23}$$

  - Slopes

$$\pi_{1i} = \beta_{10} + \sum_{m=1}^{M} \beta_{1m} X_{mi} + r_{1i} \tag{24}$$

$$r_{ij} \sim N(\mathbf{0}, \mathbf{\Theta_r}) \tag{25}$$

In LGM, two growth factors, representing intercept and slope, completely capture individual growth trajectories.

2.2.5   Algebraic representation of the multilevel model

The formulation of a longitudinal multilevel model (MLM) is similar to LGM; Equations 26 through 28 below are almost identical to Equations 22-24 for LGM. The formulation for a two-level unconditional (i.e., without covariates or explanatory variables) MLM is:

- Level-1

$$Y_{it} = \pi_{0i} + \pi_{1i}T_{ti} + e_{ti}, \; e_{ti} \sim N(0, \sigma^2) \tag{26}$$

- Level-2

$$\pi_{0j} = \beta_{00} + \beta_{01}X_i + r_{0i} \tag{27}$$

$$\pi_{1j} = \beta_{10} + \beta_{11}X_i + r_{1i} \tag{28}$$

where Y is a response variable, T is a time variable, t is a time or measurement occasion, i is an individual, and X is a time-invariant covariate.

2.2.6   Contrasting multilevel (MLM) and latent growth (LGM) models

The selection of one model from the family of latent growth models falling within the MLGMM classification might be dictated by the quality of data (e.g., in case of missing individual data; insufficient sample size for the model complexity; etc. e.g., Muthén, 2004, 2006) and/or by the research questions under study. In LGM, time or measurement occasions are fixed, with values that must be pre-specified, while with MLM, time is a variable reflecting any values representing a time, visit, or occasion. Therefore, LGM and MLM will have identical specifications and equivalent estimates when time or measurement occasions are fixed (e.g., t= 0, 1, 2, 3 in Equation 15 and 17).

2.2.7   Value-Added Model as a multilevel model

The value-added model (VAM) is a special case of multilevel model wherein the data have the specific hierarchical structure with students nested within a classroom and classrooms nested within a school.
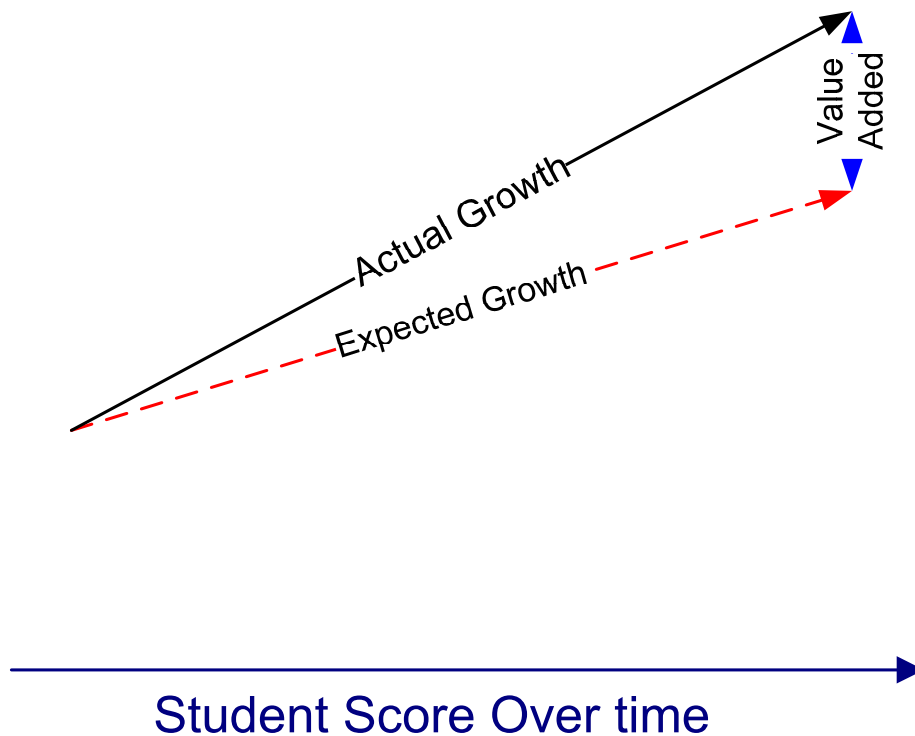
Figure 8. Conceptual representation of value added model

Figure 8 shows a conceptual representation of VAM. The difference in a student's achievement between that predicted by the model (red dotted line) and the actual achievement (black solid line), is that "value-added" by the external factor (e.g., school and teacher, shown in blue line) to be estimated. The term "value-added" represents the emphasis on estimating the contribution of higher-level effects such as teachers (level 2) and schools (level 3) on the student's achievement and/or improvement (Sanders & Rivers, 1996).

The simplest case of VAM is an unconditional two-level MLM (Doran & Lockwood, 2006; Singer, 1999), very similar to Equations 27 through 29 if terms for predictors **X** are removed. Value-added modeling is currently being used in states such as Tennessee (Tennessee Value Added Assessment System; TVASS; Sanders & Rivers, 1996) and North Carolina (Education Value Added Assessment System; EVAAS, SAS Inc. 2010). These models are far more complex than the unconditional VAM, and are intended, and are being used, for high stakes evaluation, potentially influencing employment status of teachers (see Springer et al., 2010 for current use of VAM for teacher evaluation).

The primary focus of evaluation of performance using VAM is to identify and quantify the contributions of higher-level variables such as teachers, schools, and district to the observed growth in level-1 (e.g., change in student scores over time); this is similar to the general objectives of the conventional MLM with a focus on the 2nd – and higher-level estimates. The main difference between VAM and the 2-level unconditional MLM lies in both the addition of another level of hierarchy (e.g., schools within district) and estimation of the effects of covariates associated with the additional level (e.g., level-3) on the level-1 outcomes.

2.2.8   Effects of level-1 heterogeneity on the estimates of level-2 effects in MLGMM

Muthén and Asparouhov applied the multilevel mixture model to simulated data and real data to demonstrate the use and utility of mixture modeling in educational contexts. Muthén and Asparouhov (2009) progressively added complexity to models. First, they showed that a simple regression model could not fully explain group differences in math achievement between males and females due to the underlying

heterogeneity in scores arising from a latent class variable with two levels (i.e., low and high achievers).

The second example in Muthén and Asparouhov (2009) was a conventional MLM with a single (manifest) level-1 predictor, student-level socio-economic status. Muthén and Asparouhov (2009) showed the different conclusions derived from regression with and without consideration of the latent class, effectively comparing results from a conventional multilevel model against those of a multilevel *mixture* model. Muthén and Asparouhov (2009) showed that the effect of student-level covariates can affect the interpretation of the results from a conventional multilevel regression, since the student-level latent class variable interacted with the school-level predictor. They also showed that in the presence of a substantive latent class variable at the student level, especially when the class levels interacts with the covariate, interpretation of results will depend on the latent class membership at level 1 and the value of the school-level covariate (i.e., at level 2).

Finally, Muthén and Asparouhov (2009) used actual data to fit three multilevel mixture models that varied in complexity. There were both within-level and between-level predictors (i.e., covariates) in all three models. A "plausible" null model, the simplest one fit to the data, was an unconditional MLM. Three mixture models were fit to the data, each with latent class variables. In one of these, both within-group and between-group latent class variables, the other two with only within-group latent class variable where the more complex of two allow both the intercept and slope from covariates to the predictor to be different whereas the simplest model only allows the intercept to be different.

The fit of these models to the data was estimated using a variety of indices designed to capture how well the models represented the relationships and variability in the data. The main fit statistic used to compare these models was Bayesian Information Criteria (BIC; Schwartz, 1978) representing the information in the data that was lost with each model's respective formulation (see Anderson, 2008), but also by comparing the fit to the data by each model against that of a model without a latent class variable (a plausible "null" alternative model).

Muthén and Asparouhov (2009) found three specific impacts to estimates and inferences, as compared to the conventional MLM, were derived from the three mixture models:

1. The degree of precision in level-2 estimates from the MLM was limited by the failure to capture the level-1 latent class variable.

2. Estimates of level-2 effects were inflated in the conventional MLM compared to those from the three mixture models.

3. The effects of predictors were significantly different between the conventional and mixture models; these effects were attenuated in all mixture models as compared to the conventional MLM estimates. This supports the importance of modeling the level-1 heterogeneity with latent classes in order to avoid reaching the wrong conclusion by inflating the effect of covariates.

Based on their exploration of the simple regression, conventional MLM, and MLGMM models and their respective fits to the data, in addition to the differing results and inferences supported under each analysis, the authors stated that "level-1 heterogeneity in the form of latent classes is mistaken for level 2 heterogeneity in the

form of the random effects that are used in conventional two-level regression analysis" (p. 655).

Muthén and Asparouhov (2009) had used BIC to identify/select the version of each mixture model that contained the number of levels within the latent class variable that was most consistent with the data (i.e., to select the model with the number of latent class levels associated with the lowest BIC value for that particular model specification). For each of the three mixture models, Muthén and Asparouhov (2009) also explored varying numbers of levels for the latent class representing the "mixture". They used BIC to identify the version of each mixture model with the number of class levels that best (lowest BIC of the set) captured the heterogeneity in the data. They reported that all three mixture models fit the data significantly better than the model without a latent class variable, but the differences in fit among the three alternative mixture models were less pronounced. In fact, all three mixture models yielded substantively interpretable results, with a single number of latent class levels identified by BIC for each. Thus, in this case, fit and interpretability of classes (i.e.., yielding classes that could be assigned substantively interpretable labels) did not identify a single "best" model. Therefore, in addition to demonstrating the utility and incremental improvements in interpretability that MLGMM brings to, Muthén and Asparouhov (2009) also underscored new challenges that can arise from the application of this technique, namely, that fit and interpretability, which usually drives model selection, may not clearly differentiate reasonable alternative models derived under MLGMM. This is an additional consideration for adoption of models that could be used in teacher evaluation and for decisions or policies made on the basis of such evaluations.

Muthén and Asparouhov (2009) identified the importance of accounting for heterogeneity attributable to a latent class variable in the context of the MLM framework; this partly informed the design of this simulation study. Their use of information criteria for model selection was also integrated into this study, as described in Chapter 3. Specifically, this study included BIC, as they did, but also an assortment of other information criteria, in order to further understand BIC's specific functionality under MLGMM.

2.2.9 Latent class variable identification in MLGMM

Palardy and Vermunt (2010) and Chen, Kwok, Luo, and Willson (2010) investigated the issues of latent class variable identification and the precision of classification of individuals into the latent class variable's levels in MLGMM. The issue of latent class variable identification in MLGMM, have been studied by several investigators (Muthén & Asparouhov, 2009; Palardy & Vermunt, 2010). Palardy and Vermunt (2010) reported that manifest covariates have the potential to change the distribution of random effects by which latent class variables are identified, thereby affecting identification of substantively interpretable latent class variables. Palardy and Vermunt (2010) recommended that manifest covariates should be identified *a priori,* and substantively, and that they not be derived via exploratory analysis of the data under study in mixture modeling.

The other major issue identified by Palardy and Vermunt (2010) is that the random effects for the intercepts and slopes of growth trajectories that are estimated in a growth modeling framework will interact with the identification of latent class variables because they all compete to explain the same variability in the student-level data. Consistent with

other studies (e.g., Bauer & Curran, 2004; Lubke & Neale, 2006), Palardy and Vermunt (2010) found that fixing random effects (i.e., fixing the error term on a group level equation for intercepts and/or slopes to zero) is likely to cause over-extraction of latent class variable levels. So, although the results from Muthén and Asparouhov (2009) underscored the importance of mixtures for this model type, namely the latent class variable at level 1, and its influence on estimates and inferences on level 2 variable effects, Palardy and Vermunt (2010) identified pronounced challenges to the use of mixtures in the MLGMM context. This underscores the earlier point that increasingly complex models can serve important purposes for improving precision and decreasing bias in estimation and decisions based on these estimates, but the more complex models often lead to other problems or issues.

A different but related challenge is the effect of using mixtures, but not multi-level approaches, in growth modeling. Chen et al. (2010) investigated the effect of ignoring the nested structure on identifying the latent classes at level-1 in MLGMM. That is, they focused on the effects of erroneously treating hierarchical data as if there was no hierarchy – running a GMM instead of MLGMM. Chen et al. (2010) found the nested structure had relatively minor effects on the latent class variable identification in that a given mis-specified GMM did correctly identify the latent class variable and class level membership for individuals in 80% to 90% of simulation conditions. When compared to results in simulation conditions with the correctly-specified model, MLGMM, the GMM results were not off by much as the MLGMM class levels were correctly recovered in 87% to 92% of the MLGMM conditions.

However, Chen et al. (2010) found that the intraclass correlation of group, magnitude of within-class variance, and latent class mixture proportions each had a substantial effect on the latent class level identification when the model was mis-specified (i.e., when data were analyzed with GMM), but not with the MLGMM. Other effects of ignoring the nested structure inherent in data were reported to be: 1) less precise fixed-effect estimates with greater standard error; 2) overestimated variance estimates for effects of lower-level variables; and 3) less accurate standard error estimates for all parameter estimates. In general, incorrectly ignoring the nesting structure was determined to have less of an impact on the fixed-effect estimates than on random-effect estimates, but this particular type of misspecification (i.e., ignoring the nested structure in data) led to bias and imprecision that would have important implications for VAM applications. As the random effects are the estimates of interest in VAM applications, a failure to capture the nesting of the data could adversely impact policy and other high stakes decisions (as was alluded to by Ballous, 2002). In general, incorrectly ignoring the nesting structure was determined to have limited impact upon the identification of latent class but to have an impact on the bias and precision of parameter estimates.

2.3 Substantively interpretable latent class structure in VAM

Two of the papers described above, Muthén and Asparouhov (2009) and Palardy and Vermunt (2010), have several important implications for VAM in terms of the correct – MLGMM – analytic approach. A latent class variable, representing student performance and development, has been identified by two independent studies (Chudowsky et al., 2007; Lazarus et al., 2010). Both studies identified a subgroup of students that persistently performs at the lowest level. Students are known to be

heterogeneous in their performance and their development (Chudowsky, Chudowsky, & Kober, 2007; Lockwood & McCaffrey, 2007; Lazarus et al., 2010), but they may also fall into more predictable (latent) classes that can complicate estimation with growth curve modeling – particularly if this predictable source of variability is ignored (e.g., as reported by Muthén and Asparouhov, 2009). Students who chronically perform at a low level over time have been characterized as "permanently low performing" students (PLP; Chudowsky et al., 2007; Lazarus et al., 2010). These students start off, and remain, at a low performance level over time, and are often distinct from students who start off at a higher level and remain at that level over time and from those who start higher or lower and exhibit change over time.

In their study of student types, Lazarus et al. (2010) identified two groups of low performing students, low performing (LP) and persistently low performing (PLP) (see also Chudowsky et al., 2007). LP students were defined to be those who scored at the 10th percentile or lower on the state wide standardized test in <u>one</u> of the past three years. Persistently low performing (PLP) students were those who scored at the 10th percentile or below on the statewide standardized test for <u>all three</u> years. Those students identified as PLP were not performing so badly overall that they were eligible to take the alternate form of assessment (i.e., a test for students who are in the special education program), but their performance suggests that the regular achievement tests are simply too difficult for them. Lazarus et al. discovered two demographic (manifest) variables that tended to characterize the PLP student type: they were more likely to be minorities, and more likely to be receiving free or reduced lunch (a proxy variable for low socio-economic status). Although these trends were observed for the manifest demographic variables, neither was

statistically significantly predictive of belonging to the PLP student type. As Palardy and Vermunt (2010) suggested, predictor variables or covariates should not be included for exploration of latent class variables in MLGMM due to the potential interaction between them obscuring the identification of latent classes. Together with the PLP results of Lazarus et al. (2010), indicating that manifest covariates are not sufficient, or sufficiently explanatory, the results and recommendations by Palardy and Vermunt (2010) suggest that a latent class variable – based on slopes and intercepts – may be a more efficient and effective method of identifying students in this class.

2.3.1   Potential impact of latent classes

Palardy and Vermunt (2010) and Chen et al. (2010) are recent studies showing the impacts of inappropriate modeling of latent class variables (Palardy & Vermunt, 2010) or of the nested data structure (Chen et al., 2010) on the estimates of individual and group effects (i.e., slopes and intercepts) as well as their predictors. As stated before, growth curve (and related) modeling methods have incredible potential for educational research as well as for decision making and policies that are based on evaluations, but for these methods to be both useful and used appropriately, the impacts of latent class variables and hierarchical data within the growth curve modeling framework need to be fully investigated, particularly at the level of individual estimates (i.e., a parameter for each case) rather than at the effect level (e.g., overall group effect).

2.4 Effects of un-accounted for heterogeneity at level 1 on the precision of level-2 estimates in multilevel data

This study builds on the results of these three key studies (Chen et al., 2010; Muthén & Asparouhov, 2009; Palardy & Vermunt, 2010), and incorporated the PLP

student type (Chudowsky et al., 2007; Lazarus et al., 2010) so as to estimate, and understand the magnitude of, bias in estimates at each stratum of the analysis. As described above, there are two issues in the identification of latent class variables, namely the assignment of individuals to levels of these variables and the appropriate estimation of effects of interest in MLGMM:

1) Covariates affect the identification of latent class variables;

2) Nested structure has limited impact upon the identification of latent class variables but can influence estimation and interpretation of random effects.

Coupled with the potential importance of the MLGMM for education research and decision-making, and particularly the salience of the latent class variable described by Muthén and Asparouhov (2009) and the substantively important class of PLP students indentified by Lazarus et al. (2010) and Chudowsky et al. (2007) in their analyses, this recent body of work motivated this effort to quantify these effects in the simulation study described in Chapter 3. However, identification of class membership at level 1 has not been shown to be influenced by these factors, nor is it often a consideration for decision making or VAM interpretability; therefore, assignment of individuals (at level 1) to levels of these variables was not pursued in this study.

Chapter 3: Methods

Chapter 2 provided the background supporting the objectives of this study, which were to: (1) investigate the effect of unaccounted-for heterogeneity in growth at level 1 on level-2 effects by comparing the level-2 effect estimates derived from a conventional MLM and from a multilevel growth mixture model (MLGMM); (2) examine the stability of level-2 effect estimates in MLGMM models; and (3) estimate the likelihood of class misidentification at level 1 in MLGMM and its consequences for level-2 estimates and their interpretation. To meet these research objectives, two research questions were investigated:

1) Are the level-2 parameter estimates in the multilevel model affected (in terms of bias, and precision) by incorrectly-modeled level 1 effects?

2) What information criteria can be used to identify the true number of latent class variable levels in the MLGMM context?

This chapter describes the simulation study used to answer these questions.

3.1 Characteristics of the simulation

Table 1 shows the details of the proposed simulation. Each combination of four simulation conditions in Table 1 represents a longitudinal model from which 100 datasets were sampled. Each condition has three time points (t=0, 1, 2). The simulation conditions combined to represent a total of 120 models, and with 100 "samples" or replications from each we arrived at 12,000 datasets that were built using SAS. The pilot study (see Chapter 4) results led to the increase in number of replications to 100. The 120 models were built according to the combination of characteristics representing that cell in Table 1.

Table 1

Simulation Manipulated Conditions

Conditions	Number of Levels
Custer Size	2
Cluster Number	3
Mixture Proportion	4
Cluster Effect	5
Total	120

3.2 Characteristics of individual (level 1) data

Two different growth trajectories in individuals (i.e., at level 1) represented the level-1 heterogeneity in this simulation. Growth profiles of these two groups followed Chen et al. (2010) – which in turn is based on Nylund et al. (2007) – namely, one with steeper, one with shallower slope (see Figure 1). Table 2 shows the parameter settings that were used to represent growth profiles of the two class levels ("fast growing", "slow growing") that were included within every model. These profiles were held constant across the simulation conditions as shown in Table 2, by fixing the parameters listed in Table 2 (leftmost column) to the respective values shown under the right columns (Growth profile, Fast and Slow) in Table 2.

Table 2

Setting of Growth Parameters in order to obtain the two latent classes for every sample

	Growth profile	
Parameters	Fast	Slow
Intercept mean	2.5	1
Slope Mean	0.6	0.1

The fast growing individuals have intercepts (i.e., initial level) varying around 2.5 and slopes (i.e., growth rate) varying around 0.5, while the slow growing individuals have intercepts varying around 1.00 and growth rates varying around zero. These settings were selected because they create a clear separation of the two groups, and the parameter settings characterizing the slow growing group correspond to the PLP students (Lazarus et al., 2010) described in Chapter 2. A simulation study identifying two growth profiles utilizing GMM, with N=2400 (i.e., 1200 for each growth profile) and 100 replications found the average correct identification of two growth profiles at 91% with the minimum of 89% and the maximum of 93%.

3.3 Characteristics of cluster (level-2) data

Four attributes define the characteristics of level-2 data, representing clusters in the hierarchy: 1) the cluster number; 2) the cluster size; 3) Cluster Types and the mixture proportion of individuals within a Cluster Type; and 4) the cluster effect as shown in Figures 2 through 4. Each attribute, and its role in the simulation, is described below.

3.3.1   Sample size is determined by cluster size and cluster number

The sample sizes that were used in this simulation are determined by the cluster number and cluster size (i.e., the level-2 characteristics), illustrated on figure 2. For this simulation, three levels of cluster number (CN) were chosen to represent small, medium and large (CN=30, 60, 90, respectively) districts from which clusters might be drawn in

VAM contexts. Chen et al. (2010) used CN = 30, 50, and 80, but because three Cluster

Types within a cluster (see section 3.3.2, below) need to have equal numbers of

individuals within cluster (i.e., Cluster Type size of 10, 20, or 30), the simulation

involved equal numbers per Cluster Type. An equal number of Cluster Types controls for

the potential effects of different Cluster Type numbers within a cluster on the estimation

of the value-added effect (i.e., the cluster effect).

The cluster size (CS) is the number of observations, or individuals, within a

cluster. For this simulation, CS was also based on the design used by Chen et al. (2010),

namely, values of 20 and 40. Sanders and Rivers (1996) used a cluster size of 20 and

Wright, Horn and Saunders (1997) used a cluster size of 25 on their respective simulation

studies of VAM – in both cases, they argued these cluster sizes represent average

classroom size in the U.S (at that time). This study included a cluster size of 40 because

some schools and districts tend to have larger class sizes.

3.3.2   Cluster type

In a VAM context, if the level-2 data are conceptualized as representing the

between-group level model (e.g., Equations 19-21 in Ch 2), then the cluster type can be

thought of as schools having a different proportions of student types (i.e., fast and slow

growth). In their study, Chen et al. (2010) only included clusters with equal proportions

of fast and slow growth students (i.e., mixture proportion conditions 4 and 5 of this

study). This simulation included three Cluster Types (see table 3). As noted in chapter 1

and the previous section, when each Cluster Type accounts for one-third of any given

cluster, it permits the evaluation of potential biases for the cluster effect estimates across

Cluster Types having different proportions of students in our two growth classes. Further,

in the evaluation of an effect of school in a VAM context, it is unrealistic to expect all schools to have the same proportion of students in these two growth classes. Thus, unlike the Chen et al. (2010) study, this simulation included three equal sized Cluster Types per cluster but within each Cluster Type, different mixture proportions (described below) were included.

3.3.3   Mixture proportion

The mixture proportion characterizes the prevalence of membership in the latent class' different levels. As laid out above, the growth profiles (fast/slow growing) represent the two levels of the latent class variable. The mixture proportion dictates what proportion of the sample belongs to each of these levels (fast, slow). This study uses five patterns of mixture proportion. Three patterns involve different mixture proportions based on the cluster type, i.e., the cluster type are each 1/3 of the cluster, but within each of these Cluster Types, the mixture proportions of fast and slow growers vary (see Figure 3). Two patterns of mixture proportions, taken from the Chen et al. (2010) study, represent fixed parameters within each Cluster Type. Table 3 shows the mixture proportion for these four simulation conditions. In this study, the data were generated by creating populations representing the two classes (growth profiles) and then sampling – as dictated by the condition's mixture proportion – the appropriate number of observations from each of these classes.

Table 3

Definition of mixture proportion by Cluster Type

Mixture proportion Pattern (MP)	Level-2 features Cluster Type	Growth (latent class, level-1 feature) Fast	Slow
1	1	50	50
	2	75	25
	3	100	0
2	1	25	75
	2	50	50
	3	75	25
3	1,2,3	50	50
4	1,2,3	75	25

Mixture proportion conditions 1 and 2 investigate the influence of differential mixture proportion among cluster types, where condition 1 has less variability in proportion and condition 2 has more variability. Mixture proportion conditions 3 and 4 have consistent mixture proportions among cluster types. Simulation studies conducted by Muthén and Asparouhov (2009) and Chen et al. (2010) used settings similar to conditions 3 and 4. These two conditions assess the effect of mixture proportion across the other simulation conditions.

3.3.4   Cluster effects

The fourth feature of the level-2 data is the cluster, or cluster-level, effect. In the context of VAM analysis, the cluster-level effect represents the value-added effect, γ_{11}, shown in Equation 36. The cluster effects are the parameters of interest in this study. As can be inferred from Figure 8, and from Equations 19-21, the cluster effect is only defined/estimable for individuals in the fast growth group in this simulation – because the slow growth group has zero slope (see Table 2). That is, since VAM seeks to estimate the

impact of higher-level variables on the development or change in the first level variable

(i.e., at the individual level), if there is no change, there can be no value-added effect

estimated.

The cluster effect varied based on the makeup of the cluster type –even when

cluster types are equal sizes (i.e., 33% of the given cluster per Cluster Type), as in this

simulation. Table 4 shows the five cluster effects used in the simulation, in the third

column (Cluster Effect) of the table. The same number of individuals was assigned to

each cluster effect condition proportionally depending on the size of cluster (CS) and the

number of cluster effects (e.g., five cluster effects for the first cluster effect condition).

As Table 4 shows, this study included five patterns of level-2 effects, three as

fixed effects and two as random effects. This permitted the systematic investigation of the

parameter recovery of cluster-level effects in the various simulation conditions. These

effects – representing the value added effects – are parameters of interest, critical to

address the research questions.

Table 4

Cluster effects defined by pattern of Cluster Type

Cluster Effect Pattern (CE)	Cluster Type	Cluster effect parameters
1	1	(-1, -0.5, 0, 0.5, 1)
	2	(-1, -0.5, 0, 0.5, 1)
	3	(-1, -0.5, 0, 0.5, 1)
2	1	(-1, -0.5)
	2	0
	3	(0.5, 1)
3	1	(0.5, 1)
	2	0
	3	(-1, -0.5)
4	1, 2, 3	$\gamma_{11} \sim N(0, 0.5)$
5	1, 2, 3	$\gamma_{11} \sim N(0, 1.0)$

Cluster effect condition (CE) 1 has the same five parameter values across three cluster types. This condition is specifically design to evaluate the influence of differential mixture proportion (i.e., MP1 and MP2 conditions) among cluster types in terms of the direction of biases (i.e., positive or negative) and the precision of estimates. The cluster effect condition 2 (CE2) and 3 (CE3) also investigate the cluster type level bias and precision of parameter estimates between cluster type 1 and 3 (i.e., fixed parameters are reversed between cluster type 1 and 3). These conditions were included to systematically investigate the extent of positive and negative bias in the parameter estimates. The random effects based on different variances were generated for the cluster effect condition 4 and 5 (small variation for CE4 and large variation for CE5).

3.4 Data simulation

All data were generated in SAS (9.2, SAS Inc., Cary, NC). Data generation was based on 2-class MLGMM, varying parameters for each simulation condition as outlined in Tables 1 through 4. Based on the background given in Chapter 2, the following

Equations (29-37) show how the data for this simulation were generated for the models

with characteristics outlined in Table 2 over three time points ($t=0, 1, 2$)

Level 1

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}a_{1tij} + e_{tij}, e_{tij} \sim N(0,1) \tag{29}$$

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}Class_{ij} + r_{0ij} \tag{30}$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}Class_{ij} + r_{1ij} \tag{31}$$

$$where \begin{bmatrix} r_{0ij} \\ r_{1ij} \end{bmatrix} \sim MVN \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{\pi_{00}} & \tau_{\pi_{01}} \\ \tau_{\pi_{10}} & \tau_{\pi_{11}} \end{bmatrix} \right) \tag{32}$$

Level 2

$$\beta_{00j} = \gamma_{00} + \mu_{0j} \tag{33}$$

$$\beta_{01j} = \gamma_{01} \tag{34}$$

$$\beta_{10j} = \gamma_{10} \tag{35}$$

$$\beta_{11j} = \gamma_{11} \tag{36}$$

$$where \ \mu_{0j} \sim N(0, \tau_{\beta_{000}}) \tag{37}$$

Based on the growth profiles shown in Table 3, in all simulation conditions the

Level-1 variance, e_{tij}, is set to one (i.e., it is fixed). Equation 33 specifies the magnitude

of within-class variation (i.e., variance-covariance of slopes and intercepts), which

follows the Chen et al. (2010) specification of a medium magnitude or "low separation"

condition (after Tofighi & Enders, 2008) as: $\tau_{\pi_{00}} = 0.20$, $\tau_{\pi_{10}} = \tau_{\pi_{01}} = 0.05$, and

$\tau_{\pi_{11}} = 0.05$. The four group-level (Level 2) growth parameters (Equations 33-36) are set

to the following values: $\gamma_{00} = 1.0$, $\gamma_{01} = 1.5$, $\gamma_{10} = 0.1$, and $\gamma_{11} = 0.5$, where Class on

Equations 30 and 31 is a dichotomous indicator (1=High Growth, 0=Low Growth), so that growth parameters β_{11j} are correctly represented. An intraclass correlation (ICC), representing the magnitude of intercept random effect (error/variability) over the total variability, of 0.10 translates to a parameter setting of $\tau_{\beta_{000}} = 0.133$ (Chen et al., 2010). For the fixed Cluster Type effects are manipulated based on the specification on Table 4. Equal numbers of fixed parameters (e.g., -1, -0.5, 0, 0.5, or 1) are assigned to individuals within each Cluster Type.

To contextualize these features within Table 1, in the simulation condition with cluster number (CN) =30, cluster size (CS) = 20, mixture proportion condition 1, and cluster effect condition 1, there are 10 clusters for each Cluster Type, with the proportion of individuals in the fast growth type across Cluster Types being set to 25%, 50%, and 100%, respectively. The cluster effects are set to (-1, -0.5, 0, 0.5, and 1) for all cluster subtypes, having two clusters for each cluster effect condition (i.e., the number of Cluster Types (10) divided by the number of cluster effect conditions (5); 10/5 = 2).

3.5 Model Fitting

The preceding section describes how the data were generated for the model features that were studied in this project. Given those 120 models, for each of the 100 data sets, four models were fit in the way outlined by Chen et al. (2010): 1) mis-specified model (i.e., MLLGM latent class unmodeled; 2-3) two incorrect mixture models (i.e., MLGMM with 1- and 3-latent classes); and4) correct mixture model (i.e., MLGMM with 2-latent classes) to evaluate the effect of unmodeled latent class (i.e., heterogeneity at students' level) and the latent class identification issues. M*Plus* 6.1 (Muthén & Muthén, 2010) was used to fit these four models to each of the 12,000 samples that are generated

by the 120 models that were generated as shown in Table 1. The growth profile on table 2 was used as the starting value for the corresponding parameters on the mixture models.

3.6 Analysis of results of model fitting

The identification of the presence of, and levels in, a latent class variable in mixture models should be based on more than one statistical index (Bauer & Curran, 2004; Nylund et al., 2007; Palardy & Vermunt, 2010). This study utilized six indices – information criteria – to identify the model with the number of latent class variables, and its levels, that are most representative of the data (according to information in the data represented by the model) (see Anderson, 2008). It is important for the model selection criteria to be robust since, as outlined in the foregoing study design elements, there were mis-specified models fit to data. The six information criteria that were used are:

- Akaike Information criteria (AIC; Akaike 1987)

- Modified Akaike information criteria (AIC3; Bozdogan, 1993)

- Second order bias corrected AIC (AICc; McQuarrie & Tsai, 1998; after Akaike, 1987)

- Bayesian information criteria (BIC; Schwarz, 1978)

- Bayesian information criteria with a cluster number for sample size adjustment factor (BICB; Parlady & Vermunt, 2010)

- Sample size adjusted BIC (SABIC; Sclove, 1987)

All information criteria are defined as a function of log-likelihood of the model; they differ in terms of the penalty each imposes depending on the number of parameters estimated and/or sample size. Lower values of any information criterion indicate that the model for which it was computed fits the data better than do models with higher criterion

values. Equation 38 is AIC (Akaike, 1987), on which many of the modern information

criteria are based (see Anderson, 2008); the following equations define each information

criterion that were used in this study:

$$AIC = -2\log LL + 2P \tag{38}$$

$$AIC = -2\log LL + 3P \tag{39}$$

$$AICc = -2LL + 2p[\frac{N}{N - p - 1}] \tag{40}$$

$$BIC = -2LL + P\log(N) \tag{41}$$

$$BICB = -2LL + P\log(N_{cluster}) \tag{42}$$

$$SABIC = -2LL + P\log(\frac{N + 2}{24}) \tag{43}$$

where P is the number of estimated parameters, N is the sample size, and $N_{cluster}$ is the

number of clusters.

Muthén and Asparouhov (2009) and Nylund et al. (2007) reported BIC to be one

of the most effective information criteria to determine the correct number of latent classes

with GMMs. By contrast, Palardy and Vermunt (2010) found BICB to be more effective

than BIC and AIC3 to be more effective than AIC. Anderson (2008) recommends *against*

using BIC for multimodel selection exercises (see also Burnham & Anderson, 2002) but

its performance has been shown to be quite reliable and robust when used in simulations

involving the types of models that were built and tested in this simulation, specifically

because the correct model is known to be among those in the model space (see Anderson,

2008).

Results of model fitting with M*Plus* Ver 6.1 (Muthén & Muthén, 2010) were

summarized as the percent of occasions, of 100 samples fitted, that each index identified

a given model (of the four fitted) to be the best model over the replication. This summary is shown in Table 5 in Chapter 4, where pilot results supporting this research are described.

3.7 Analysis

Parameter estimates from M*Plus* analysis were processed in SAS 9.2 (SAS Institutes, 2009-2010). M*Plus* provides the estimates of individual growth parameters (level 1), cluster level effects (i.e., intercepts and slope –level 2), and fit information (overall model). Group level effects for the MLLGM were derived from the fast-growth latent class. A SAS program then converted group level effects (i.e., γ_{11}) to quintile rank, computed the bias using Equation 44, the variance of the group level effect, and constructed 90% confidence interval (90% CI) over the replications for that model.

$$B(\bar{\hat{\theta}}) = \bar{\hat{\theta}}_{est} - \theta_{true} \tag{44}$$

3.7.1   Outcomes of Interest: Parameter Recovery

This section describes the methods used to summarize bias, variance, and 90% CI for the mean bias estimate over the replications for the true (i.e., two class MLGMM) and mis-specified (i.e., MLLGM) models. The purpose of this analysis was to investigate systematic trend in biases, not to identify "significant effects" of simulation conditions (i.e., cluster subtype, cluster size, and model types), which were evaluated as described in section 3.5. Pilot work, described in the next chapter, provides an example of how parameter recovery was summarized, shown in Tables 7 and 8. For the main study, visual representations of the information described by summaries like those in Tables 7 and 8 were constructed in order to facilitate the interpretation of overall trends in bias, if any emerged from the simulations and the model fitting in the main study.

3.7.2   Outcome of Interest: Classification error at quintile level

Sanders and Rivers (1996) evaluated the stability of estimated teacher's effect at the quintile level. This study proposes to also utilize the quintile level evaluation in order to examine the estimation accuracy between the incorrectly specified model (i.e., MLLGM) and correctly specified model (i.e., 2-class MLGMM). The classification rate at the quintile level comparing true (i.e., quintile rank based on simulation criteria) and estimated (i.e., rank estimated in the model that is identified) values were summarized using weighted Kappa, which penalizes disagreements more when the classification falls further from the diagonal (perfect agreement). This is shown using the pilot data in Chapter 4, Table 9.

3.8   Evaluation of simulation: Achievement of stated design aims

Analyses of variance (ANOVAs) were conducted in the pilot study described in Chapter 4 in order to examine the four design factors that had also been studied by Chen et al. (2010): cluster number, cluster size, mixture proportion, and cluster effect and their interactions. The present study investigated the significance of various factors that may influence the biases in parameter estimates among simulation conditions and their interactions, as well as exploring the functionality of BIC, relative to other information criteria, in the MLGMM context – thereby integrating and refining results from Muthén and Asparouhov (2009) and Palardy and Vermunt (2010). Including the ANOVA established how and whether the results from this study are comparable to those of Chen et al. (2010). The present study was therefore contextualized with prior work and poised to build on it.

3.9 Summary of Methods

This chapter described the design features of this simulation study to investigate the effect of unmodeled heterogeneity at the individual level on the precision of estimation at higher levels in an MLGMM framework representing a generic VAM type analysis. The focus of this study was not to examine the effect of design factor at global level (e.g., significance with ANOVA) but rather, to identify the potential patterns of estimation biases and imprecision at the group parameter estimates level (e.g., level 2 parameter estimation) that arise when mis-specified mixture modeling is used. One of the purposes of this study was to determine if the precision of estimates from MLGMM warrants further investigation in real data, particularly in the context of the teacher evaluation with VAM. The pilot study illustrating these methods and testing the fidelity of the code to the design features is described in Chapter 4.

Chapter 4: Pilot Study: testing simulation features and analysis plan

A pilot study was performed to test the accuracy of programs designed for the simulation study and to identify potential flaws in the simulation design including the presence and extent of any model convergence problems.

4.1 Testing simulation features

There were three main steps in this pilot study: 1) data generation, 2) estimation, and 3) analysis. SAS macro programs were written for each step: for data generation; to run M*Plus* for the estimation of models; to process results from the M*Plus* model estimation; and to generate the analysis. An additional program was written to test the quality/ensure the fidelity of the simulation code by comparing simulated data against the specification of simulation for all conditions. An error identification program was also written to examine all M*Plus* output to identify convergence issues. This program created a list of simulation conditions resulting in convergence issues, and also automatically reran analyses whenever a convergence issue was encountered. All programs were written to automate these processes and were controlled by the Excel specification file for simulation conditions that were run and analyzed. The Excel file roughly approximated Table 1. All simulation conditions were tested with at least five replications until no modifications were indicated; a very small 40 sample pilot was then run.

The purpose of these tests, run with just five replications, was to verify that all codes – including data generation, estimation, error identification, and analysis – were working properly. These tests led to code modifications. When no further modifications were deemed necessary, the code was run on 40 replications, and the pilot results described below are based on these 40 replications of each model.

The following sections summarize the preliminary results from the mixture proportion (MP=1) including all conditions on three other effects, three cluster numbers , two cluster sizes , and the five mixture proportions. The results were summarized for the 40 replications. This pilot study represented 1/5 of the entire study and took roughly 30 hours.

4.2 Preliminary results on the model identification and class identification

Convergence issues were anticipated with the simulation conditions with the smaller cluster size (i.e., 20) and the mixture proportion condition 3 (MP=3) and with more variable random cluster effect condition (i.e., $\gamma_{11} \sim N(0,1.0)$). Table 5 summarizes the model identification rate by six information criteria described in chapter 3. The preliminary results clearly show that BIC does not perform well for most conditions, which is surprising given its applicability to simulation studies (i.e., it works only when the true model is known to be among those in question) and its excellent performance in other work (e.g., Muthén & Asparouhov; 2009). Table 6 shows the class identification (i.e., fast or slow growth) for the mixture proportion, MP=1, condition. The models with four- and five- latent classes are also performed for the model identification study. However, due to the large proportion of models with the convergence issues including the negative variance in the parameter estimates, zero case in one or more latent classes, or the model convergence problem (i.e., model did not converge). Only 15% of four class models (i.e., 6 out of 40 replications) and none of five class models converged without the issues, therefore only two- and three-class models were used in the main study described in Chapter 5.

Table 5

Pilot Study Results: Percent of Correct Model Identification by Six Information Criteria

for Specified Simulation Conditions: MP1

Simulation Condition				Information Criteria					
Mixture Proportion	Cluster Effect	Cluster Number	Cluster Size	AIC	AIC3	AICc	BIC	BICB	SABIC
1	1	30	20	32.5	22.5	27.5	0	12.5	17.5
			40	45	27.5	47.5	0	22.5	15
		60	20	62.5	60	65	0	22.5	22.5
			40	60	65	62.5	0	52.5	35
		90	20	62.5	77.5	62.5	0	62.5	67.5
			40	67.5	87.5	67.5	17.5	87.5	72.5
	2	30	20	25	10	22.5	0	5	10
			40	60	17.5	60	0	15	10
		60	20	50	37.5	55	0	25	27.5
			40	60	77.5	60	2.5	47.5	25
		90	20	62.5	75	67.5	2.5	37.5	47.5
			40	70	87.5	70	7.5	77.5	52.5
	3	30	20	17.5	7.5	17.5	0	0	5
			40	57.5	45	55	0	37.5	22.5
		60	20	52.5	47.5	52.5	0	15	17.5
			40	67.5	80	70	2.5	60	32.5
		90	20	50	75	50	0	35	42.5
			40	67.5	87.5	67.5	7.5	70	57.5
	4	30	20	30	12.5	25	0	7.5	12.5
			40	37.5	25	35	0	12.5	7.5
		60	20	52.5	50	55	0	15	25
			40	72.5	87.5	77.5	10	50	37.5
		90	20	72.5	75	72.5	5	42.5	47.5
			40	72.5	92.5	75	2.5	82.5	72.5
	5	30	20	52.5	27.5	45	0	17.5	20
			40	62.5	55	62.5	0	47.5	37.5
		60	20	45	50	50	2.5	50	50
			40	57.5	90	60	32.5	95	85
		90	20	60	92.5	62.5	7.5	75	77.5
			40	70	87.5	70	47.5	90	87.5

Table 6

Pilot study results: Recovery Rate of Latent Class for Specified Simulation Conditions:

MP1

Simulation Condition				
Mixture Proportion	Cluster Effect	Cluster Number	Cluster Size	Class Identification Rate (%)
1	1	30	20	81.04
			40	81.27
		60	20	81
			40	82.6
		90	20	81.42
			40	83.44
	2	30	20	79.53
			40	80.83
		60	20	82.96
			40	80.86
		90	20	81.5
			40	82.01
	3	30	20	78.8
			40	82.79
		60	20	82.8
			40	82.11
		90	20	82.44
			40	83.22
	4	30	20	79.08
			40	77.03
		60	20	79.95
			40	82.03
		90	20	80.97
			40	82.53
	5	30	20	76.83
			40	82.93
		60	20	82.59
			40	82.16
		90	20	83.57
			40	83.74

The average latent class recovery rate was around 80%, and large cluster size seemed to moderately increase the recovery rate. There were no convergence issues for MLLGM and the two-class MLGMM, but there were a few convergence issues for the mis-specified three-class MLGMM, when the cluster size was small (i.e., CS=20). It was found that MLGMM improved the estimates of level-2 parameters even without strong identification of individual latent classes. Therefore the latent class recovery study was not included in the final analysis.

4.3 Preliminary results on precision of estimates

Tables 7 and 8 show the mean bias and standard deviation of bias and corresponding 95% CIs for each true cluster effect (i.e., -1, -0.5, 0, 0.5 and 1) for all cluster sizes and numbers on MP=1 (Table 3) and CE=1 (Table 4) conditions. The biases were smaller, considerably smaller for some cases, for MLGMM as compared to MLLGM; the variability of bias was slightly larger for MLGMM, which could be due to the mis-identification of the individuals (i.e., fast and slow growth). In addition, the estimation parameters coded into the M*Plus* programs used here were still preliminary, which might have contributed to the inaccuracy in estimates.

Overall, these pilot results suggest that MLGMM could be a useful method to reduce the bias in the estimate of VAM effects in this simulation condition.

Table 7

Bias, Error, CI of group estimates: Mixture Proportion 1 (MP1) and Cluster Effect 1

(CE1) for the true model

			Cluster Type																		
			1						2						3						
			Bias			Error			Bias			Error			Bias			Error			
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
30	20	-1	-0.34	-0.69	-0.06	0.2	0.01	0.29	-0.22	-0.58	0.05	0.18	0.01	0.25	-0.12	-0.43	0.17	0.18	0.01	0.26
		-0.5	-0.13	-0.39	0.17	0.19	0	0.25	-0.12	-0.48	0.17	0.19	0	0.32	-0.1	-0.36	0.13	0.17	0	0.3
		0	0.08	-0.08	0.25	0.38	0.22	0.49	0.02	-0.21	0.29	0.36	0.22	0.44	-0.03	-0.23	0.18	0.15	0.01	0.2
		0.5	0.24	-0.04	0.55	0.22	0	0.39	0.1	-0.23	0.37	0.19	0	0.34	0	-0.25	0.24	0.16	0	0.33
		1	0.42	0.05	0.96	0.23	0.01	0.25	0.24	-0.05	0.72	0.2	0.01	0.25	0.03	-0.19	0.28	0.17	0	0.25
	40	-1	-0.28	-0.52	0.02	0.17	0	0.23	-0.19	-0.49	0.05	0.15	0	0.23	-0.08	-0.24	0.12	0.13	0	0.2
		-0.5	-0.1	-0.25	0.13	0.14	0.01	0.21	-0.09	-0.28	0.12	0.13	0	0.15	-0.07	-0.26	0.13	0.13	0.01	0.23
		0	0.09	-0.04	0.2	0.39	0.27	0.49	0.02	-0.23	0.2	0.37	0.2	0.45	-0.04	-0.2	0.11	0.12	0.01	0.22
		0.5	0.24	-0.12	0.46	0.17	0	0.26	0.07	-0.09	0.35	0.14	0	0.23	-0.01	-0.29	0.21	0.15	0	0.29
		1	0.39	0.03	0.81	0.19	0	0.23	0.11	-0.07	0.4	0.15	0.01	0.26	0	-0.17	0.17	0.11	0	0.19
60	20	-1	-0.31	-0.63	-0.11	0.2	0.03	0.25	-0.22	-0.67	-0.01	0.21	0.07	0.3	-0.15	-0.38	0	0.17	0.03	0.26
		-0.5	-0.14	-0.31	0.06	0.19	0.04	0.27	-0.1	-0.29	0.06	0.17	0.04	0.27	-0.12	-0.23	0.06	0.16	0.04	0.22
		0	0.09	-0.03	0.19	0.38	0.2	0.45	0	-0.13	0.18	0.36	0.24	0.43	-0.04	-0.21	0.14	0.17	0.06	0.26
		0.5	0.21	0.03	0.53	0.21	0.06	0.31	0.08	-0.07	0.28	0.18	0.05	0.27	0.02	-0.14	0.25	0.18	0.05	0.25
		1	0.44	0.16	0.91	0.24	0.05	0.32	0.19	-0.05	0.58	0.2	0.06	0.29	0.06	-0.09	0.28	0.17	0.03	0.25
	40	-1	-0.29	-0.63	-0.14	0.15	0.02	0.2	-0.18	-0.52	-0.01	0.15	0.02	0.18	-0.11	-0.28	0.05	0.13	0.01	0.21
		-0.5	-0.09	-0.32	0.05	0.15	0.03	0.22	-0.09	-0.25	0.03	0.14	0.03	0.21	-0.07	-0.29	0.13	0.14	0.04	0.19
		0	0.1	-0.01	0.18	0.39	0.2	0.46	0.02	-0.08	0.1	0.37	0.27	0.43	-0.04	-0.2	0.11	0.12	0.02	0.22
		0.5	0.24	0.09	0.56	0.15	0.04	0.2	0.08	-0.03	0.33	0.14	0.02	0.23	-0.02	-0.15	0.1	0.12	0.03	0.17
		1	0.38	0.13	0.88	0.18	0.03	0.25	0.11	0	0.43	0.14	0.01	0.2	-0.01	-0.13	0.13	0.14	0.03	0.24
90	20	-1	-0.32	-0.62	-0.14	0.2	0.1	0.24	-0.2	-0.51	-0.04	0.2	0.06	0.27	-0.13	-0.32	0.02	0.17	0.06	0.23
		-0.5	-0.12	-0.31	0.09	0.21	0.07	0.29	-0.1	-0.27	0.04	0.19	0.07	0.27	-0.11	-0.27	0.06	0.18	0.09	0.27
		0	0.09	0	0.2	0.39	0.23	0.47	0.02	-0.12	0.16	0.37	0.27	0.44	-0.03	-0.13	0.12	0.17	0.08	0.24
		0.5	0.22	0.01	0.43	0.21	0.08	0.34	0.09	-0.05	0.37	0.2	0.09	0.27	0.02	-0.08	0.13	0.16	0.07	0.27
		1	0.41	0.21	0.85	0.24	0.09	0.34	0.19	0.04	0.46	0.2	0.06	0.3	0.04	-0.15	0.15	0.17	0.03	0.25
	40	-1	-0.28	-0.4	-0.09	0.15	0.04	0.22	-0.15	-0.29	-0.03	0.14	0.05	0.19	-0.09	-0.2	0	0.13	0.05	0.19
		-0.5	-0.1	-0.2	0.03	0.13	0.05	0.18	-0.08	-0.21	0.07	0.13	0.05	0.19	-0.07	-0.17	0	0.12	0.04	0.19
		0	0.1	0.03	0.15	0.41	0.34	0.45	0.03	-0.03	0.1	0.38	0.27	0.43	-0.06	-0.18	0.05	0.12	0.05	0.18
		0.5	0.22	0.05	0.36	0.16	0.07	0.21	0.06	-0.07	0.17	0.13	0.05	0.2	-0.02	-0.13	0.08	0.13	0.06	0.19
		1	0.38	0.19	0.57	0.18	0.05	0.24	0.13	0.01	0.22	0.14	0.06	0.16	-0.01	-0.12	0.1	0.12	0.06	0.17

Table 8

Bias, Error, CI of Group Estimates: Mixture Proportion 1 (MP1) and Cluster Effect (CE1) for Misspecified Model

			Cluster Type															
			1						2						3			
			Bias			Error			Bias			Error			Bias			Error		
Cluster Effect	Cluster Number	Cluster Size	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
30	20	-1	-0.46	-0.78	-0.23	0.16	0	0.29	-0.32	-0.56	0.02	0.14	0	0.25	-0.15	-0.42	0.19	0.17	0	0.24
		-0.5	-0.19	-0.42	0.01	0.15	0.01	0.24	-0.17	-0.54	0.07	0.18	0.01	0.31	-0.13	-0.41	0.08	0.16	0	0.24
		0	0.08	-0.05	0.3	0.32	0.22	0.41	0.01	-0.14	0.22	0.35	0.21	0.51	-0.05	-0.26	0.18	0.13	0	0.21
		0.5	0.32	0.06	0.56	0.16	0	0.35	0.15	-0.12	0.39	0.16	0.01	0.23	-0.02	-0.24	0.19	0.15	0.01	0.27
		1	0.58	0.26	0.84	0.16	0	0.19	0.33	0.05	0.54	0.16	0.01	0.23	0.02	-0.32	0.26	0.17	0	0.25
	40	-1	-0.41	-0.57	-0.16	0.13	0.01	0.16	-0.31	-0.48	-0.17	0.11	0	0.21	-0.12	-0.3	0.04	0.11	0.01	0.2
		-0.5	-0.16	-0.4	0.04	0.14	0.01	0.2	-0.15	-0.29	0.05	0.12	0	0.17	-0.1	-0.26	0.1	0.12	0	0.24
		0	0.1	-0.02	0.21	0.31	0.23	0.38	0.01	-0.09	0.14	0.33	0.17	0.45	-0.06	-0.29	0.12	0.12	0	0.18
		0.5	0.33	0.14	0.55	0.13	0.01	0.21	0.14	-0.07	0.34	0.12	0.01	0.2	-0.04	-0.25	0.15	0.14	0	0.26
		1	0.6	0.31	0.84	0.14	0	0.2	0.25	0.11	0.4	0.11	0	0.21	-0.03	-0.17	0.15	0.11	0	0.24
60	20	-1	-0.45	-0.78	-0.27	0.16	0.04	0.22	-0.32	-0.68	-0.16	0.17	0.04	0.26	-0.18	-0.46	0.01	0.16	0.03	0.24
		-0.5	-0.2	-0.49	-0.05	0.17	0.03	0.28	-0.15	-0.36	0.04	0.15	0.03	0.25	-0.13	-0.4	0.05	0.16	0.03	0.24
		0	0.1	-0.07	0.19	0.32	0.22	0.4	0.01	-0.2	0.14	0.34	0.2	0.43	-0.06	-0.33	0.07	0.17	0.04	0.27
		0.5	0.31	0.06	0.48	0.16	0.04	0.23	0.16	-0.17	0.32	0.17	0.07	0.27	0	-0.26	0.19	0.16	0.04	0.24
		1	0.63	0.34	0.81	0.16	0.03	0.24	0.33	0.07	0.52	0.15	0.04	0.25	0.06	-0.22	0.2	0.17	0.04	0.25
	40	-1	-0.43	-0.58	-0.29	0.12	0.03	0.2	-0.29	-0.42	-0.1	0.12	0.02	0.17	-0.16	-0.29	0.06	0.11	0.02	0.19
		-0.5	-0.16	-0.31	0.02	0.13	0.02	0.2	-0.15	-0.26	0.02	0.12	0.02	0.2	-0.12	-0.25	0.12	0.13	0.03	0.19
		0	0.1	0.04	0.24	0.32	0.22	0.36	0.01	-0.06	0.11	0.33	0.21	0.49	-0.09	-0.2	0.09	0.11	0.02	0.16
		0.5	0.33	0.2	0.47	0.11	0.02	0.17	0.14	-0.02	0.4	0.13	0.02	0.23	-0.06	-0.16	0.15	0.11	0.03	0.15
		1	0.57	0.41	0.73	0.12	0.03	0.17	0.25	0.13	0.4	0.1	0.02	0.15	-0.04	-0.14	0.19	0.13	0.04	0.22
90	20	-1	-0.47	-0.58	-0.08	0.17	0.06	0.24	-0.32	-0.45	-0.01	0.17	0.04	0.24	-0.17	-0.3	0.14	0.15	0.07	0.2
		-0.5	-0.2	-0.37	0.24	0.17	0.08	0.22	-0.15	-0.29	0.07	0.15	0.07	0.25	-0.14	-0.27	0.25	0.17	0.05	0.23
		0	0.1	0	0.32	0.33	0.26	0.42	0.02	-0.09	0.2	0.35	0.23	0.47	-0.05	-0.16	0.14	0.15	0.04	0.23
		0.5	0.33	0.24	0.73	0.16	0.1	0.26	0.16	0.02	0.52	0.17	0.03	0.27	0	-0.11	0.34	0.16	0.06	0.44
		1	0.61	0.45	0.76	0.15	0.04	0.24	0.32	0.18	0.89	0.16	0.07	0.24	0.02	-0.14	0.41	0.16	0.04	0.22
	40	-1	-0.42	-0.55	-0.21	0.12	0.06	0.16	-0.27	-0.38	0.01	0.13	0.05	0.19	-0.14	-0.23	0.16	0.12	0.05	0.16
		-0.5	-0.17	-0.28	0.08	0.12	0.04	0.16	-0.14	-0.25	0.06	0.12	0.04	0.16	-0.12	-0.2	0.17	0.11	0.04	0.16
		0	0.11	0.06	0.22	0.32	0.25	0.37	0.03	-0.03	0.14	0.34	0.22	0.53	-0.1	-0.2	0.13	0.12	0.05	0.17
		0.5	0.33	0.24	0.53	0.12	0.05	0.16	0.13	0.03	0.42	0.11	0.04	0.2	-0.06	-0.17	0.21	0.12	0.05	0.18
		1	0.59	0.5	0.82	0.12	0.06	0.19	0.28	0.17	0.67	0.12	0.04	0.15	-0.04	-0.15	0.22	0.12	0.05	0.16

4.4 Preliminary results on classification accuracy

Table 9 shows the classification accuracy of cluster effects (i.e., value-added effect) ranked by quintile level. The range of weighted kappa varied from .61 to .83 with

an average of .73 for the true model (i.e., two-class MLGMM), For MLLGM, the range

for weighted kappa was .54 to.71 with an average of .64. The MLGMM had a higher

classification rate than MLLGM for this simulation.

Table 9

Rate of Misclassification at the Quintile Level

Simulation Condition				Kappa	
Mixture Proportion	Cluster Effect	Cluster Number	Cluster Size	MLGMM	MLLGM
1	1	30	20	0.76	0.67
			40	0.82	0.71
		60	20	0.77	0.67
			40	0.82	0.71
		90	20	0.77	0.67
			40	0.83	0.70
	2	30	20	0.71	0.66
			40	0.77	0.68
		60	20	0.73	0.66
			40	0.76	0.68
		90	20	0.73	0.66
			40	0.76	0.68
	3	30	20	0.67	0.61
			40	0.75	0.65
		60	20	0.70	0.62
			40	0.75	0.65
		90	20	0.70	0.61
			40	0.75	0.65
	4	30	20	0.61	0.54
			40	0.65	0.57
		60	20	0.63	0.55
			40	0.69	0.59
		90	20	0.62	0.54
			40	0.69	0.58
	5	30	20	0.67	0.60
			40	0.73	0.63
		60	20	0.73	0.64
			40	0.75	0.66
		90	20	0.74	0.64
			40	0.75	0.65

4.5 Preliminary results on ANOVA over the simulation condition

There were not enough simulation conditions to conduct the full ANOVA as was

done for the main study. However, the preliminary results indicated that the cluster effect,

cluster size, and all two-way interaction effects had a significant effect (and $p<0.01$) on the mean bias estimates. The cluster number appeared not to have a significant effect. The ANOVA effects were used for the generation of a plot representing, and so facilitating, the interpretation of biases in estimates between the true and mis-specified models.

4.6 Determination of number of replications for the study

The number of replications to be used in the main study was determined by examining the precision of estimates described in section 4.3 over different numbers of replications: 20, 40, 80, 100, 200, and 400. Table 10 shows the summary of 90% confidence interval and the standard deviation of bias estimates from cluster effect condition 1 and mixture proportion condition 1. The variation of standard deviations and the range of confidence interval converged around 100 replications, indicating that 100 replication would be sufficient for the main study.

Table 10

Precision of Bias Estimates Over a Different Number of Replications

		Number of Replications					
Cluster Type	Bias Estimates	20	40	80	100	200	400
1	5%	-0.36	-0.43	-0.37	-0.44	-0.43	-0.43
	90%	4.01	2.2	2.53	2.17	2.33	2.35
	SD	1.25	0.81	1.23	1.04	1.11	1.15
2	5%	-0.89	-0.76	-0.65	-0.65	-0.65	-0.65
	90%	3.57	2.13	2.2	2.28	2.15	2.25
	SD	1.31	0.85	1.29	1.1	1.14	1.17
3	5%	-1.47	-1.93	-1.64	-1.79	-1.66	-1.7
	90%	3.28	0.82	1.26	1.06	1.08	1.16
	SD	1.5	0.8	1.36	1.15	1.13	1.19

4.7 Other Analysis Issues

The convergence and analysis issues were identified by examining M*plus* output.

The first step was to examine if the output files included models for which estimates had

not been generated, which indicates non-convergence of that model. The second step was

to read in estimates to: 1) identify negative variances and 2) latent classes with zero

cases, both of which represent non-informative convergence of that model. New data was

generated to replace data that had resulted in one or more of these indicators of

convergence problems, and this process was continued until 100 successful replications

for each simulation condition were completed.

4.8 Pilot study summary

This small pilot study demonstrated that the automated procedures for generating,

manipulating, and analyzing data according to the simulation characteristics outlined in

Chapter 3/Table1 worked, and that the fidelity of simulated data to the simulation design

was high. Convergence issues were only encountered in the contexts in which they were

anticipated (i.e., model fitting with over-fitting 3-class MLGMM with small cluster size

(20)), and in no other context. It is important to note that there was no convergence issue

for the model of interest, two-class MLGMM, in which the parameters of interest were

estimated. Therefore the impact of the minor convergence issues with 3-class MLGMM

was deemed to be minimal. Each of the summary features described in Chapter 3

functioned in this pilot data to yield interpretable outcomes and no issues were

encountered that were not A) expected and B) easily addressed. In summary, the pilot

results supported the likelihood that the main study would be completed as planned and

that the impacts on effects and estimates would be representative of the simulation design

outlined in Chapter 3.

Chapter 5: Main Study Results

The main study commenced once the pilot study concluded. As outlined in the previous chapter, the code was found to provide high fidelity to the simulation objectives; specific changes to the design were to generate 100 replications of each model, and to utilize four (reduced from five) mixture proportion conditions. Therefore, there were 120 simulation conditions for the main study, with 100 trials (or samples) per condition, yielding 12,000 datasets (i.e., 120 simulation conditions × 100 replications) that were generated and analyzed, as outlined in Chapter 3, with three MLGMMs, that is, with 1-, 2-, or 3- latent classes (n.b., the 1-latent class MLGMM is equivalent to MLLGM)**.** The bias estimates (i.e., the mean of bias over 100 replications) were computed using Equation 44 and 90% confidence intervals were constructed, using the variance of these 100 cluster level estimates as the measure of error, from the mis-specified model (i.e., MLLGM or 1-class MLGMM) and from the true model (i.e., 2-class MLGMM). ANOVA was performed to compare bias in cluster level effect estimates between the mis-specified model and the true model across the simulation conditions of interest. This study took approximately 580 hours of continuous computation time and, with the exceptions noted above, was carried out exactly as described in Chapters 3 and 4, using the code that had been written for the pilot study in Chapter 4.

Failure of models to converge can be an issue in mixture models (i.e., the 2- or 3-class MLGMMs in this simulation), particularly when more latent classes are extracted than the true model has, as indicated in the pilot study. However, in the main study, no convergence issues occurred for MLGMMs with 1- or 2-latent classes. Table 11

summarizes the convergence problems that were encountered over the three sets of 12,000 replications (i.e., 36,000 total).

There were 32 convergence issues out of total of 12,000 estimations with the MLGMM with 3 latent classes, and these only occurred for the cluster effect conditions 4 and 5; the majority (22 out of 32) of errors happened when the cluster size was 20. These convergence issues were addressed as described in Chapter 4 and so the results that follow describe model fits and parameter estimates from the 3-class MLGMM fit to all 12,000 replications.

Table 11

Convergence Rate by Simulation Conditions

Mixture Proportion	Cluster Effect	Cluster Number	Cluster Size	Error
1	4	30	20	2
			40	1
		60	20	4
2	4	60	20	1
			40	1
		90	40	1
2	5	30	40	1
		60	20	1
3	4	30	40	1
		60	40	1
3	5	30	20	2
			40	1
		60	20	2
		90	20	2
4	4	30	20	1
			40	1
		60	20	1
			40	2
4	5	30	20	5
		60	20	1

5.1 Results on model identification

The first step of the analysis was to identify the number of latent classes extracted from the model. The true number of latent classes was always two. Therefore, model identification in this study was summarized as the rate of correctly selecting the two-class MLGMM across all model fits, utilizing the six information criteria described in Chapter 3. The recovery of the individual-level latent class membership was deemed in the pilot study not to be important for the research questions, and so this data was not captured for any replication, and did not contribute to the estimation of correct model identification.

Tables 12 to 16 summarize the model identification rates derived from the six information criteria described in Chapter 3, ordered by cluster effect (CE1 through 5). These five tables each include all mixture proportion (MP1 to 4) conditions, cluster numbers (CN=30, 60, and 90), and cluster sizes (CS=20 and 40). The rates of successful model identification were computed as each information criterion's (correct) selection of the two-class MLGMM over the 100 replications in each set of conditions; higher values are better identification rates.

AIC and AICc performed very similarly in almost all conditions, except that AICc performed slightly better when the cluster size was smallest (i.e., CS=20). The model identification rates of AIC and AICc were relatively consistent across simulation conditions, but they performed best, where both the cluster number and cluster size were small (i.e., CN=30 and CS=20 or 40). However, the AIC and AICc identification rates were moderate, 35-65% in all conditions.

AIC3, BICB, and SABIC had similar identification rates in all conditions. AIC3 consistently had higher rates for mixture proportion conditions 1 and 2 (i.e., different mixture proportion among cluster types) and when the sample size is smaller (i.e., the sample size, $CN \times CS < 1200$). BICB and SABIC performed particularly well when the mixture proportion was consistent among cluster types (MP3 and MP4) and sample size is large (i.e., $CN \times CS > 1800$). SABIC performed slightly worse than BICB on mixture proportion conditions 1 and 2, and with large cluster size (CS=40), but this difference between SABIC and BICB was very small with small cluster size (CS=20).

Table 12

Recovery Rate of Latent Class for Specified Simulation Conditions: CE1

Simulation Condition				Information Criteria					
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	AIC	AIC3	AICc	BIC	BICB	SABIC
1	1	30	20	35	31	34	10	25	28
			40	44	36	45	8	27	21
		60	20	46	56	48	7	32	36
			40	62	59	62	9	42	33
		90	20	56	68	60	11	47	49
			40	63	89	64	32	86	79
	2	30	20	42	26	45	12	19	24
			40	41	26	41	9	22	16
		60	20	49	43	48	6	26	29
			40	68	64	67	19	49	36
		90	20	51	67	52	6	42	42
			40	54	75	54	11	69	54
	3	30	20	38	42	39	7	38	39
			40	67	74	67	16	75	64
		60	20	57	80	59	13	66	68
			40	67	90	71	65	96	96
		90	20	65	84	68	43	93	93
			40	72	87	73	92	93	94
	4	30	20	53	56	59	8	52	53
			40	61	86	66	42	88	89
		60	20	64	85	67	48	90	91
			40	69	85	72	92	89	89
		90	20	57	78	58	74	91	91
			40	63	81	63	93	89	92

Table 13

Recovery Rate of Latent Class for Specified Simulation Conditions: CE2

Simulation Condition				Information Criteria					
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	AIC	AIC3	AICc	BIC	BICB	SABIC
2	1	30	20	34	26	36	10	22	23
			40	42	34	41	12	26	15
		60	20	46	46	47	15	27	33
			40	60	58	61	16	46	35
		90	20	63	74	66	14	47	50
			40	61	80	62	10	76	60
	2	30	20	38	27	38	10	26	27
			40	46	51	49	11	44	35
		60	20	60	62	62	24	46	49
			40	50	72	51	15	71	65
		90	20	53	65	54	11	45	49
			40	45	73	46	39	86	82
	3	30	20	47	44	48	5	33	38
			40	66	73	68	10	74	67
		60	20	71	83	71	14	72	74
			40	63	86	63	61	92	92
		90	20	57	77	59	32	78	80
			40	60	86	60	93	93	93
	4	30	20	57	59	59	8	51	56
			40	67	85	69	43	91	92
		60	20	65	86	69	38	87	86
			40	51	75	52	87	89	90
		90	20	68	91	73	73	94	94
			40	70	85	71	95	92	93

Table 14

Recovery Rate of Latent Class for Specified Simulation Conditions: CE3

Simulation Condition				Information Criteria					
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	AIC	AIC3	AICc	BIC	BICB	SABIC
3	1	30	20	46	19	48	9	17	17
			40	42	23	41	4	21	13
		60	20	47	43	48	10	28	29
			40	60	69	60	7	51	42
		90	20	59	75	59	13	44	46
			40	62	88	63	13	79	71
	2	30	20	51	42	55	20	38	39
			40	53	57	54	25	54	48
		60	20	57	63	57	22	51	54
			40	54	62	54	41	65	65
		90	20	57	72	58	29	55	57
			40	57	71	58	60	73	73
	3	30	20	48	42	49	12	34	38
			40	65	75	68	5	72	60
		60	20	59	74	64	7	60	63
			40	59	83	61	62	90	91
		90	20	67	84	67	33	90	90
			40	60	80	60	88	91	91
	4	30	20	59	47	60	6	38	44
			40	64	84	68	31	82	85
		60	20	66	81	72	29	81	84
			40	59	83	59	94	94	95
		90	20	66	84	66	76	93	93
			40	68	86	70	92	91	91

Table 15

Recovery Rate of Latent Class for Specified Simulation Conditions: CE4

Simulation Condition				Information Criteria					
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	AIC	AIC3	AICc	BIC	BICB	SABIC
4	1	30	20	31	14	30	3	10	12
			40	35	16	35	2	13	10
		60	20	44	38	43	4	22	22
			40	45	57	47	3	39	31
		90	20	41	53	42	1	33	34
			40	54	65	54	5	60	49
	2	30	20	39	26	37	11	23	25
			40	40	26	39	4	22	14
		60	20	54	40	53	10	21	25
			40	57	61	57	5	41	34
		90	20	42	58	42	9	30	33
			40	53	70	53	8	62	54
	3	30	20	46	34	46	10	29	31
			40	50	63	52	4	54	45
		60	20	65	71	65	11	62	65
			40	59	87	63	60	90	91
		90	20	52	81	53	30	80	82
			40	45	65	46	78	82	82
	4	30	20	56	54	57	9	49	53
			40	59	79	60	34	78	78
		60	20	53	74	59	24	70	70
			40	63	82	65	85	87	87
		90	20	46	70	47	52	77	77
			40	60	79	60	87	87	87

Table 16

Recovery Rate of Latent Class for Specified Simulation Conditions: CE5

Simulation Condition				Information Criteria					
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	AIC	AIC3	AICc	BIC	BICB	SABIC
5	1	30	20	27	13	27	0	12	12
			40	40	45	42	3	39	31
		60	20	41	53	42	5	40	42
			40	50	67	50	16	61	51
		90	20	40	56	42	11	51	55
			40	32	52	32	25	56	54
	2	30	20	17	12	18	2	8	10
			40	23	19	22	3	18	9
		60	20	28	35	31	5	18	19
			40	23	33	25	0	25	16
		90	20	33	48	35	5	35	40
			40	21	43	22	7	36	32
	3	30	20	43	36	45	9	31	33
			40	53	75	56	6	76	65
		60	20	50	66	54	11	57	59
			40	52	81	53	55	89	90
		90	20	56	81	58	29	83	84
			40	44	74	46	84	83	83
	4	30	20	45	61	49	16	62	62
			40	52	77	54	72	80	80
		60	20	53	79	54	62	84	82
			40	34	59	35	66	65	65
		90	20	54	71	55	76	77	77
			40	37	50	37	60	60	60

5.2 Results on bias of estimates

As described in Chapter 3, the bias of the cluster effect estimates, computed as the difference between the fixed effect (i.e., simulation value) and the cluster effect estimates, from the mis-specified model (i.e., 1-class MLGMM) and from the true model (i.e., 2-class MLGMM), were examined over the simulation conditions. These results, fully presented in Appendix A, are summarized here and discussed in Chapter 6.

Appendix A shows the mean, 90% CIs for the mean, and the standard deviations of bias

as estimated in all simulation conditions. Tables in Appendix A are ordered first by the

cluster effect, then by the mixture proportion, and finally by model (true, mis-specified).

Following is the order of tables:

- True Model (2-latent class MLGMM)
    - Appendix A.1 to1.4 for cluster effect condition 1 and mixture proportion 1 through 4
    - Appendix A.5 to 1.8 for cluster effect condition 2 and mixture proportion 1 through 4
    - Appendix A.9 to 1.12 for cluster effect condition 3 and mixture proportion 1 through 4
    - Appendix A.13 for cluster effect condition 4
    - Appendix A.14 for cluster effect condition 5

- Mis-specified Model (1-latent class MLGMM)
    - Appendix A.15 to 1.18 for cluster effect condition 1 and mixture proportion 1 through 4
    - Appendix A.19 to 1.22 for cluster effect condition 2 and mixture proportion 1 through 4
    - Appendix A.23 to 1.26 for cluster effect condition 3 and mixture proportion 1 through 4
    - Appendix A.27 for cluster effect condition 4
    - Appendix A.28 for cluster effect condition 5

The results are discussed below by cluster effect in the subsequent section, and

summarized by figures capturing the salient features of, and trends in, estimates in order

to address the research questions. The term "true effect" (TE) is used in this section to

represent the individual fixed parameters within each cluster effect (e.g., -1, -0.5, 0, 0.5

and 1 for cluster effect condition 1). Bias in the cluster estimates from each of the 100

replications of the mis-specified and the true models were summarized (mean, standard

deviation), representing the results for each of 120 conditions of this study. These results

were analyzed by 5×3×4×3×2 (i.e., TE×CT×MP×CN×CS) ANOVA, and the relevant

results from the ANOVAs that answer the research questions are described below. Focus

on the simulation conditions reduced the number of figures to use while providing

sufficient information to answer the research questions.

5.2.1   Cluster Effect Condition 1 (CE=1)

Cluster effect condition 1 (CE1) was included to evaluate the effect of differential

mixture proportions, among three cluster types, on the cluster effect estimates from the

mis-specified and the true models. CE1 has the same cluster effects (i.e., -1, -0.5, 0, 0.5,

and 1) in each of three cluster types. The cluster types were defined based on the mixture

proportion. Therefore, differences in bias were not expected among the four mixture

proportion conditions (and this was not tested). Instead, the term of interest is the 3-way

interaction among the three cluster types (CT), four mixture proportions (MP), and five

effects on the cluster effect 1 (TE); the interaction term was included in ANOVA models

that were run both with and without cluster size (CS) and number (CN) included. Both

the 3-way CT×MP×TE and 4-way CS×CT×MP×TE interaction terms were significant at

the $p<.0001$ level, but the 4-way CN×CT×MP×TE term was not significant ($p=0.97$).

These results are summarized in Figures 9 and 10 (see Appendix A for full results).

Each of Figures 9 and 10 includes 15 plots (five TE conditions × three cluster

types) with either CS=20 (Figure 10) or CS=40 (Figure 10). Each plot has two lines

representing the estimates from the true model (Model T, a line with squares) and the

mis-specified model (Model M, a line with circles), for four MP conditions. Each row of

figures is based on the cluster size (20 for figure 9 and 40 for figure 10) and cluster types

(1st row is cluster type 1, 2nd row is cluster type 2, and 3rd row is cluster type 3). Each

column of figures is organized by five TE (i.e., -1, -0.5, 0, 0.5, and 1).
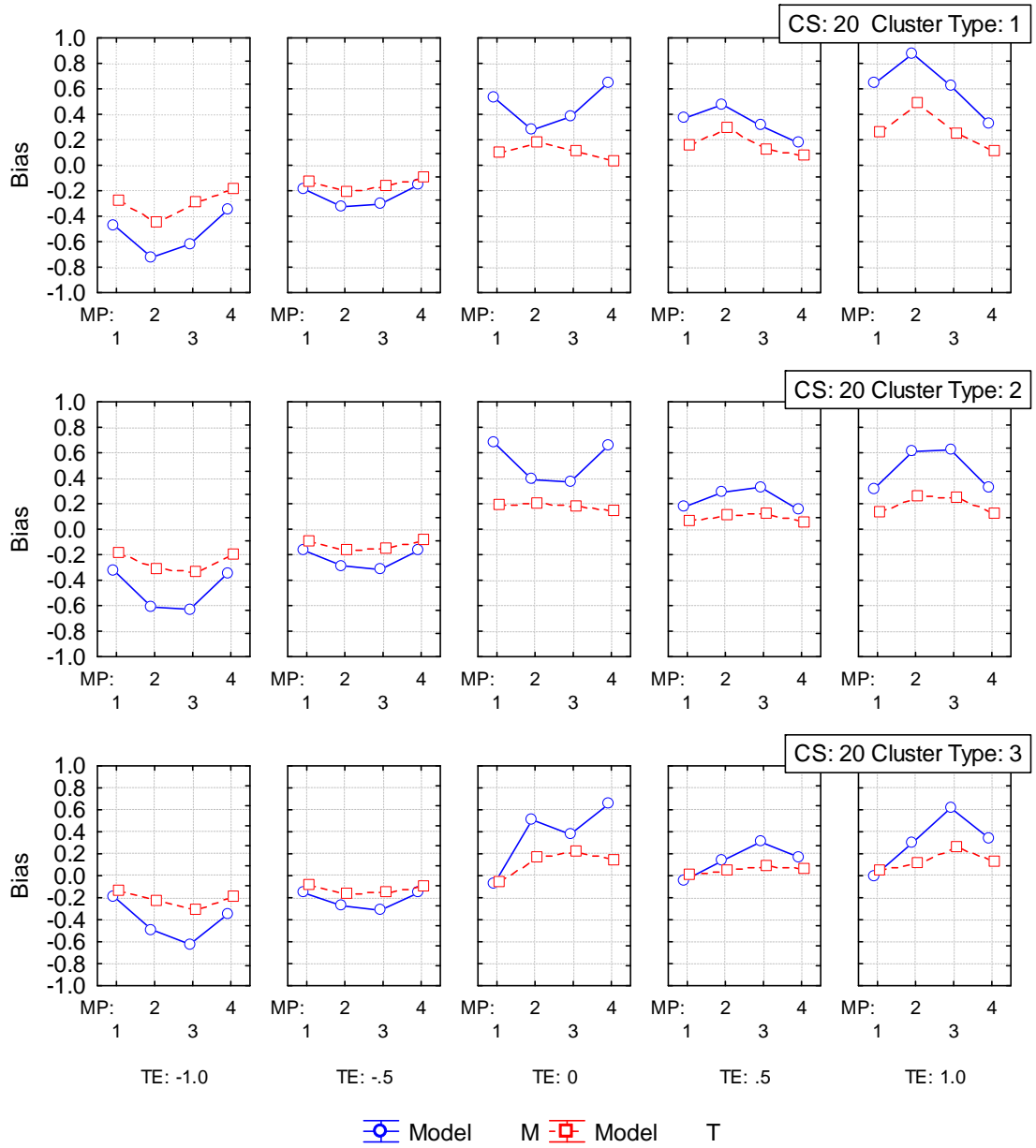
Figure 9. Bias estimates for cluster effect 1 (CE1) and cluster size 20 (CS20)
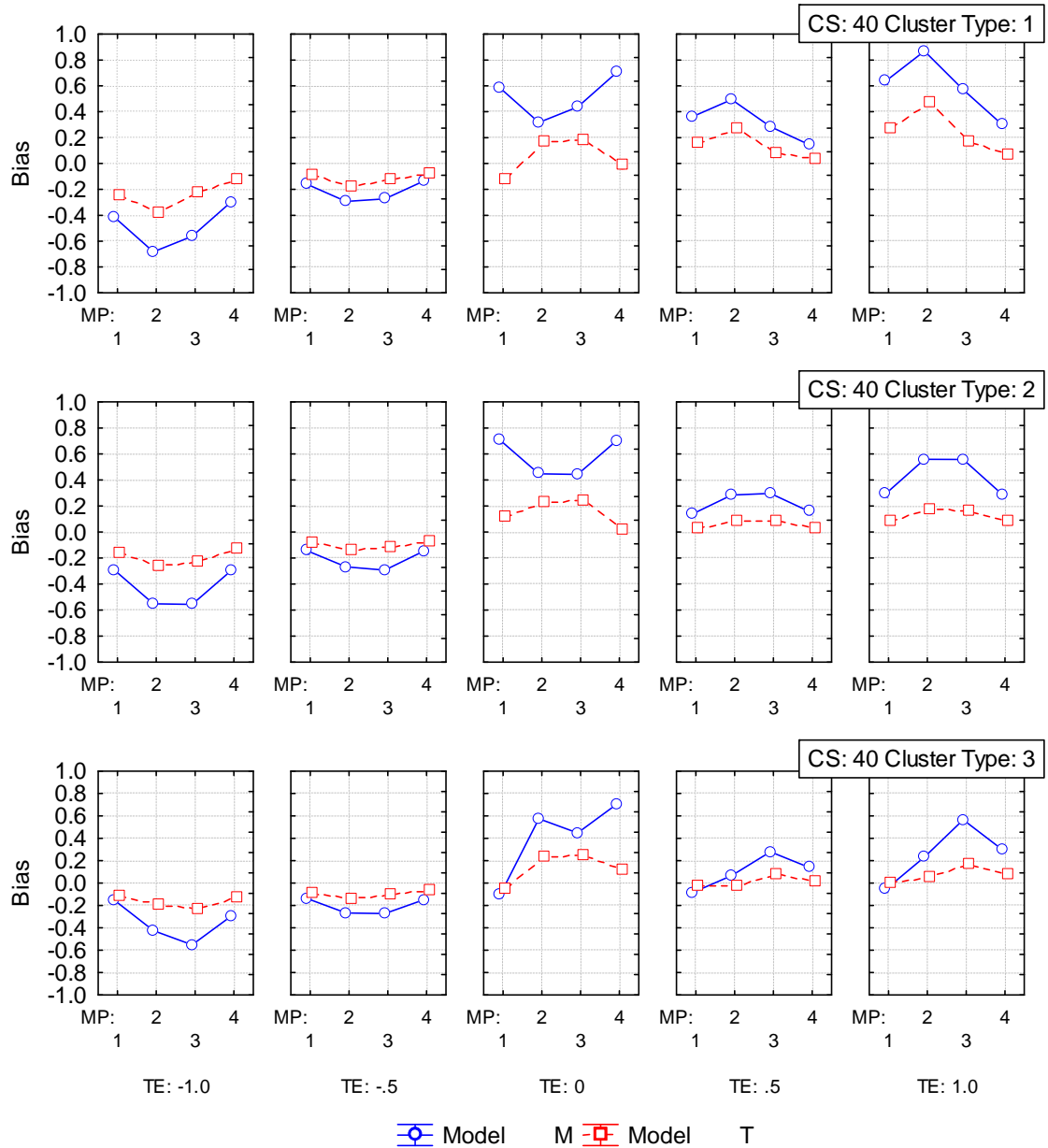
Figure 10. Bias estimates for cluster effect 1 (CE1) and cluster size 40 (CS40)

Positive bias represents the overestimation of TEs and negative bias represents

underestimation of TEs. Figures 9 and 10 show that the bias from the true model was

consistently closer to zero, compared to that of the mis-specified model, for the same data

from all conditions (i.e., the true model yielded more accurate estimation of parameters in

terms of recovery of TE values). The effect of cluster size was minimal (comparing

85

Figures 9 and 10), although the magnitude of differences in bias between the true and mis-specified models was greater when the cluster size was larger (i.e., CS=40, see Appendix A for the detail).

Figures 9 and 10 also show that the patterns of bias over the four mixture proportions in each cluster type was consistent when cluster effects were non-zero, namely, positive bias was observed with positive cluster effects (i.e., TE=0.5 and 1) and negative bias observed with negative effects (i.e., TE=-0.5 and -1). The magnitude of bias was greater for MP2 than MP1, which differed in the amount of variation in mixture proportion. These findings were consistent for the true and mis-specified models although the true model yielded much less bias and the difference between bias in estimates derived from the true and mis-specified models increased as the variability of mixture proportion increased (i.e., from MP1 to MP2). The trend in bias was reversed for the mis-specified model between MP1 and MP2 on cluster types 1 and 2 when the cluster effect was zero because there were more cases with the cluster effect of zero for these conditions due to the cases in the non-growth group. All bias was positive, i.e., all TE were overestimated, when the cluster effect was zero.

The comparison of MP3 and 4 showed that the magnitude of bias decreased as the proportion of cases in the fast growth group increased, except for the mis-specified model when the cluster effect was zero. The difference in bias was greater between the true and mis-specified model as the proportion of the fast group decreased for all non-zero cluster effects.

There was virtually no difference in bias between the true and mis-specified models on cluster type 3 of MP1. All cases were from the fast growth group in this

condition, suggesting that cluster effect estimates were not influenced by the differential mixture proportions in other cluster types.

5.2.2   Cluster Effect Condition 2 and 3 (CE2 and 3)

Cluster effect conditions 2 and 3 were designed to evaluate the potential for systematic bias in the cluster effect estimates that could arise from the same cluster effects in the presence of different mixture proportions. CE2 and CE3 had the same overall cluster effects (i.e., -1, -0.5, 0, 0.5, and 1) but the assignment of effects were reversed between cluster type 1 and 3, while cluster type 2 had zero effects in both CE2 and CE3. The cluster types were again based on the mixture proportions. Therefore, no differences in bias were expected and found between CE2 and CE3 for the mixture proportion conditions MP3 and MP4 (see Figure 11).

The term of interest is the interaction among the two cluster effect conditions (CE), four mixture proportions (MP), and five effects (TE) condition, with or without cluster size (CS) and number (CN). The ANOVA determined whether these interaction terms affected the amount of bias in estimating the five cluster effect conditions by the true and mis-specified models. This $2\times4\times5\times3\times2$ ANOVA found the three-way CE×MP×TE term significant at $p<.0001$, but the four-way CN×CE×MP×TE ($p>.92$) and CS×CE×MP×TE ($p>.49$) terms were not significant. Figure 11 summarizes these results (see Appendix A for full results).

Figure 11 includes a total of eight plots (two cluster effect conditions × four mixture proportions) with two lines representing the estimates from the true model (Model T, a line with squares) and the mis-specified model (Model M, a line with circles).
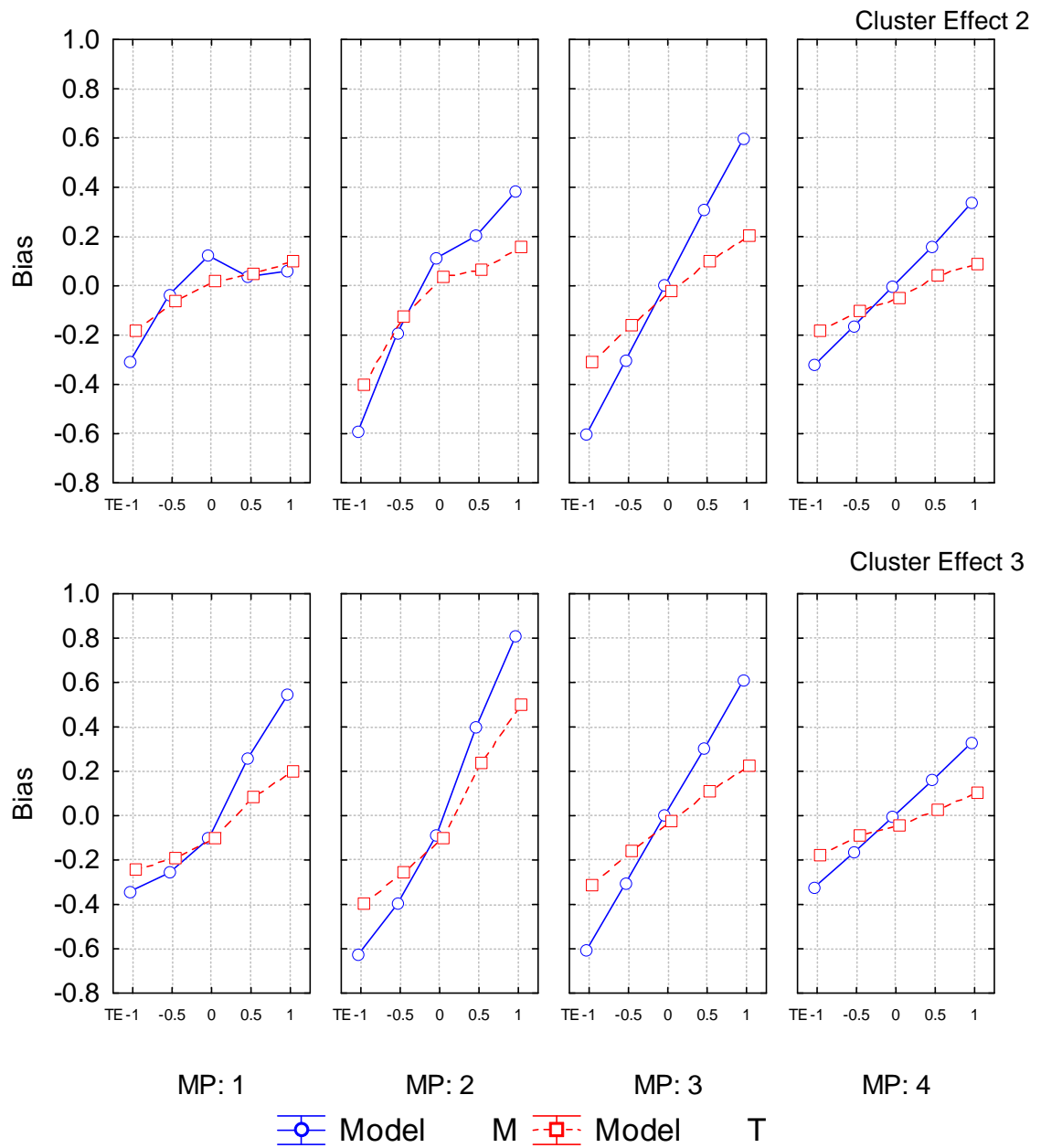
Figure 11. Bias estimates for cluster effects 2 and 3 (CE2 and CE3)

The x-axis of each figure represents five TEs. Each row of plots represents a cluster effect condition (CE2 and CE3). Cluster type was not included in the figure because the true effect (TE) is a direct indicator of cluster types.

A comparison of bias in estimates derived from CE3 vs. CE4 on MP1 and MP2 magnified the effect of cluster types (i.e., differential mixture proportion) that was observed in Figure 11. Cluster type 1 was indicated by TE = -1 and TE= -0.5 for CE3, and TE = 1 and TE=0.5 for CE4. Cluster type 2 had zero TE on both CE levels, and cluster type 3 had the reverse TEs compared to cluster type 1. The MP1 condition represented lower variation in mixture proportion than MP2, and the magnitude of difference in bias was greater between true and mis-specified models in the MP2 condition as compared to the MP1 condition. The negative bias derived from TE -1 and -0.5 was greater for the condition with MP1 and CE3, where more cases were in the fast growth group. The positive bias derived from TE 0.5 and 1 on MP2 was much more pronounced for CE3 (only 25% of cases in the fast growth group) whereas all cases were in the fast growth group in CE2, which had minimal bias in estimates from both the true and mis-specified models. The trends in bias under the MP2 condition were similar to that under MP1, but the magnitude of bias was greater for MP2. The magnitudes of both positive and negative bias increased as the variation in the mixture proportion increased. The most pronounced effect was observed in the MP2 and CE3 conditions where the positive bias with TE 0.5 and 1 was the highest (only 25% of cases were in the fast growth group). The overall variation in mixture proportions increased the magnitude of bias, especially on the non-zero positive TEs.

Similar effects of mixture proportion on bias were observed for MP3 and MP4, where MP3 had a greater magnitude of bias. There was no difference between CE3 and CE4 on MP3 and MP4, as expected, because both MP3 and MP4 had constant mixture proportions across the three cluster types. The variation of mixture proportion between

the fast and slow growth cases was greater on MP3 (50 fast/50 slow) than MP4 (75

fast/25 slow). The trend of bias was symmetric for the negative and positive TEs,

centered around zero bias on TE=0 from both the true and mis-specified models on MP3,

whereas the magnitude of negative bias was greater on MP4 for the true model. That is,

positive bias was attenuated when a greater proportion of the cases were in the fast

growth group.

5.2.3   Cluster Effect Condition 4 and 5 (CE=4 and CE=5)

Cluster effect conditions 4 and 5 were designed to assess the impact of variability

in cluster effect on bias of estimation impact of variability on the cluster effect estimates.

The cluster effects in these two conditions were randomly generated from the normal

distribution with the mean of 0 and a variance of 0.5 for CE4 and 1.0 for CE5.

The term of interest in these analyses was the interaction between the cluster

types (CT), mixture proportion (MP) and cluster effect conditions (CE), with or without

cluster size (CS) and cluster number (CN). The question is whether these interaction

terms are associated with the amount of bias in estimating random cluster effects under

the true and mis-specified models. ANOVA found that the two-way MP×CT term

significant ($p<.0001$), but other terms were not (all $p>0.3$), including that of the CE.

Figure 12 is the only one of these that includes statistically not significant simulation

conditions (i.e., CE) because for this particular analysis, these conditions directly

addressed a research question (see Appendix A for full results). These are discussed in

Chapter 6.
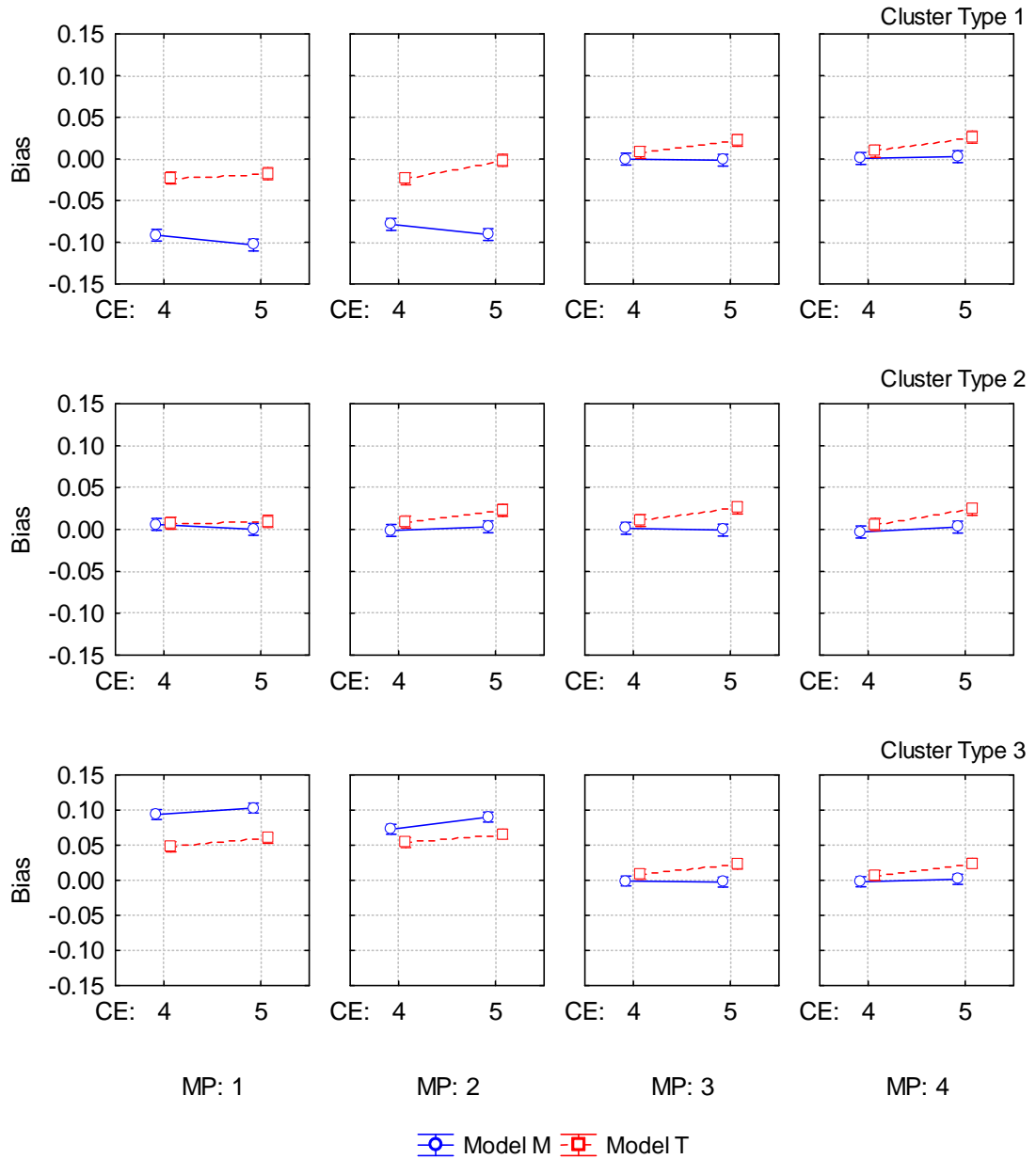
Figure 12. Bias estimates for cluster effect 4 and 5 (CE4 and CE5)

Figure 12 includes a total of 12 plots with four mixture proportions and three cluster types on the x-axis with two lines in each plot representing the estimates from the true model (Model T, a line with squares) and the mis-specified model (Model M, a line with circles). Four figures in each row represent the mixture proportion conditions. Each row of plots represents a cluster type (CT=1, 2, or 3).

The ANOVA results indicated that the variation in true cluster effects (i.e., N(0,1) and N(0,0.5) random cluster effects) did not have a significant impact upon the bias of estimates, with only a slight increase in the magnitude of bias on CE5 (i.e., the effects with a higher variance) on MP1 and MP2.

Bias in estimates was minimal in the MP3 and MP4 conditions for both CE4 and CE5, indicating that the variability of the cluster effects had limited, if any, impact when the mixture proportion was constant. The variability of the cluster effect estimates was also small on MP1 and MP2 conditions (Figure 11). The magnitude of bias was very similar between CE4 and CE5 for the mis-specified model, while for the true model, the magnitude of bias was greater for CE5 than for CE4.

5.3 Precision of estimates

The mis-specified model condition consistently resulted in precision that was equal to or greater than that of the true model, as expected. The difference was smallest when the most cases were in the fast growth group (i.e., MP1 and cluster type 3), and was largest where the fewest cases were in the fast growth group (i.e., MP2 and cluster type 1). For both the mis-specified and true models, the precision of estimates increased as the effective sample size increased. As was described in Chapters 2 and 3, the mis-specified model always utilized 100% of the sample for the estimation of parameters, whereas the effective sample size was dependent on the mixture proportion for the true model. When combined with the previous set of results about bias, this simulation has shown that the mis-specified model consistently yields greater bias, with higher precision for those biased estimates, as compared to the true model.

5.4 Results on classification accuracy

Tables 17-21 show the classification accuracy of MLGMMs with five cluster effects ranked by quintile level. The cluster number did not have significant effects on classification accuracy as indicated by kappa, and kappa values tended to increase as cluster size increased, for both the true and mis-specified models. The true model performed much better on CE1 and moderately better on CE3, whereas the mis-specified model performed significantly better than the true model on CE4 and CE5 conditions. The CE2 condition led to virtually identical classification accuracy by both the true and mis-specified models.

Table 17

Rate of Misclassification at the Quintile Level: Cluster Effect 1 (CE1)

Simulation Condition				Kappa (95% CI)	
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	MLGMM	MLLGM
1	1	30	20	0.78 (0.76 ,0.80)*	0.64 (0.62 ,0.67)
			40	0.80 (0.78 ,0.82)*	0.63 (0.61 ,0.66)
		60	20	0.77 (0.75 ,0.79)*	0.65 (0.63 ,0.67)
			40	0.81 (0.79 ,0.82)*	0.64 (0.61 ,0.66)
		90	20	0.77 (0.75 ,0.79)*	0.63 (0.61 ,0.66)
			40	0.80 (0.78 ,0.82)*	0.64 (0.61 ,0.66)
	2	30	20	0.68 (0.66 ,0.70)*	0.53 (0.50 ,0.56)
			40	0.70 (0.68 ,0.72)*	0.56 (0.53 ,0.58)
		60	20	0.68 (0.66 ,0.70)*	0.53 (0.51 ,0.56)
			40	0.70 (0.68 ,0.72)*	0.55 (0.52 ,0.57)
		90	20	0.65 (0.63 ,0.68)*	0.54 (0.51 ,0.56)
			40	0.70 (0.68 ,0.72)*	0.54 (0.52 ,0.57)
	3	30	20	0.70 (0.68 ,0.72)*	0.58 (0.56 ,0.61)
			40	0.72 (0.70 ,0.74)*	0.61 (0.58 ,0.63)
		60	20	0.70 (0.67 ,0.72)*	0.57 (0.54 ,0.59)
			40	0.72 (0.70 ,0.74)*	0.61 (0.58 ,0.63)
		90	20	0.68 (0.66 ,0.70)*	0.58 (0.55 ,0.60)
			40	0.72 (0.70 ,0.74)*	0.60 (0.58 ,0.63)
	4	30	20	0.75 (0.73 ,0.77)*	0.61 (0.59 ,0.64)
			40	0.75 (0.73 ,0.77)*	0.62 (0.59 ,0.64)
		60	20	0.75 (0.73 ,0.77)*	0.62 (0.59 ,0.64)
			40	0.75 (0.73 ,0.77)*	0.62 (0.60 ,0.65)
		90	20	0.73 (0.71 ,0.75)*	0.61 (0.58 ,0.63)
			40	0.74 (0.72 ,0.76)*	0.62 (0.60 ,0.65)

* denote that Kappa for the group is significantly higher at $p<0.05$ level

Table 18

Rate of Misclassification at the Quintile Level: Cluster Effect 2 (CE2)

Simulation Condition				Kappa (95% CI)	
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	MLGMM	MLLGM
2	1	30	20	0.85 (0.82 ,0.87)	0.89 (0.87 ,0.91)
			40	0.95 (0.93 ,0.96)	0.93 (0.91 ,0.95)
		60	20	0.85 (0.82 ,0.87)	0.87 (0.85 ,0.90)
			40	0.93 (0.91 ,0.95)	0.94 (0.92 ,0.95)
		90	20	0.85 (0.83 ,0.87)	0.90 (0.88 ,0.92)*
			40	0.94 (0.92 ,0.95)	0.95 (0.94 ,0.97)
	2	30	20	0.73 (0.70 ,0.77)	0.76 (0.73 ,0.80)
			40	0.85 (0.83 ,0.87)	0.85 (0.83 ,0.88)
		60	20	0.75 (0.71 ,0.78)	0.75 (0.72 ,0.78)
			40	0.88 (0.86 ,0.90)	0.85 (0.82 ,0.88)
		90	20	0.75 (0.71 ,0.78)	0.76 (0.73 ,0.79)
			40	0.85 (0.82 ,0.87)	0.86 (0.83 ,0.88)
	3	30	20	0.77 (0.74 ,0.80)	0.75 (0.72 ,0.78)
			40	0.89 (0.86 ,0.91)	0.86 (0.83 ,0.88)
		60	20	0.80 (0.77 ,0.83)	0.75 (0.72 ,0.78)
			40	0.87 (0.84 ,0.89)	0.86 (0.83 ,0.88)
		90	20	0.76 (0.73 ,0.79)	0.77 (0.74 ,0.80)
			40	0.87 (0.85 ,0.89)	0.84 (0.82 ,0.87)
	4	30	20	0.85 (0.83 ,0.88)	0.86 (0.84 ,0.89)
			40	0.93 (0.91 ,0.95)	0.94 (0.93 ,0.96)
		60	20	0.86 (0.83 ,0.88)	0.86 (0.84 ,0.88)
			40	0.93 (0.92 ,0.95)	0.91 (0.89 ,0.93)
		90	20	0.87 (0.85 ,0.90)	0.83 (0.81 ,0.86)
			40	0.93 (0.91 ,0.94)	0.93 (0.91 ,0.94)

* denote that Kappa for the group is significantly higher at $p<0.05$ level

Table 19

Rate of Misclassification at the Quintile Level: Cluster Effect 3 (CE3)

Simulation Condition				Kappa (95% CI)	
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	MLGMM	MLLGM
3	1	30	20	0.82 (0.79 ,0.85)	0.82 (0.79 ,0.85)
			40	0.92 (0.90 ,0.94)	0.89 (0.87 ,0.91)
		60	20	0.84 (0.81 ,0.86)	0.81 (0.78 ,0.84)
			40	0.92 (0.90 ,0.93)*	0.86 (0.84 ,0.89)
		90	20	0.84 (0.82 ,0.87)	0.79 (0.76 ,0.82)
			40	0.91 (0.89 ,0.93)*	0.85 (0.82 ,0.87)
	2	30	20	0.71 (0.68 ,0.75)*	0.62 (0.58 ,0.66)
			40	0.76 (0.73 ,0.80)	0.71 (0.68 ,0.75)
		60	20	0.75 (0.72 ,0.79)*	0.62 (0.58 ,0.67)
			40	0.78 (0.75 ,0.81)	0.72 (0.69 ,0.76)
		90	20	0.69 (0.65 ,0.72)	0.64 (0.59 ,0.68)
			40	0.80 (0.77 ,0.83)*	0.70 (0.66 ,0.74)
	3	30	20	0.77 (0.74 ,0.80)	0.75 (0.72 ,0.78)
			40	0.87 (0.85 ,0.89)	0.84 (0.82 ,0.87)
		60	20	0.79 (0.76 ,0.82)	0.74 (0.70 ,0.77)
			40	0.86 (0.84 ,0.89)	0.86 (0.83 ,0.88)
		90	20	0.78 (0.75 ,0.81)	0.75 (0.71 ,0.78)
			40	0.89 (0.86 ,0.91)	0.84 (0.82 ,0.87)
	4	30	20	0.86 (0.83 ,0.88)	0.85 (0.83 ,0.88)
			40	0.93 (0.92 ,0.95)	0.94 (0.92 ,0.95)
		60	20	0.85 (0.83 ,0.87)	0.86 (0.83 ,0.88)
			40	0.94 (0.92 ,0.95)	0.94 (0.92 ,0.96)
		90	20	0.85 (0.82 ,0.87)	0.86 (0.84 ,0.89)
			40	0.95 (0.93 ,0.96)	0.94 (0.92 ,0.95)

* denote that Kappa for the group is significantly higher at $p<0.05$ level

Table 20

Rate of Misclassification at the Quintile Level: Cluster Effect 4 (CE4)

Simulation Condition				Kappa (95% CI)	
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	MLGMM	MLLGM
4	1	30	20	0.47 (0.40 ,0.53)	0.66 (0.61 ,0.71)*
			40	0.60 (0.55 ,0.66)	0.69 (0.65 ,0.74)
		60	20	0.47 (0.41 ,0.54)	0.69 (0.64 ,0.74)*
			40	0.57 (0.51 ,0.63)	0.71 (0.67 ,0.76)*
		90	20	0.49 (0.42 ,0.55)	0.67 (0.62 ,0.72)*
			40	0.61 (0.55 ,0.67)	0.74 (0.69 ,0.78)*
	2	30	20	0.44 (0.37 ,0.51)	0.52 (0.46 ,0.58)
			40	0.50 (0.43 ,0.56)	0.59 (0.53 ,0.65)
		60	20	0.49 (0.42 ,0.55)	0.50 (0.44 ,0.57)
			40	0.52 (0.46 ,0.59)	0.55 (0.49 ,0.60)
		90	20	0.42 (0.35 ,0.49)	0.56 (0.50 ,0.62)*
			40	0.53 (0.46 ,0.59)	0.64 (0.59 ,0.69)
	3	30	20	0.51 (0.44 ,0.58)	0.52 (0.46 ,0.58)
			40	0.53 (0.47 ,0.60)	0.63 (0.57 ,0.68)
		60	20	0.52 (0.46 ,0.58)	0.63 (0.57 ,0.68)
			40	0.69 (0.64 ,0.74)	0.68 (0.63 ,0.73)
		90	20	0.53 (0.46 ,0.59)	0.54 (0.47 ,0.60)
			40	0.61 (0.56 ,0.66)	0.72 (0.67 ,0.77)*
	4	30	20	0.59 (0.53 ,0.65)	0.70 (0.65 ,0.75)*
			40	0.68 (0.63 ,0.74)	0.79 (0.75 ,0.83)*
		60	20	0.57 (0.51 ,0.63)	0.68 (0.63 ,0.73)*
			40	0.66 (0.61 ,0.71)	0.75 (0.71 ,0.80)
		90	20	0.56 (0.50 ,0.63)	0.74 (0.69 ,0.78)*
			40	0.71 (0.66 ,0.76)	0.77 (0.72 ,0.81)

* denote that Kappa for the group is significantly higher at $p<0.05$ level

Table 21

Rate of Misclassification at the Quintile Level: Cluster Effect 5 (CE5)

Simulation Condition				Kappa (95% CI)	
Cluster Effect	Mixture Proportion	Cluster Number	Cluster Size	MLGMM	MLLGM
5	1	30	20	0.60 (0.54 ,0.66)	0.77 (0.72 ,0.81)*
			40	0.57 (0.51 ,0.63)	0.79 (0.75 ,0.83)*
		60	20	0.67 (0.62 ,0.72)	0.78 (0.74 ,0.82)*
			40	0.72 (0.67 ,0.77)	0.79 (0.75 ,0.83)
		90	20	0.64 (0.58 ,0.69)	0.78 (0.74 ,0.82)*
			40	0.59 (0.53 ,0.65)	0.80 (0.76 ,0.83)*
	2	30	20	0.41 (0.34 ,0.48)	0.68 (0.63 ,0.73)*
			40	0.41 (0.34 ,0.48)	0.70 (0.66 ,0.75)*
		60	20	0.41 (0.34 ,0.48)	0.70 (0.66 ,0.75)*
			40	0.48 (0.41 ,0.55)	0.74 (0.70 ,0.79)*
		90	20	0.46 (0.39 ,0.52)	0.73 (0.68 ,0.77)*
			40	0.42 (0.35 ,0.48)	0.75 (0.70 ,0.79)*
	3	30	20	0.56 (0.50 ,0.62)	0.70 (0.66 ,0.75)*
			40	0.74 (0.70 ,0.79)	0.81 (0.78 ,0.85)
		60	20	0.58 (0.52 ,0.64)	0.78 (0.74 ,0.82)*
			40	0.76 (0.72 ,0.81)	0.80 (0.77 ,0.84)
		90	20	0.68 (0.63 ,0.73)	0.77 (0.73 ,0.81)
			40	0.70 (0.65 ,0.75)	0.82 (0.78 ,0.86)*
	4	30	20	0.68 (0.63 ,0.73)	0.82 (0.79 ,0.86)*
			40	0.75 (0.71 ,0.80)	0.86 (0.83 ,0.89)*
		60	20	0.75 (0.71 ,0.80)	0.80 (0.76 ,0.84)
			40	0.66 (0.61 ,0.72)	0.88 (0.85 ,0.90)*
		90	20	0.72 (0.67 ,0.77)	0.88 (0.84 ,0.91)*
			40	0.58 (0.51 ,0.64)	0.89 (0.86 ,0.91)*

* denote that Kappa for the group is significantly higher at $p<0.05$ level

Table 22 shows the average kappa obtained over the cluster effects. The average values reflect the finding above. The classification rate by the true model, but not by the mis-specified model, was affected by the random cluster effects (i.e., CE4 and 5).

Table 22

Average Kappa by Cluster Effect

	Average Kappa	
Cluster Effect	MLGMM	MLLGM
1	0.73	0.60
2	0.85	0.85
3	0.84	0.80
4	0.55	0.65
5	0.61	0.78

The results on the classification accuracy agreed with the evaluation of bias in the cluster effects, where lower bias and higher precision lead to the higher classification accuracy. The true model had the highest classification rate at CE1 condition where it had a minimal bias, but the classification rate suffered on CE4 and 5 where the bias were higher than the mis-specified model and the precision was lower.

Chapter 6: Discussion

The pilot study described in Chapter 4 tested the code that was used in the main study, identifying convergence issues and fidelity of the programs and the code coordinating these programs to simulate, fit models to, and analyze 36,000 individual runs of the three models that represent 100 replications for each combination of conditions shown in Table 1. The results of these models were presented in Chapter 5. Since model convergence was perfect on 1- and 2-class MLGMM (1-class MLGMM is equivalent to MLLGM), and a very low incidence of convergence problem occurred with 3-class MLGMM (32 issues in these 12,000 runs), this chapter discusses how the results in Chapter 5 address the main research questions that motivated this study.

The goal of this study was to investigate the impact on the teacher's (or cluster) effect estimates that might arise from having different proportions of students in two growth groups within a single classroom, as described in the example in Chapter 1. Fairness in evaluation could not be established if there was systematic bias in estimates of any teacher's effect or effectiveness. This simulation study manipulated a variety of conditions in order to investigate the magnitude of bias in teacher's effect estimates resulting from heterogeneous student growth within a class. In particular, the research questions were:

1) What information criteria can be used to identify the true number of latent class variable levels in the MLGMM context?

2) Are the level-2 parameter estimates in the multilevel model affected (in terms of bias, and precision) by incorrectly-modeled level 1 effects?

The brief summary of the findings presented in Chapter 5, and exemplified in Tables 11-22, Figures 9-12, and Appendix A is that:

  a.   AIC3, BICB, and SABIC performed well to identify MLGMM with the correct number of latent classes, although AIC and AICc were the only information criteria to perform well with smaller sample size. BIC performed poorly, contrary to previous research findings. BIC over-penalized model because a total sample size did not reflect the true sample size of data.

  b.   Model misspecification leads to systematic bias in level-2 parameter estimates in multi-level models, especially when there is more variability in some classroom (represented by mixture proportions). This bias is attenuated when the proportion of students belonging to a high-growth group is equal to, or greater than, that of the slow growth (e.g., PLP) group. However, when MLGMM is used instead of simple MLLGM for the level-2 parameter estimates; the bias is greatly reduced, loses all systematicity, and appears unaffected by any of the other features that were manipulated in the simulation.

  c.   Bias in estimation of teacher effects was significantly reduced by accounting for the student level heterogeneity in most simulation conditions, except for a few conditions described later in this discussion (Section 6.3.3).

  d.   Precision of the estimated teacher effects was affected systematically by each of the conditions under study in this project.

Effects of the various conditions on precision tended to vary depending

on the proportion of students in the fast growth group, for all sample

sizes, underscoring a specific effect that unmodeled heterogeneity in the

classroom can have on the estimation of teacher effects.

Taken together, these results suggest that the evaluation of teachers, in terms of

their effects/effectiveness, using VAM, can proceed fairly across a wide spectrum of

contexts (and school, class or district sizes) – *but only if bias can be controlled* as

discussed below. In fact, the potential for controlling bias is the most important feature of

MLGMM, specifically because high levels of precision for biased estimates could lead to

greater (misplaced) confidence in such incorrect estimates.

More worrisome is the pattern of bias in the results for better (positive cluster

effect estimates) and worse (negative cluster effect) teaching. Figures 9-12 show that if a

cluster effect is positive, then the bias tends to be positive (overestimation), and that the

greater the absolute value of this cluster effect, the greater the bias. Increasingly better

teachers will appear even better due to the bias and overestimation of positive cluster

(teacher) effects. Figures 9-12 shows that, if a cluster effect is negative, then the bias

tends to be negative, representing overestimation of a *negative* effect of the teacher.

Similar to the overestimation of positive cluster effects, the overestimation (bias) of

negative effects also increases with the absolute value of the cluster effect estimate. Thus,

increasingly worse teachers will appear even worse due to this bias and overestimation of

negative teacher effects.

In the following sections, the results presented in Chapter 5 are discussed with

respect to their contributions to these conclusions and the future steps suggested by the

results and their implications for the effective and fair application of VAM, using

MLGMM, for policy making and teacher evaluations.

In the following sections, the results presented in Chapter 5 are discussed with

respect to their contributions to these conclusions and the future steps suggested by the

results and their implications for the effective and fair application of VAM, using

MLGMM, for policy making and teacher evaluations. As described in Chapter 2, the PLP

students identified by Lazarus et al. (2010) may well describe the slow growth group

simulated in this study. If so, then it highlights the importance of accounting for the

presence of PLP students in the estimation of teacher effects. However, the exclusion of a

subgroup of students (i.e., PLP) from the estimation of a teacher's effect assumes a great

deal of the "truth" of the statistical identification of such a class of students. Including

some demographic indicators (observed variables), for example, those identified by

Lazarus et al. (2010), in the model as covariates may help to reduce the impact of PLP

students without assuming that the latent classes inferred from the data have identified

this class correctly, although as Palardy and Vermunt (2009) indicated, including

covariates in VAM analyses can make the identification of latent classes more difficult.

As noted in the introduction, the collection and analysis of any data must be driven by

clear statements of the assumptions being made and the sources of variability to be

modeled; users of the VAM approach must justify whether observed variables – proxies

for potentially important latent classes – are closer to the "truth" than the latent variables

might be.

6.1 Model convergence of the multilevel growth mixture model

The convergence rates of MLGMMs were much better than expected, resulting in very few issues (32/12,000, all within a small fraction of the 12,000 model estimations). Convergence issues only occurred for the over-estimating condition (i.e., MLGMM with more latent classes than the true model). Combining multilevel structure with GMM did not seem to affect the model convergence, as was shown in Table 10. Problems occurred most often when the cluster size was small (i.e., CS=20) and had fewer cases in the high growth, relative to the low growth, group (i.e., MP2). The effect of cluster number was minimal and consistent across all conditions involving MLGMMs. However, due to modeling constraints for MLGMMs, the cluster size was held constant across all clusters, and while this inflated the rate of model convergence, equal cluster size is not realistic. Therefore, this work supports the combination of MLM and GMM approaches, but future work should proceed with more realistic data (with different sized clusters), which might have an impact on the model convergence and interpretability.

6.2 Information criteria performance

Six information criteria were used to identify the model that fit the data best among the 1-, 2-, and 3-class MLGMMs (recall that the 1-class MLGMM is MLLGM). BIC performed quite poorly, but not uniformly worst, despite performing extremely well in previous research (e.g., Chen et al., 2010; Muthén & Asparouhov, 2009; Nylund, Asparouhov, & Muthén, 2007). Palardy and Vermunt (2010) proposed BICB, which utilizes the number of clusters instead of overall sample size (as BIC does) for the sample size adjustment factor. It is possible that BIC was hampered in the multilevel modeling conditions of this simulation because the sample size penalty factor was too severe (i.e.,

selecting the simpler model too often). BIC only worked well when the cluster effect was random (i.e., CE4 or 5) with a large sample size; with random cluster effects and small samples together, it performed poorly and inconsistently (see Tables 15 and 16). BICB and SABIC both worked well, especially when the cluster size (i.e., CS=40) and cluster number (i.e., CN=60 or 90) were larger. The sample size penalty adjustment of SABIC, as compared to the over-penalization of BIC, led to superior and more consistent performance of SABIC. However, BICB generally outperformed SABIC on almost all simulation settings, and especially with smaller sample sizes. Overall, the cluster number seems to have been a good proxy for the sample size in MLGMM, and in Table 12 this is most obvious. However, both BICB and SABIC still tended to over-penalize models in conditions with both smaller cluster size (i.e., 20) and cluster number (i.e., 40 or 60).

AIC and AICc performed similarly, correctly identifying models in conditions with smaller cluster number and sizes, whereas BIC did not function well in these conditions. The correct identification rates of AIC and AICc were not as good as BICB or SABIC as the overall sample size increased from 600 to 3,600, but all four of these criteria performed fairly consistently across all conditions. AIC3 had a model identification profile very similar to BICB and SABIC, with marginally better identification rates than these two criteria in conditions with smaller overall sample size, but significantly worse identification than AIC and AICc in these same conditions, An interesting observation was that AIC3 performed slightly worse than AIC and AICc in conditions with larger overall sample sizes, defined by cluster number and size (i.e., sample size = cluster number $\times$ cluster size).

The six information criteria each performed differently across conditions; none of them was consistently best (or worst). AIC and AICc performed well with smaller sample sizes, AIC3 performed well in mixture proportions conditions 1 and 2 with a smaller sample sizes, and BICB and SABIC performed well in mixture conditions 3 and 4 with larger sample size. The pattern of results for BIC was similar to those of BICB and SABIC, but at a lower level; therefore BIC may have a very limited use in the model identification in MLGMM.

Within both the pilot and main studies, BIC did not perform well for most conditions, which is surprising given its applicability to simulation studies (i.e., it works only when the true model is known to be among those in question) and its excellent performance in other work (e.g., Muthén & Asparouhov, 2009). BIC has tendency to prefer a simple model and only performs well where the cluster size is large (i.e., 40) and the mixture proportion is consistent (i.e., MP = 3 or 4).

Based on these results, given the motivation for the conditions that were included in the simulation (i.e., emphasizing the accuracy of estimation of the teacher's effect or value-added after taking account the growth profiles, clusters and cluster sizes), the best model selection performance will be obtained by combining either AIC or AICc – which were indistinguishable in these results – together with either AIC3 or BICB. AIC and AICc are recommended for cases involving smaller sample sizes, roughly less than 1,200 total cases, and AIC3 and BICB are recommended in cases with a large sample size (1,200 cases or more). Contrary to previous findings, these findings suggest that BIC should not be used for MLGMM model identification.

6.3 Evaluation of systematic biases across the simulation condition

This section is divided to discuss cluster effects (CE=1, 2 and 3, and 4 and 5) as they were presented in Chapter 5. The purpose of this study was to identify potential systematic biases introduced by model misspecification, which should always be considered in any modeling enterprise and which may negatively influence the otherwise fair comparison of parameters. As noted in Chapter 1, the simulation was set up to address this question by determining whether level-2 parameter estimates in a multilevel model were or could be affected (in terms of bias and precision) by incorrectly-modeled level-1 effects (i.e., by ignoring the heterogeneity in the student population). A systematic manipulation of the cluster effects and the mixture proportion described in Chapter 2, one way to estimate these effects, was employed in the simulation design, and the question was addressed by examining the patterns of bias on the fixed cluster effects introduced through the cluster effect conditions 1 through 3.

6.3.1   Cluster effect condition 1 (CE=1)

Cluster effect condition 1 had the same five true effects (i.e., -1, -0.5, 0, 0.5, 1) across three cluster types in the mixture proportion conditions, and this was designed to investigate how the bias on each true effect behaved when mixture proportions vary (see Chapter 3).

As seen in Figures 9 and 10, model misspecification led to greater, and systematic, bias, as compared to conditions involving the true model. If a cluster effect is negative, then the bias tends to be negative, representing overestimation of a *negative* effect of the teacher. If a cluster effect is positive, then the bias tends to be positive

107

(overestimation), and that the greater the absolute value of this cluster effect, the greater the bias.

In addition to varying with the sign of the true effect (TE), bias was also greater in magnitude for greater TE –whether positive or negative. (i.e., bias increased in absolute value as TE moved away from zero). In conditions with increased overall variability in the mixture proportion (i.e., MP1 and 2) the magnitude of bias increased substantially. The within-sample variability defined by cluster type exhibited the same pattern, namely, that magnitude of bias increased with mixture proportion variability. Greater bias was observed for conditions with greater variability in mixture proportion (MP2) as compared to conditions with less variability in this proportion (MP1).

In addition to these systematic effects of mixture proportion, true effect of cluster, or their combination on the magnitude and sign of the bias on the cluster effect estimates, there were also effects on bias coming from model misspecification. Significant positive bias was observed in conditions with zero TE when the model was mis-specified, but not for the true model in this condition. The most pronounced positive bias at zero TE occurred for MP1 and MP4 conditions – suggesting sensitivity to a higher proportion of fast growth cases within any cluster. This means that student heterogeneity contributes to the inflation of the cluster effect estimates. The only exception was at cluster type 3 on MP1, in which all of cases were in the fast growth group (i.e., 1 class); bias in estimates was negligible for both mis-specified and true models in this condition.

These findings suggest that the potential for bias in estimating a teacher's effect (represented by cluster effect estimates) is greatest in the following conditions:

1. Higher overall variation, between cluster type, in terms of in mixture proportion (MP=2).

2. Smaller proportion of cases in the fast growth group (cluster type 1 in MP2 and 3).

3. Higher overall variation in the mixture proportion (MP3).

The effect of model misspecification on the bias in estimating cluster effect is similar; namely, the magnitude of bias with zero TE was influenced by the overall relative proportions of fast and slow growth groups, especially for the mis-specified model. Unlike in the true model conditions, in the misspecification condition, all cases in the slow growth group had zero TEs and these mis-specified models included all cases in their estimation, whereas the true models attempted to separate the fast and slow growth groups during estimation.

The true model reduced bias much better than the mis-specified model, implying that the effects of model misspecification would be greatest in a school district having schools with a wide range of performances and/or classes within a school encompassing a wide performance range. For instance, the highest magnitude of bias would occur in a classroom with the smallest proportion of fast growth students within a school that also has a small proportion of fast growth students. This might represent a heterogeneous urban school district. All teachers, good or bad, would be most affected if evaluated in the context of schools with few fast growers, but bad teachers would be more negatively affected (i.e., further lowering the negative cluster effect estimates) than good teachers being evaluated in schools with many fast growers. This situation would likely occur within more heterogeneous urban school districts. The evaluation of teachers in more

homogeneous suburban school districts where majority of students belonged to the fast

growth group, and where most schools in the district maintain a high achievement level,

would yield the least biased estimates of teacher effect, even with a mis-specified VAM.

6.3.2   Cluster Effect Conditions 2 and 3 (CE=2 and 3)

Cluster effect conditions 2 and 3 were designed to evaluate the potential

systematic biases comparing the same cluster effects in the different mixture proportions.

CE2 and CE3 had the same overall cluster effects (i.e., -1, -0.5, 0, 0.5, and 1) but the

assignment of effects was reversed between cluster types 1 and 3, while cluster type 2

had zero effects on both CE conditions. The cluster types were defined based on the

mixture proportion. Therefore, no differences in bias were observed comparing CE2 and

CE3 for mixture proportion conditions MP3 and MP4.

The positive TEs were more strongly associated with differential mixture

proportions, in terms of bias resulting from conditions CE2 and CE3 with both MP1 and

MP2. The positive bias was most pronounced when the fast growth group had the

smallest proportion. The negative bias on the negative TEs was less pronounced when the

overall variation in mixture proportion was low (MP1). The magnitude of this negative

bias was fairly low in estimates derived from both true and mis-specified models.

However, this was not observed when the variation in mixture proportion was high

(MP2), as outline in the preceding section.

There was significant negative bias on the negative TEs regardless of the

proportion of cases in the fast growth group (compare TE -1 and -0.5 in MP2 between

CE2 and CE3 in Figure 11). The inclusion of lower proportions of the fast growth group

greatly increased positive bias on the positive TEs, especially for MP2. Comparison of

MP3 and MP4 profiles in Figure 11 confirms that the lower proportion of cases in the fast growth group was the source of this bias.

The pattern of bias on the cluster effect estimates across conditions was informative:

1. Low overall variation in growth profile (MP1) conferred the least penalty on the cluster effect estimates (i.e., a low magnitude of decrease in the cluster effect estimates) when combined with negative TEs, regardless of the actual proportion of cases in the fast (or either) growth group (cluster type 1 & 3), but the bias on positive TEs was greatly increased (i.e., higher magnitude of parameter change than for negative TE) when the proportion of fast growth cases was low (MP1, CE3, and cluster type 1).

2. High overall variation in growth group proportions (MP2) led to strong negative bias for negative TEs (i.e., a higher magnitude of decrease in the cluster effect estimates), compare to MP1. The positive bias on the cluster effect estimates for positive TEs on MP2 had the same pattern as MP1.

3. Overall bias was greatest in MP2 and CE3 conditions, where the variation in the sample and among cluster types were the highest, exaggerating overestimation of a negative effect of poor teachers and overestimation of positive effect for good teachers. Greater variation in the mixture proportion increased bias.

4. The potential for unfairness, if VAM without accounting for student-level heterogeneity is employed to estimate teacher effect, is very high due to the tendency for increasing student-level heterogeneity to lead to overestimation

of positive effects for good teachers and overestimation of a negative effect of poor teachers.

The true models yielded estimates that were less biased than those of the mis-specified models, but the magnitudes of bias were greater for CE2 and 3 conditions than were those from CE1 conditions. If the TE*s* from conditions similar to CE2 and CE3 existed when a given teacher was being evaluated using VAM (by MLGMM), then there is a strong likelihood of significantly overestimating teachers' positive effects, while overestimation of teachers' negative effect to a lesser degree. That is, in a district with a wide range of students in terms of growth profiles, including low-starting, fast growth students in low performing schools and high-starting, fast growth students in high performing schools, or in a single school with these characteristics in the classrooms. These mixtures will create bias in evaluation that heavily favors, and also inflates the effects of good performing teachers. They do much better, in a sense, at identifying poor performing teacher by overestimating the negative effects of poorer performing teachers. Differential bias that depends on the actual capability of teachers cannot provide fair evaluations. In cases where a class has fewer students in the fast growth group, the VAM approach will strongly favor teachers with a positive effect and will severely penalize those teachers with negative effects.

In addition to having significant implications for the fairness of decision-making and policy based on VAM results, these results can also affect the choices that teachers make – they might feel that schools with higher proportions of fast growing students are the only contexts in which they have a chance of being evaluated fairly. The issue of fairness – and its perception – in evaluation affects all parties in these decisions.

112

6.3.3   Cluster Effect Condition 4 and 5 (CE=4 and 5)

Cluster effect condition 4 and 5 were included to assess the impact of overall variability in the sample on the cluster effect estimates. As described in Chapter 3, the cluster effects were randomly generated from the normal distribution with the mean of 0 and a different variance (i.e., 0.5 for CE4 and 1.0 for CE5).

The bias in the cluster effect estimates was quite limited for CE4 and 5 conditions, and this was observed for both the true and mis-specified models. The effects of MP, CS, and CN were also very limited in these conditions. By contrast, increased variance on the TE was the only manipulated feature that actually increased the bias derived from the true model in a meaningful way.

Interestingly, the BIC criterion only functioned as expected for CE4 and 5 conditions when the sample size was large (e.g., CN=90 and CS=40) and on MP3 or 4 (i.e., constant mixture proportion). A constant mixture proportion was used in the reviewed research by Nylund et al.,(2007), Muthén and Asparouhov (2009), and Chen et al. (2010). This might account for the strong performance of BIC reported in those analyses, and explain why BIC performed so poorly in virtually all of the analyses reported here (as discussed above in Section 6.2).

6.3.4   Precision of estimates

As expected, the precision of estimates was greater whenever the number of cases in the condition was higher. This finding is important for the teacher evaluation example described in Chapter 1 because of the potential for this same higher precision to also arise when the estimate is *biased*. The estimates from the mis-specified models tended to be biased, substantially in some cases (see Figures 9-11), which only serves to compound

113

the problems associated with the mis-specified model. Additionally, because the

precision for mis-specified models tends to be improved by larger sample sizes, just as

that of true models, if the model misspecification is undetected, it will give erroneous

confidence in the biased estimates.

However, if the cluster effects (e.g., teacher's effect) within a cluster unit (e.g.,

school district) are similar to CE4 and CE5 (i.e., normally distributed around zero) with

equal proportions of students in the growth profiles, then model misspecification is

performs equally well or better than the true model. As discussed in Section 6.5 below,

the zero cluster effects assigned for the slow growth group could have reduced the bias in

the cluster effect estimates, particularly for the mis-specified model. The samples in CE4

and CE5 had greater numbers of cases with a cluster effect of zero, which acts to further

reduce the variance of overall cluster effects in these conditions. This combination of

variance "shrinkage" effects explains the reduction in bias.

6.4 Results on classification accuracy

Classification accuracy at the quintile level was included as an alternative

measure of bias and precision of cluster effects because it takes both bias and precision of

cluster effect estimates into account (analyzed by kappa) to summarize model

performance. This study found that the true and mis-specified models each performed

better in specific simulation conditions; the true model outperformed the mis-specified

model (in terms of bias and precision) in CE1 conditions, but the true model performed

poorly in terms of bias on CE4 and 5 conditions, most likely due to the condition where

the mean of TEs for the fast growth group and the effect of slow growth group were both

zero. The classification method would be particularly useful when the evaluating the

114

effect of a certain criterion, such as the threshold or cutoff to be proficient or non-proficient, which was the beyond the scope of this research.

6.5 Limitations of the research

This study was designed to address specific questions as outlined earlier. Simulation projects require fixed characteristics, and as such, these led to several limitations. One such limitation is the use of only two growth profiles. This might be more realistic than assuming homogeneous growth within a cluster, but it is far more likely that there are more than just two growth profiles in any classroom or school. A related challenge was that no latent classes were included to represent the cluster level (e.g., between-level or teacher's level) where interactions between individuals and teachers are very likely. Further, some mixture proportions were unrealistic (i.e., MP3 and 4) because they represent homogeneous growth within clusters; these conditions were needed in order to contextualize these results with those published previously. The mixture proportions used for MP1 and 2 might not reflect reality either, but they do represent the assumption that there is variation in these growth class proportions (i.e., proportions of student in each growth profile) within a given cluster, and that this variation is unlikely to be consistent across all clusters in a given modeling situation. The results do suggest that variation in those proportions has a significant impact on estimation and thereby, on decision-making that might be based on those teacher effect estimates. Future studies could explore whether a wider, more realistic, range of variation in growth class proportions yields a clearer picture of this impact and possible ways of addressing it in simulations.

The impact of higher proportions of slow growth group members, which had zero cluster effect, was especially apparent in simulation conditions CE4 and 5 and was not expected. The use of positive mean true random effects (i.e., N(1,1)) or a negative non-zero mean true effect to represent the slow growth group could potentially alleviate the issue encountered for CE4 and 5 (see section 6.3.3), and could more clearly demonstrate the differences in inferences that are supported by the true and mis-specified model conditions. An option for realizing these features, while not causing the issues described, is to center the true effect of the fast growth group for CE1 through CE3 on a positive value (e.g., 1) instead of zero, which was the value used in this study. In spite of these limitations, the conclusions outlined at the start of the chapter support a general argument about the impact of using MLGMM over MLLGM. Coupled with the results and lessons learned from the zero cluster effect characteristics, this research could be a useful guide for further investigations, as well as applications, utilizing MLGMM.

An issue of inferential robustness, when alternative models with similar model fit (i.e., information criteria select alternative models instead of true model) lead to a different interpretations or conclusions, has not fully addressed in this dissertation. The simulation conditions of this study were designed to minimize the influence of uncontrolled effects to avoid this issue. However in more complex real life data, it is extremely important to carefully investigate alternative models in order to make a valid interpretation of results.

Finally, although this study was designed to investigate the impact of unmodeled heterogeneity at the classroom level on the potential for fair VAM-derived teacher evaluations, the greatest challenges to fair decision-making that is based on teacher

116

effects (or value-added effect by teachers) is not the actual values of these estimates, but rather, it is the distinction between proficient and not-proficient teachers – a two-level classification. The simulation, and therefore, results, do not speak to that two-level situation, but the finding that teachers with more positive and more negative cluster effects will actually generate differentially biased estimates suggests that any proficient/non-proficient classification will require very careful attention to the "non-proficient" characterization. Further, the estimation of changes in teacher effects would be critical, because these results suggest that "improvement" in teacher effect would be more easily recognizable in better teachers and would be more difficult to recognize in those who may need, or indeed may be struggling, to improve the most.

6.6 Future directions

This research was limited in scope but it achieved the stated goal of providing evidence supporting the use, and interpretability, of MLGMM as a tool to control bias and improve fairness in the evaluation of teacher effects in value-added modeling contexts. In addition to different approaches to address the limitations outlined above, future work in this domain should test the effect of <u>unequal</u> cluster sizes in the estimation and identification of MLGMM. Unequal cluster sizes, together with a more realistic variety in the latent growth profiles, have not been studied and would represent a greater range of real-world conditions in which MLGMM should be tested. The introduction of the between-level latent classes to incorporate, or explore, interaction between the between-level (e.g., teachers) and the within-level (e.g., students) latent classes might be useful projects, depending on the type of evaluations that are of interest (and on the emphasis on potential sources of the value that is believed to have been added in

decision-making). Studying the effects of different growth profile parameters, including the shape and rate of growth and the number of growth profiles, could also strengthen the estimation, applicability, and interpretability of cluster effects that are estimated with MLGMM. At some point, estimation of change in teacher effect will become a very important topic, possibly supporting the proficient/not-proficient classification based on VAM estimates, as long as the bias is controlled and is no longer differential depending on whether the teacher is stronger or weaker. In sum, this study supports the continued exploration of MLGMM for fair decision-making in educational contexts.

Appendix A

Estimation Results for All Simulation Conditions

Appendix A.1: Bias and Error of group estimates: Mixture Proportion 1 (MP1) and Cluster Effect 1 (CE1) for the true model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.29	-0.56	0.08	0.21	-0.19	-0.56	0.15	0.2	-0.12	-0.46	0.18	0.2
		-0.5	-0.12	-0.56	0.18	0.23	-0.07	-0.42	0.24	0.19	-0.09	-0.42	0.19	0.19
		0	0.14	-0.76	0.91	0.51	0.07	-0.84	1	0.54	-0.05	-0.37	0.25	0.19
		0.5	0.15	-0.19	0.55	0.21	0.06	-0.22	0.34	0.18	-0.01	-0.28	0.25	0.18
		1	0.28	-0.06	0.63	0.22	0.17	-0.17	0.5	0.21	0.05	-0.21	0.29	0.16
	40	-1	-0.23	-0.48	-0.03	0.15	-0.17	-0.41	0.09	0.15	-0.13	-0.31	0.1	0.12
		-0.5	-0.1	-0.34	0.14	0.14	-0.08	-0.36	0.16	0.15	-0.08	-0.32	0.14	0.14
		0	-0.1	-0.9	0.89	0.56	0.15	-0.89	1.03	0.64	-0.04	-0.28	0.18	0.14
		0.5	0.17	-0.1	0.36	0.14	0.02	-0.18	0.27	0.14	-0.01	-0.24	0.22	0.15
		1	0.28	-0.01	0.63	0.19	0.11	-0.15	0.38	0.15	0.01	-0.25	0.27	0.14
60	20	-1	-0.28	-0.64	0.01	0.22	-0.2	-0.5	0.1	0.19	-0.17	-0.51	0.13	0.19
		-0.5	-0.13	-0.43	0.22	0.2	-0.14	-0.45	0.2	0.2	-0.09	-0.41	0.22	0.19
		0	0.04	-0.92	0.9	0.55	0.2	-0.57	0.91	0.49	-0.03	-0.33	0.24	0.18
		0.5	0.14	-0.2	0.49	0.22	0.04	-0.29	0.37	0.2	0	-0.33	0.29	0.2
		1	0.25	-0.04	0.59	0.21	0.11	-0.23	0.43	0.19	0.04	-0.27	0.36	0.18
	40	-1	-0.22	-0.48	0.09	0.16	-0.16	-0.44	0.07	0.14	-0.09	-0.34	0.14	0.14
		-0.5	-0.1	-0.33	0.16	0.15	-0.08	-0.25	0.13	0.11	-0.1	-0.32	0.08	0.13
		0	-0.12	-0.94	0.97	0.59	0.14	-0.9	0.93	0.58	-0.06	-0.27	0.13	0.13
		0.5	0.15	-0.12	0.4	0.15	0.03	-0.23	0.3	0.15	-0.02	-0.22	0.16	0.12
		1	0.25	-0.11	0.57	0.2	0.07	-0.19	0.33	0.15	0.01	-0.19	0.26	0.14
90	20	-1	-0.26	-0.65	0.13	0.24	-0.17	-0.5	0.18	0.2	-0.14	-0.45	0.14	0.18
		-0.5	-0.16	-0.59	0.2	0.23	-0.09	-0.34	0.23	0.18	-0.09	-0.4	0.14	0.18
		0	0.1	-0.93	0.95	0.6	0.29	-0.54	1.02	0.51	-0.1	-0.37	0.2	0.2
		0.5	0.16	-0.18	0.53	0.23	0.07	-0.21	0.37	0.18	0.02	-0.31	0.27	0.18
		1	0.25	-0.14	0.59	0.23	0.11	-0.23	0.41	0.2	0.04	-0.25	0.29	0.18
	40	-1	-0.27	-0.5	0	0.15	-0.15	-0.46	0.16	0.16	-0.11	-0.34	0.11	0.14
		-0.5	-0.08	-0.36	0.18	0.16	-0.09	-0.29	0.18	0.14	-0.08	-0.28	0.1	0.13
		0	-0.14	-0.88	0.74	0.52	0.08	-1.03	0.99	0.64	-0.06	-0.29	0.16	0.13
		0.5	0.17	-0.12	0.42	0.17	0.01	-0.24	0.27	0.15	-0.03	-0.31	0.18	0.15
		1	0.28	-0.01	0.56	0.19	0.08	-0.17	0.33	0.14	-0.02	-0.23	0.19	0.12

Appendix A.2: Bias and Error of group estimates: Mixture Proportion 2 (MP2) and Cluster Effect 1 (CE1) for the true model

			Cluster Type									
			1				2				3		
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.26	-0.65	0.13	0.24	-0.17	-0.5	0.18	0.2	-0.14	-0.45	0.14	0.18
		-0.5	-0.16	-0.59	0.2	0.23	-0.09	-0.34	0.23	0.18	-0.09	-0.4	0.14	0.18
		0	0.1	-0.93	0.95	0.6	0.29	-0.54	1.02	0.51	-0.1	-0.37	0.2	0.2
		0.5	0.16	-0.18	0.53	0.23	0.07	-0.21	0.37	0.18	0.02	-0.31	0.27	0.18
		1	0.25	-0.14	0.59	0.23	0.11	-0.23	0.41	0.2	0.04	-0.25	0.29	0.18
	40	-1	-0.27	-0.5	0	0.15	-0.15	-0.46	0.16	0.16	-0.11	-0.34	0.11	0.14
		-0.5	-0.08	-0.36	0.18	0.16	-0.09	-0.29	0.18	0.14	-0.08	-0.28	0.1	0.13
		0	-0.14	-0.88	0.74	0.52	0.08	-1.03	0.99	0.64	-0.06	-0.29	0.16	0.13
		0.5	0.17	-0.12	0.42	0.17	0.01	-0.24	0.27	0.15	-0.03	-0.31	0.18	0.15
		1	0.28	-0.01	0.56	0.19	0.08	-0.17	0.33	0.14	-0.02	-0.23	0.19	0.12
60	20	-1	-0.46	-0.83	-0.07	0.23	-0.31	-0.72	0.07	0.25	-0.22	-0.58	0.13	0.22
		-0.5	-0.23	-0.6	0.12	0.22	-0.17	-0.55	0.2	0.23	-0.17	-0.42	0.1	0.16
		0	0.12	-0.54	0.74	0.4	0.12	-0.91	0.94	0.5	0.18	-0.5	0.91	0.43
		0.5	0.31	-0.13	0.71	0.24	0.1	-0.31	0.39	0.23	0.03	-0.27	0.33	0.19
		1	0.53	0.03	0.93	0.28	0.29	-0.04	0.71	0.24	0.11	-0.19	0.46	0.2
	40	-1	-0.36	-0.68	-0.01	0.2	-0.27	-0.54	0.01	0.16	-0.2	-0.47	0.11	0.17
		-0.5	-0.17	-0.51	0.1	0.18	-0.13	-0.42	0.11	0.16	-0.14	-0.37	0.14	0.16
		0	0.2	-0.59	0.94	0.46	0.19	-0.76	0.84	0.52	0.2	-0.58	1	0.5
		0.5	0.28	-0.06	0.54	0.18	0.09	-0.23	0.38	0.2	-0.01	-0.26	0.24	0.14
		1	0.5	0.05	0.91	0.25	0.21	-0.13	0.48	0.18	0.05	-0.19	0.28	0.16
90	20	-1	-0.43	-0.79	0	0.24	-0.33	-0.65	0.22	0.24	-0.25	-0.58	0.03	0.2
		-0.5	-0.18	-0.51	0.2	0.23	-0.16	-0.53	0.23	0.22	-0.17	-0.49	0.21	0.21
		0	0.21	-0.49	1	0.43	0.21	-0.67	0.82	0.44	0.16	-0.38	0.89	0.42
		0.5	0.29	-0.14	0.71	0.26	0.14	-0.22	0.58	0.22	0.05	-0.22	0.34	0.17
		1	0.45	0	0.78	0.25	0.24	-0.06	0.56	0.19	0.12	-0.19	0.42	0.2
	40	-1	-0.4	-0.68	-0.04	0.2	-0.27	-0.61	0.01	0.19	-0.2	-0.47	0.07	0.16
		-0.5	-0.18	-0.41	0.13	0.17	-0.14	-0.39	0.16	0.17	-0.13	-0.39	0.14	0.17
		0	0.16	-0.69	0.8	0.47	0.22	-0.85	0.97	0.54	0.19	-0.48	0.83	0.44
		0.5	0.29	0.01	0.59	0.18	0.1	-0.17	0.39	0.17	-0.03	-0.3	0.22	0.16
		1	0.46	0.13	0.81	0.21	0.17	-0.14	0.5	0.19	0.03	-0.24	0.34	0.17

Appendix A.3: Bias and Error of group estimates: Mixture Proportion 3 (MP3) and Cluster Effect 1 (CE1) for the true model

Cluster Number	Cluster Size	Cluster Effect	Cluster Type 1 Bias			Cluster Type 1 Error	Cluster Type 2 Bias			Cluster Type 2 Error	Cluster Type 3 Bias			Cluster Type 3 Error
			Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.43	-0.79	0	0.24	-0.33	-0.65	0.22	0.24	-0.25	-0.58	0.03	0.2
		-0.5	-0.18	-0.51	0.2	0.23	-0.16	-0.53	0.23	0.22	-0.17	-0.49	0.21	0.21
		0	0.21	-0.49	1	0.43	0.21	-0.67	0.82	0.44	0.16	-0.38	0.89	0.42
		0.5	0.29	-0.14	0.71	0.26	0.14	-0.22	0.58	0.22	0.05	-0.22	0.34	0.17
		1	0.45	0	0.78	0.25	0.24	-0.06	0.56	0.19	0.12	-0.19	0.42	0.2
	40	-1	-0.4	-0.68	-0.04	0.2	-0.27	-0.61	0.01	0.19	-0.2	-0.47	0.07	0.16
		-0.5	-0.18	-0.41	0.13	0.17	-0.14	-0.39	0.16	0.17	-0.13	-0.39	0.14	0.17
		0	0.16	-0.69	0.8	0.47	0.22	-0.85	0.97	0.54	0.19	-0.48	0.83	0.44
		0.5	0.29	0.01	0.59	0.18	0.1	-0.17	0.39	0.17	-0.03	-0.3	0.22	0.16
		1	0.46	0.13	0.81	0.21	0.17	-0.14	0.5	0.19	0.03	-0.24	0.34	0.17
60	20	-1	-0.47	-0.96	-0.04	0.26	-0.32	-0.67	0.05	0.22	-0.24	-0.59	0.12	0.23
		-0.5	-0.21	-0.66	0.26	0.27	-0.18	-0.56	0.22	0.22	-0.15	-0.5	0.24	0.22
		0	0.22	-0.49	0.79	0.39	0.28	-0.49	0.93	0.45	0.17	-0.56	0.85	0.42
		0.5	0.29	-0.2	0.67	0.27	0.09	-0.28	0.51	0.25	0.05	-0.28	0.39	0.2
		1	0.48	0.01	0.93	0.27	0.25	-0.13	0.56	0.22	0.12	-0.27	0.51	0.23
	40	-1	-0.4	-0.71	-0.07	0.2	-0.26	-0.5	0.06	0.17	-0.17	-0.43	0.1	0.16
		-0.5	-0.18	-0.42	0.1	0.15	-0.15	-0.41	0.12	0.16	-0.15	-0.43	0.09	0.16
		0	0.14	-0.8	0.8	0.49	0.27	-0.92	0.9	0.53	0.3	-0.37	0.95	0.46
		0.5	0.24	-0.08	0.52	0.19	0.06	-0.18	0.29	0.15	-0.03	-0.31	0.24	0.16
		1	0.44	0.07	0.81	0.23	0.14	-0.16	0.44	0.19	0.06	-0.19	0.34	0.15
90	20	-1	-0.29	-0.64	0.05	0.21	-0.32	-0.72	0.14	0.24	-0.32	-0.66	0.08	0.22
		-0.5	-0.19	-0.57	0.18	0.22	-0.15	-0.48	0.23	0.22	-0.18	-0.51	0.2	0.23
		0	0.13	-0.63	0.78	0.45	0.12	-0.59	0.81	0.44	0.16	-0.52	0.76	0.45
		0.5	0.12	-0.31	0.49	0.23	0.12	-0.21	0.49	0.22	0.08	-0.34	0.41	0.22
		1	0.31	-0.07	0.66	0.22	0.26	-0.15	0.59	0.22	0.28	-0.14	0.59	0.23
	40	-1	-0.24	-0.54	0	0.17	-0.24	-0.56	0.09	0.19	-0.25	-0.53	0.04	0.17
		-0.5	-0.11	-0.42	0.2	0.19	-0.13	-0.37	0.13	0.15	-0.11	-0.36	0.16	0.16
		0	0.15	-0.84	0.88	0.53	0.18	-0.68	0.94	0.54	0.16	-0.9	0.92	0.54
		0.5	0.09	-0.2	0.35	0.17	0.11	-0.21	0.42	0.18	0.08	-0.19	0.34	0.16
		1	0.19	-0.08	0.44	0.17	0.19	-0.1	0.52	0.19	0.18	-0.11	0.5	0.19

Appendix A.4: Bias and Error of group estimates: Mixture Proportion 4 (MP4) and Cluster Effect 1 (CE1) for the true model

			Cluster Type												
			1				2				3				
			Bias			Error	Bias			Error	Bias			Error	
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.29	-0.64	0.05	0.21	-0.32	-0.72	0.14	0.24	-0.32	-0.66	0.08	0.22
		-0.5	-0.19	-0.57	0.18	0.22	-0.15	-0.48	0.23	0.22	-0.18	-0.51	0.2	0.23
		0	0.13	-0.63	0.78	0.45	0.12	-0.59	0.81	0.44	0.16	-0.52	0.76	0.45
		0.5	0.12	-0.31	0.49	0.23	0.12	-0.21	0.49	0.22	0.08	-0.34	0.41	0.22
		1	0.31	-0.07	0.66	0.22	0.26	-0.15	0.59	0.22	0.28	-0.14	0.59	0.23
	40	-1	-0.24	-0.54	0	0.17	-0.24	-0.56	0.09	0.19	-0.25	-0.53	0.04	0.17
		-0.5	-0.11	-0.42	0.2	0.19	-0.13	-0.37	0.13	0.15	-0.11	-0.36	0.16	0.16
		0	0.15	-0.84	0.88	0.53	0.18	-0.68	0.94	0.54	0.16	-0.9	0.92	0.54
		0.5	0.09	-0.2	0.35	0.17	0.11	-0.21	0.42	0.18	0.08	-0.19	0.34	0.16
		1	0.19	-0.08	0.44	0.17	0.19	-0.1	0.52	0.19	0.18	-0.11	0.5	0.19
60	20	-1	-0.26	-0.68	0.12	0.24	-0.33	-0.72	0.12	0.24	-0.3	-0.67	0.15	0.26
		-0.5	-0.15	-0.49	0.34	0.25	-0.15	-0.54	0.31	0.26	-0.12	-0.49	0.27	0.22
		0	0.07	-0.76	0.86	0.48	0.16	-0.67	0.86	0.46	0.3	-0.46	0.89	0.4
		0.5	0.12	-0.21	0.45	0.2	0.12	-0.25	0.47	0.22	0.1	-0.34	0.45	0.24
		1	0.2	-0.2	0.58	0.24	0.25	-0.15	0.72	0.26	0.27	-0.09	0.64	0.22
	40	-1	-0.23	-0.5	0.08	0.18	-0.23	-0.48	0.02	0.16	-0.21	-0.5	0.09	0.18
		-0.5	-0.14	-0.44	0.15	0.19	-0.12	-0.36	0.16	0.16	-0.11	-0.42	0.17	0.16
		0	0.2	-0.76	0.95	0.54	0.29	-0.74	0.95	0.54	0.27	-0.49	0.96	0.49
		0.5	0.08	-0.21	0.37	0.18	0.09	-0.15	0.32	0.15	0.09	-0.19	0.34	0.16
		1	0.17	-0.08	0.49	0.17	0.18	-0.16	0.48	0.19	0.18	-0.09	0.44	0.16
90	20	-1	-0.32	-0.72	0.1	0.26	-0.35	-0.71	-0.04	0.23	-0.32	-0.7	0.05	0.23
		-0.5	-0.16	-0.5	0.2	0.24	-0.16	-0.53	0.18	0.21	-0.16	-0.54	0.19	0.22
		0	0.13	-0.76	0.81	0.48	0.24	-0.69	0.86	0.44	0.21	-0.64	0.81	0.43
		0.5	0.12	-0.3	0.53	0.24	0.11	-0.25	0.48	0.21	0.09	-0.24	0.49	0.22
		1	0.24	-0.13	0.66	0.24	0.22	-0.18	0.65	0.24	0.22	-0.17	0.55	0.21
	40	-1	-0.21	-0.56	0.1	0.2	-0.21	-0.46	0.11	0.18	-0.24	-0.53	0.02	0.17
		-0.5	-0.14	-0.44	0.22	0.19	-0.11	-0.4	0.22	0.19	-0.11	-0.41	0.16	0.17
		0	0.2	-0.85	1	0.54	0.26	-0.73	1.05	0.53	0.33	-0.66	0.93	0.49
		0.5	0.06	-0.23	0.34	0.18	0.04	-0.27	0.31	0.18	0.05	-0.26	0.37	0.18
		1	0.13	-0.21	0.41	0.19	0.14	-0.13	0.42	0.18	0.13	-0.18	0.44	0.17

Appendix A.5: Bias and Error of group estimates: Mixture Proportion 1 (MP1) and Cluster Effect 2 (CE2) for the true model

Cluster Number	Cluster Size	Cluster Effect	Cluster Type 1 Bias			Error	Cluster Type 2 Bias			Error	Cluster Type 3 Bias			Error
			Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.21	-0.56	0.18	0.22	0	0	0	0	0	0	0	0
		-0.5	-0.09	-0.48	0.28	0.23	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.01	-0.3	0.35	0.2	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.03	-0.26	0.26	0.16
		1	0	0	0	0	0	0	0	0	0.14	-0.16	0.41	0.19
	40	-1	-0.16	-0.36	0.05	0.13	0	0	0	0	0	0	0	0
		-0.5	-0.03	-0.28	0.24	0.16	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.04	-0.19	0.22	0.12	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.05	-0.17	0.31	0.16
		1	0	0	0	0	0	0	0	0	0.05	-0.2	0.29	0.14
60	20	-1	-0.19	-0.53	0.21	0.22	0	0	0	0	0	0	0	0
		-0.5	-0.07	-0.43	0.24	0.21	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	-0.33	0.26	0.2	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.07	-0.23	0.38	0.19
		1	0	0	0	0	0	0	0	0	0.11	-0.2	0.41	0.2
	40	-1	-0.2	-0.44	0.01	0.16	0	0	0	0	0	0	0	0
		-0.5	-0.06	-0.29	0.2	0.14	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.03	-0.21	0.28	0.14	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.05	-0.16	0.26	0.13
		1	0	0	0	0	0	0	0	0	0.08	-0.14	0.25	0.12
90	20	-1	-0.19	-0.54	0.15	0.22	0	0	0	0	0	0	0	0
		-0.5	-0.07	-0.42	0.29	0.21	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.01	-0.3	0.37	0.2	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.06	-0.25	0.35	0.19
		1	0	0	0	0	0	0	0	0	0.13	-0.17	0.37	0.16
	40	-1	-0.15	-0.41	0.12	0.16	0	0	0	0	0	0	0	0
		-0.5	-0.05	-0.27	0.19	0.14	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.03	-0.24	0.31	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.03	-0.21	0.24	0.14
		1	0	0	0	0	0	0	0	0	0.08	-0.15	0.3	0.14

Appendix A.6: Bias and Error of group estimates: Mixture Proportion 2 (MP2) and Cluster Effect 2 (CE2) for the true model

Cluster Number	Cluster Size	Cluster Effect	Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
			Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.19	-0.54	0.15	0.22	0	0	0	0	0	0	0	0
		-0.5	-0.07	-0.42	0.29	0.21	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.01	-0.3	0.37	0.2	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.06	-0.25	0.35	0.19
		1	0	0	0	0	0	0	0	0	0.13	-0.17	0.37	0.16
	40	-1	-0.15	-0.41	0.12	0.16	0	0	0	0	0	0	0	0
		-0.5	-0.05	-0.27	0.19	0.14	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.03	-0.24	0.31	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.03	-0.21	0.24	0.14
		1	0	0	0	0	0	0	0	0	0.08	-0.15	0.3	0.14
60	20	-1	-0.47	-0.87	-0.03	0.27	0	0	0	0	0	0	0	0
		-0.5	-0.12	-0.57	0.43	0.29	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.04	-0.38	0.45	0.25	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.08	-0.29	0.41	0.21
		1	0	0	0	0	0	0	0	0	0.22	-0.12	0.62	0.22
	40	-1	-0.35	-0.66	-0.02	0.19	0	0	0	0	0	0	0	0
		-0.5	-0.07	-0.41	0.23	0.19	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.04	-0.3	0.27	0.19	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.06	-0.25	0.33	0.17
		1	0	0	0	0	0	0	0	0	0.13	-0.12	0.37	0.14
90	20	-1	-0.44	-0.89	-0.03	0.27	0	0	0	0	0	0	0	0
		-0.5	-0.17	-0.49	0.22	0.24	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.03	-0.38	0.37	0.24	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.1	-0.26	0.48	0.21
		1	0	0	0	0	0	0	0	0	0.21	-0.15	0.65	0.24
	40	-1	-0.34	-0.69	-0.05	0.2	0	0	0	0	0	0	0	0
		-0.5	-0.1	-0.36	0.22	0.18	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.03	-0.2	0.37	0.17	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.04	-0.23	0.31	0.17
		1	0	0	0	0	0	0	0	0	0.07	-0.22	0.29	0.16

Appendix A.7: Bias and Error of group estimates: Mixture Proportion 3 (MP3) and Cluster Effect 2 (CE2) for the true model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.44	-0.89	-0.03	0.27	0	0	0	0	0	0	0	0
		-0.5	-0.17	-0.49	0.22	0.24	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.03	-0.38	0.37	0.24	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.1	-0.26	0.48	0.21
		1	0	0	0	0	0	0	0	0	0.21	-0.15	0.65	0.24
	40	-1	-0.34	-0.69	-0.05	0.2	0	0	0	0	0	0	0	0
		-0.5	-0.1	-0.36	0.22	0.18	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.03	-0.2	0.37	0.17	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.04	-0.23	0.31	0.17
		1	0	0	0	0	0	0	0	0	0.07	-0.22	0.29	0.16
60	20	-1	-0.44	-0.88	0.01	0.29	0	0	0	0	0	0	0	0
		-0.5	-0.17	-0.62	0.29	0.27	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.03	-0.41	0.4	0.24	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.1	-0.33	0.5	0.23
		1	0	0	0	0	0	0	0	0	0.21	-0.15	0.57	0.22
	40	-1	-0.36	-0.67	-0.02	0.2	0	0	0	0	0	0	0	0
		-0.5	-0.12	-0.47	0.2	0.2	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.05	-0.25	0.33	0.17	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.02	-0.23	0.32	0.16
		1	0	0	0	0	0	0	0	0	0.12	-0.18	0.41	0.17
90	20	-1	-0.35	-0.73	-0.01	0.22	0	0	0	0	0	0	0	0
		-0.5	-0.18	-0.52	0.17	0.21	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.05	-0.43	0.33	0.22	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.15	-0.21	0.54	0.24
		1	0	0	0	0	0	0	0	0	0.26	-0.18	0.67	0.26
	40	-1	-0.26	-0.56	0	0.17	0	0	0	0	0	0	0	0
		-0.5	-0.13	-0.4	0.2	0.19	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.03	-0.3	0.22	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.08	-0.19	0.3	0.16
		1	0	0	0	0	0	0	0	0	0.15	-0.2	0.44	0.19

Appendix A.8: Bias and Error of group estimates: Mixture Proportion 4 (MP4) and Cluster Effect 2 (CE2) for the true model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.35	-0.73	-0.01	0.22	0	0	0	0	0	0	0	0
		-0.5	-0.18	-0.52	0.17	0.21	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.05	-0.43	0.33	0.22	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.15	-0.21	0.54	0.24
		1	0	0	0	0	0	0	0	0	0.26	-0.18	0.67	0.26
	40	-1	-0.26	-0.56	0	0.17	0	0	0	0	0	0	0	0
		-0.5	-0.13	-0.4	0.2	0.19	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.03	-0.3	0.22	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.08	-0.19	0.3	0.16
		1	0	0	0	0	0	0	0	0	0.15	-0.2	0.44	0.19
60	20	-1	-0.35	-0.85	0.08	0.25	0	0	0	0	0	0	0	0
		-0.5	-0.16	-0.58	0.17	0.22	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.04	-0.38	0.25	0.2	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.13	-0.24	0.49	0.21
		1	0	0	0	0	0	0	0	0	0.23	-0.07	0.52	0.19
	40	-1	-0.26	-0.56	0	0.18	0	0	0	0	0	0	0	0
		-0.5	-0.12	-0.35	0.15	0.14	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.01	-0.33	0.31	0.19	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.08	-0.16	0.34	0.16
		1	0	0	0	0	0	0	0	0	0.17	-0.12	0.44	0.18
90	20	-1	-0.4	-0.85	-0.03	0.24	0	0	0	0	0	0	0	0
		-0.5	-0.21	-0.56	0.16	0.22	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.03	-0.47	0.33	0.25	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.12	-0.3	0.47	0.22
		1	0	0	0	0	0	0	0	0	0.23	-0.13	0.6	0.23
	40	-1	-0.22	-0.51	0.06	0.19	0	0	0	0	0	0	0	0
		-0.5	-0.15	-0.46	0.1	0.17	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	-0.28	0.28	0.18	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.05	-0.21	0.32	0.17
		1	0	0	0	0	0	0	0	0	0.18	-0.08	0.48	0.17

Appendix A.9: Bias and Error of group estimates: Mixture Proportion 1 (MP1) and Cluster Effect 3 (CE3) for the true model

Cluster Number	Cluster Size	Cluster Effect	Cluster Type 1 Bias Mean	2.5%	97.5%	Error Mean	Cluster Type 2 Bias Mean	2.5%	97.5%	Error Mean	Cluster Type 3 Bias Mean	2.5%	97.5%	Error Mean
30	20	-1	0	0	0	0	0	0	0	0	-0.26	-0.53	0.08	0.19
		-0.5	0	0	0	0	0	0	0	0	-0.19	-0.48	0.15	0.19
		0	0	0	0	0	-0.07	-0.38	0.23	0.18	0	0	0	0
		0.5	0.09	-0.24	0.41	0.21	0	0	0	0	0	0	0	0
		1	0.26	-0.14	0.76	0.27	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.22	-0.41	0	0.13
		-0.5	0	0	0	0	0	0	0	0	-0.17	-0.37	0.04	0.13
		0	0	0	0	0	-0.09	-0.3	0.11	0.13	0	0	0	0
		0.5	0.1	-0.18	0.36	0.15	0	0	0	0	0	0	0	0
		1	0.2	-0.08	0.48	0.17	0	0	0	0	0	0	0	0
60	20	-1	0	0	0	0	0	0	0	0	-0.3	-0.62	0.02	0.2
		-0.5	0	0	0	0	0	0	0	0	-0.19	-0.5	0.11	0.19
		0	0	0	0	0	-0.09	-0.36	0.19	0.18	0	0	0	0
		0.5	0.08	-0.25	0.43	0.21	0	0	0	0	0	0	0	0
		1	0.21	-0.09	0.54	0.2	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.21	-0.43	-0.02	0.14
		-0.5	0	0	0	0	0	0	0	0	-0.18	-0.38	0.04	0.13
		0	0	0	0	0	-0.11	-0.35	0.13	0.14	0	0	0	0
		0.5	0.07	-0.16	0.33	0.15	0	0	0	0	0	0	0	0
		1	0.16	-0.06	0.43	0.15	0	0	0	0	0	0	0	0
90	20	-1	0	0	0	0	0	0	0	0	-0.26	-0.56	0.01	0.18
		-0.5	0	0	0	0	0	0	0	0	-0.23	-0.54	0.04	0.17
		0	0	0	0	0	-0.11	-0.47	0.2	0.19	0	0	0	0
		0.5	0.09	-0.29	0.49	0.23	0	0	0	0	0	0	0	0
		1	0.19	-0.09	0.5	0.2	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.21	-0.43	0	0.14
		-0.5	0	0	0	0	0	0	0	0	-0.2	-0.44	0.03	0.14
		0	0	0	0	0	-0.13	-0.36	0.09	0.14	0	0	0	0
		0.5	0.07	-0.19	0.29	0.14	0	0	0	0	0	0	0	0
		1	0.18	-0.06	0.42	0.15	0	0	0	0	0	0	0	0

Appendix A.10: Bias and Error of group estimates: Mixture Proportion 2 (MP2) and Cluster Effect 3 (CE3) for the true model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	0	0	0	0	0	0	0	0	-0.26	-0.56	0.01	0.18
		-0.5	0	0	0	0	0	0	0	0	-0.23	-0.54	0.04	0.17
		0	0	0	0	0	-0.11	-0.47	0.2	0.19	0	0	0	0
		0.5	0.09	-0.29	0.49	0.23	0	0	0	0	0	0	0	0
		1	0.19	-0.09	0.5	0.2	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.21	-0.43	0	0.14
		-0.5	0	0	0	0	0	0	0	0	-0.2	-0.44	0.03	0.14
		0	0	0	0	0	-0.13	-0.36	0.09	0.14	0	0	0	0
		0.5	0.07	-0.19	0.29	0.14	0	0	0	0	0	0	0	0
		1	0.18	-0.06	0.42	0.15	0	0	0	0	0	0	0	0
60	20	-1	0	0	0	0	0	0	0	0	-0.44	-0.72	-0.07	0.2
		-0.5	0	0	0	0	0	0	0	0	-0.26	-0.51	0.11	0.18
		0	0	0	0	0	-0.06	-0.43	0.26	0.19	0	0	0	0
		0.5	0.25	-0.1	0.57	0.21	0	0	0	0	0	0	0	0
		1	0.52	0.01	0.87	0.26	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.36	-0.6	-0.11	0.17
		-0.5	0	0	0	0	0	0	0	0	-0.26	-0.46	-0.03	0.14
		0	0	0	0	0	-0.15	-0.4	0.1	0.15	0	0	0	0
		0.5	0.26	-0.05	0.55	0.18	0	0	0	0	0	0	0	0
		1	0.5	0.15	0.8	0.19	0	0	0	0	0	0	0	0
90	20	-1	0	0	0	0	0	0	0	0	-0.43	-0.74	-0.05	0.2
		-0.5	0	0	0	0	0	0	0	0	-0.28	-0.56	-0.03	0.17
		0	0	0	0	0	-0.1	-0.4	0.2	0.17	0	0	0	0
		0.5	0.24	-0.09	0.58	0.21	0	0	0	0	0	0	0	0
		1	0.48	0.03	0.87	0.24	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.35	-0.68	-0.05	0.19
		-0.5	0	0	0	0	0	0	0	0	-0.24	-0.45	0.03	0.15
		0	0	0	0	0	-0.1	-0.38	0.2	0.16	0	0	0	0
		0.5	0.22	-0.08	0.52	0.18	0	0	0	0	0	0	0	0
		1	0.48	0.13	0.77	0.2	0	0	0	0	0	0	0	0

Appendix A.11: Bias and Error of group estimates: Mixture Proportion 3 (MP3) and Cluster Effect 3 (CE3) for the true model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	0	0	0	0	0	0	0	0	-0.43	-0.74	-0.05	0.2
		-0.5	0	0	0	0	0	0	0	0	-0.28	-0.56	-0.03	0.17
		0	0	0	0	0	-0.1	-0.4	0.2	0.17	0	0	0	0
		0.5	0.24	-0.09	0.58	0.21	0	0	0	0	0	0	0	0
		1	0.48	0.03	0.87	0.24	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.35	-0.68	-0.05	0.19
		-0.5	0	0	0	0	0	0	0	0	-0.24	-0.45	0.03	0.15
		0	0	0	0	0	-0.1	-0.38	0.2	0.16	0	0	0	0
		0.5	0.22	-0.08	0.52	0.18	0	0	0	0	0	0	0	0
		1	0.48	0.13	0.77	0.2	0	0	0	0	0	0	0	0
60	20	-1	0	0	0	0	0	0	0	0	-0.46	-0.78	-0.12	0.21
		-0.5	0	0	0	0	0	0	0	0	-0.26	-0.52	0	0.18
		0	0	0	0	0	-0.1	-0.42	0.27	0.21	0	0	0	0
		0.5	0.25	-0.01	0.53	0.17	0	0	0	0	0	0	0	0
		1	0.54	0.17	0.93	0.23	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.34	-0.58	-0.07	0.16
		-0.5	0	0	0	0	0	0	0	0	-0.25	-0.45	-0.03	0.13
		0	0	0	0	0	-0.11	-0.35	0.11	0.15	0	0	0	0
		0.5	0.23	-0.04	0.51	0.17	0	0	0	0	0	0	0	0
		1	0.49	0.18	0.8	0.2	0	0	0	0	0	0	0	0
90	20	-1	0	0	0	0	0	0	0	0	-0.34	-0.76	0.1	0.25
		-0.5	0	0	0	0	0	0	0	0	-0.18	-0.49	0.14	0.19
		0	0	0	0	0	-0.02	-0.43	0.4	0.24	0	0	0	0
		0.5	0.12	-0.18	0.53	0.22	0	0	0	0	0	0	0	0
		1	0.3	-0.1	0.63	0.22	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.28	-0.58	0.09	0.19
		-0.5	0	0	0	0	0	0	0	0	-0.16	-0.38	0.13	0.18
		0	0	0	0	0	-0.02	-0.31	0.23	0.17	0	0	0	0
		0.5	0.11	-0.17	0.34	0.15	0	0	0	0	0	0	0	0
		1	0.2	-0.08	0.51	0.18	0	0	0	0	0	0	0	0

Appendix A.12: Bias and Error of group estimates: Mixture Proportion 4 (MP4) and Cluster Effect 3 (CE3) for the true model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	0	0	0	0	0	0	0	0	-0.34	-0.76	0.1	0.25
		-0.5	0	0	0	0	0	0	0	0	-0.18	-0.49	0.14	0.19
		0	0	0	0	0	-0.02	-0.43	0.4	0.24	0	0	0	0
		0.5	0.12	-0.18	0.53	0.22	0	0	0	0	0	0	0	0
		1	0.3	-0.1	0.63	0.22	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.28	-0.58	0.09	0.19
		-0.5	0	0	0	0	0	0	0	0	-0.16	-0.38	0.13	0.18
		0	0	0	0	0	-0.02	-0.31	0.23	0.17	0	0	0	0
		0.5	0.11	-0.17	0.34	0.15	0	0	0	0	0	0	0	0
		1	0.2	-0.08	0.51	0.18	0	0	0	0	0	0	0	0
60	20	-1	0	0	0	0	0	0	0	0	-0.37	-0.74	-0.03	0.22
		-0.5	0	0	0	0	0	0	0	0	-0.21	-0.59	0.1	0.21
		0	0	0	0	0	-0.01	-0.39	0.36	0.22	0	0	0	0
		0.5	0.13	-0.17	0.51	0.21	0	0	0	0	0	0	0	0
		1	0.25	-0.1	0.66	0.22	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.27	-0.55	0.06	0.19
		-0.5	0	0	0	0	0	0	0	0	-0.12	-0.33	0.13	0.16
		0	0	0	0	0	-0.04	-0.29	0.31	0.19	0	0	0	0
		0.5	0.1	-0.18	0.38	0.17	0	0	0	0	0	0	0	0
		1	0.18	-0.12	0.42	0.16	0	0	0	0	0	0	0	0
90	20	-1	0	0	0	0	0	0	0	0	-0.34	-0.76	0.04	0.24
		-0.5	0	0	0	0	0	0	0	0	-0.17	-0.56	0.17	0.22
		0	0	0	0	0	-0.03	-0.39	0.37	0.23	0	0	0	0
		0.5	0.1	-0.26	0.47	0.21	0	0	0	0	0	0	0	0
		1	0.26	-0.03	0.59	0.19	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.26	-0.6	0.11	0.2
		-0.5	0	0	0	0	0	0	0	0	-0.13	-0.42	0.16	0.18
		0	0	0	0	0	-0.03	-0.27	0.22	0.16	0	0	0	0
		0.5	0.1	-0.17	0.35	0.17	0	0	0	0	0	0	0	0
		1	0.17	-0.16	0.44	0.19	0	0	0	0	0	0	0	0

Appendix A.13: Bias and Error of group estimates: Cluster Effect 4 (CE4) for the true model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Mixture Proportion	Cluster Number	Cluster Size	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
1	30	20	0.06	-0.07	0.22	0.35	0.02	-0.1	0.15	0.34	-0.01	-0.17	0.13	0.32
		40	0.03	-0.07	0.18	0.36	-0.02	-0.15	0.12	0.41	-0.05	-0.2	0.09	0.43
	60	20	0.05	-0.03	0.16	0.31	0.02	-0.09	0.1	0.33	-0.01	-0.1	0.09	0.41
		40	0.05	-0.03	0.16	0.36	0.01	-0.09	0.09	0.44	-0.02	-0.12	0.06	0.39
	90	20	0.04	-0.03	0.12	0.33	0.01	-0.08	0.09	0.34	-0.01	-0.13	0.07	0.39
		40	0.05	-0.01	0.14	0.38	0	-0.08	0.08	0.34	-0.03	-0.13	0.04	0.42
2	30	20	0.03	-0.09	0.15	0.23	0.01	-0.14	0.2	0.26	-0.01	-0.18	0.12	0.31
		40	0.07	-0.04	0.21	0.27	0.01	-0.13	0.12	0.31	-0.03	-0.2	0.1	0.36
	60	20	0.05	-0.05	0.14	0.23	0.01	-0.09	0.12	0.26	-0.02	-0.12	0.07	0.31
		40	0.06	-0.01	0.15	0.2	0.01	-0.08	0.09	0.37	-0.04	-0.14	0.06	0.36
	90	20	0.04	-0.03	0.13	0.25	0.01	-0.1	0.09	0.25	-0.01	-0.09	0.06	0.32
		40	0.07	0	0.16	0.26	0.01	-0.07	0.09	0.36	-0.04	-0.13	0.05	0.35
3	30	20	0	-0.14	0.12	0.27	0.02	-0.1	0.13	0.29	0.01	-0.16	0.15	0.28
		40	0	-0.15	0.14	0.33	0.01	-0.11	0.14	0.34	0.01	-0.14	0.15	0.29
	60	20	0.02	-0.09	0.11	0.29	0.01	-0.08	0.13	0.3	0.02	-0.09	0.13	0.34
		40	0.01	-0.1	0.1	0.31	0	-0.11	0.12	0.41	0	-0.11	0.11	0.37
	90	20	0.01	-0.08	0.09	0.26	0.01	-0.06	0.1	0.36	0.01	-0.1	0.09	0.3
		40	0.01	-0.07	0.1	0.38	0.01	-0.09	0.11	0.36	0.01	-0.07	0.09	0.37
4	30	20	0.01	-0.15	0.16	0.41	0	-0.15	0.14	0.38	0.02	-0.11	0.17	0.42
		40	-0.01	-0.17	0.13	0.46	0	-0.15	0.13	0.46	0	-0.12	0.13	0.42
	60	20	0.01	-0.11	0.11	0.36	0	-0.13	0.13	0.38	0	-0.13	0.11	0.35
		40	0.02	-0.09	0.12	0.39	0.01	-0.09	0.13	0.43	0.01	-0.09	0.11	0.42
	90	20	0.01	-0.06	0.1	0.29	0.01	-0.08	0.09	0.4	0.01	-0.08	0.09	0.41
		40	0.01	-0.07	0.08	0.41	0.01	-0.08	0.08	0.41	0.01	-0.07	0.09	0.49

Appendix A.14: Bias and Error of group estimates: Cluster Effect 5 (CE5) for the true model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Mixture Proportion	Cluster Number	Cluster Size	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
1	30	20	0.05	-0.21	0.35	0.8	0.01	-0.26	0.35	0.83	-0.04	-0.31	0.29	0.66
		40	0.06	-0.19	0.32	0.82	0.01	-0.29	0.27	1.21	-0.02	-0.31	0.21	0.79
	60	20	0.06	-0.08	0.19	0.92	0.02	-0.14	0.18	0.97	-0.01	-0.2	0.13	0.89
		40	0.07	-0.11	0.33	0.96	0.01	-0.2	0.23	1.09	-0.02	-0.23	0.19	0.92
	90	20	0.05	-0.09	0.2	0.95	0.01	-0.14	0.14	0.85	-0.01	-0.2	0.13	0.88
		40	0.06	-0.04	0.23	1.07	0.01	-0.09	0.16	1.15	-0.01	-0.18	0.11	0.63
2	30	20	0.07	-0.13	0.37	0.54	0.06	-0.15	0.32	0.53	0.03	-0.22	0.31	0.61
		40	0.08	-0.09	0.31	0.36	0.03	-0.22	0.21	0.55	0.01	-0.25	0.17	0.66
	60	20	0.07	-0.06	0.24	0.41	0.01	-0.16	0.17	0.61	-0.01	-0.17	0.13	0.64
		40	0.06	-0.09	0.2	0.5	0.01	-0.16	0.14	0.62	-0.02	-0.21	0.13	0.68
	90	20	0.06	-0.06	0.21	0.59	0.01	-0.15	0.17	0.56	-0.02	-0.19	0.11	0.7
		40	0.06	-0.04	0.19	0.44	0.02	-0.1	0.11	0.68	0	-0.15	0.1	0.74
3	30	20	0.02	-0.21	0.24	0.63	0.01	-0.24	0.27	0.58	0	-0.22	0.24	0.64
		40	0.03	-0.19	0.25	0.81	0.04	-0.19	0.25	0.84	0.03	-0.16	0.3	0.69
	60	20	0.02	-0.17	0.24	0.64	0.03	-0.19	0.23	0.65	0.03	-0.18	0.25	0.58
		40	0.02	-0.16	0.21	0.73	0.03	-0.18	0.2	0.8	0.03	-0.16	0.2	0.76
	90	20	0.03	-0.11	0.16	0.74	0.03	-0.12	0.18	0.76	0.03	-0.14	0.18	0.77
		40	0.03	-0.11	0.17	0.65	0.02	-0.13	0.18	0.76	0.02	-0.13	0.16	0.78
4	30	20	0.02	-0.32	0.35	0.87	0.02	-0.3	0.28	0.8	0.03	-0.28	0.32	0.95
		40	0.04	-0.23	0.29	1.01	0.05	-0.25	0.33	0.96	0.05	-0.27	0.32	1.06
	60	20	0.02	-0.23	0.22	1.01	0.02	-0.25	0.25	0.91	0.02	-0.22	0.22	0.97
		40	0.02	-0.16	0.18	1.16	0.02	-0.14	0.19	0.97	0.01	-0.13	0.19	1.1
	90	20	0.03	-0.1	0.17	1.01	0.03	-0.11	0.2	1.16	0.03	-0.1	0.19	0.95
		40	0.01	-0.16	0.17	0.79	0.01	-0.14	0.16	0.94	0.01	-0.14	0.16	0.98

Appendix A.15: Bias and Error of group estimates: Mixture Proportion 1 (MP1) and Cluster Effect 1 (CE1) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.47	-0.67	-0.21	0.14	-0.33	-0.6	-0.06	0.16	-0.18	-0.44	0.03	0.14
		-0.5	-0.18	-0.47	0.06	0.16	-0.19	-0.42	0.01	0.14	-0.15	-0.42	0.09	0.16
		0	0.53	0.33	0.79	0.14	0.67	0.4	0.94	0.16	-0.07	-0.32	0.14	0.15
		0.5	0.36	0.09	0.62	0.15	0.17	-0.04	0.41	0.15	-0.04	-0.32	0.18	0.15
		1	0.65	0.38	0.89	0.15	0.32	0.05	0.56	0.15	-0.01	-0.25	0.23	0.15
	40	-1	-0.41	-0.58	-0.21	0.11	-0.28	-0.45	-0.12	0.11	-0.15	-0.3	0.04	0.1
		-0.5	-0.16	-0.35	0.02	0.12	-0.15	-0.34	0.02	0.12	-0.15	-0.34	0.03	0.12
		0	0.59	0.42	0.79	0.11	0.72	0.55	0.88	0.11	-0.11	-0.32	0.07	0.12
		0.5	0.36	0.19	0.54	0.11	0.14	-0.05	0.34	0.12	-0.08	-0.24	0.1	0.11
		1	0.63	0.44	0.84	0.12	0.3	0.12	0.49	0.11	-0.06	-0.26	0.15	0.12
60	20	-1	-0.47	-0.76	-0.16	0.17	-0.34	-0.56	-0.11	0.15	-0.18	-0.46	0.08	0.18
		-0.5	-0.2	-0.49	0.02	0.16	-0.17	-0.44	0.05	0.15	-0.16	-0.4	0.1	0.15
		0	0.53	0.24	0.84	0.17	0.66	0.44	0.89	0.15	-0.07	-0.32	0.15	0.15
		0.5	0.37	0.12	0.63	0.15	0.17	-0.08	0.41	0.15	-0.04	-0.32	0.2	0.16
		1	0.63	0.4	0.9	0.14	0.32	0.02	0.6	0.18	0	-0.28	0.26	0.17
	40	-1	-0.41	-0.6	-0.23	0.12	-0.29	-0.47	-0.1	0.12	-0.16	-0.31	0.02	0.11
		-0.5	-0.15	-0.4	0.07	0.13	-0.13	-0.28	0.03	0.1	-0.13	-0.32	0.08	0.12
		0	0.59	0.4	0.77	0.12	0.71	0.53	0.9	0.12	-0.1	-0.31	0.1	0.12
		0.5	0.36	0.14	0.53	0.11	0.14	-0.04	0.32	0.12	-0.08	-0.25	0.13	0.12
		1	0.63	0.45	0.84	0.11	0.3	0.1	0.49	0.12	-0.06	-0.23	0.15	0.12
90	20	-1	-0.47	-0.72	-0.22	0.14	-0.3	-0.56	-0.07	0.16	-0.2	-0.46	0.04	0.16
		-0.5	-0.18	-0.41	0.06	0.15	-0.13	-0.33	0.09	0.14	-0.16	-0.41	0.12	0.16
		0	0.53	0.28	0.78	0.14	0.7	0.44	0.93	0.16	-0.09	-0.35	0.23	0.17
		0.5	0.37	0.1	0.64	0.16	0.18	-0.1	0.43	0.16	-0.05	-0.3	0.2	0.16
		1	0.64	0.39	0.86	0.15	0.32	0.05	0.61	0.16	0.01	-0.25	0.31	0.17
	40	-1	-0.43	-0.66	-0.21	0.13	-0.31	-0.48	-0.09	0.12	-0.15	-0.33	0.01	0.11
		-0.5	-0.18	-0.35	0.02	0.11	-0.15	-0.36	0.04	0.11	-0.14	-0.31	0.04	0.11
		0	0.57	0.34	0.79	0.13	0.69	0.52	0.91	0.12	-0.11	-0.33	0.08	0.12
		0.5	0.37	0.19	0.56	0.11	0.14	-0.04	0.33	0.11	-0.1	-0.31	0.08	0.12
		1	0.65	0.45	0.87	0.12	0.28	0.06	0.46	0.12	-0.05	-0.24	0.12	0.1

Appendix A.16: Bias and Error of group estimates: Mixture Proportion 2 (MP2) and Cluster Effect 1 (CE1) for Misspecified model

Cluster Number	Cluster Size	Cluster Effect	Cluster Type 1 Bias Mean	2.5%	97.5%	Error Mean	Cluster Type 2 Bias Mean	2.5%	97.5%	Error Mean	Cluster Type 3 Bias Mean	2.5%	97.5%	Error Mean
30	20	-1	-0.47	-0.72	-0.22	0.14	-0.3	-0.56	-0.07	0.16	-0.2	-0.46	0.04	0.16
		-0.5	-0.18	-0.41	0.06	0.15	-0.13	-0.33	0.09	0.14	-0.16	-0.41	0.12	0.16
		0	0.53	0.28	0.78	0.14	0.7	0.44	0.93	0.16	-0.09	-0.35	0.23	0.17
		0.5	0.37	0.1	0.64	0.16	0.18	-0.1	0.43	0.16	-0.05	-0.3	0.2	0.16
		1	0.64	0.39	0.86	0.15	0.32	0.05	0.61	0.16	0.01	-0.25	0.31	0.17
	40	-1	-0.43	-0.66	-0.21	0.13	-0.31	-0.48	-0.09	0.12	-0.15	-0.33	0.01	0.11
		-0.5	-0.18	-0.35	0.02	0.11	-0.15	-0.36	0.04	0.11	-0.14	-0.31	0.04	0.11
		0	0.57	0.34	0.79	0.13	0.69	0.52	0.91	0.12	-0.11	-0.33	0.08	0.12
		0.5	0.37	0.19	0.56	0.11	0.14	-0.04	0.33	0.11	-0.1	-0.31	0.08	0.12
		1	0.65	0.45	0.87	0.12	0.28	0.06	0.46	0.12	-0.05	-0.24	0.12	0.1
60	20	-1	-0.72	-0.93	-0.53	0.12	-0.62	-0.86	-0.41	0.13	-0.48	-0.71	-0.23	0.15
		-0.5	-0.32	-0.5	-0.12	0.12	-0.3	-0.51	-0.03	0.14	-0.3	-0.52	-0.09	0.14
		0	0.28	0.07	0.47	0.12	0.38	0.14	0.59	0.13	0.52	0.29	0.77	0.15
		0.5	0.47	0.22	0.69	0.14	0.27	0.03	0.45	0.14	0.14	-0.09	0.34	0.13
		1	0.88	0.66	1.09	0.12	0.62	0.42	0.86	0.14	0.3	0.07	0.54	0.15
	40	-1	-0.68	-0.88	-0.47	0.12	-0.54	-0.67	-0.38	0.09	-0.42	-0.61	-0.22	0.11
		-0.5	-0.29	-0.47	-0.11	0.11	-0.27	-0.45	-0.12	0.09	-0.25	-0.43	-0.04	0.11
		0	0.32	0.12	0.53	0.12	0.46	0.33	0.62	0.09	0.58	0.39	0.78	0.11
		0.5	0.49	0.32	0.67	0.11	0.29	0.1	0.46	0.11	0.07	-0.13	0.27	0.12
		1	0.87	0.67	1.06	0.12	0.55	0.32	0.69	0.11	0.23	-0.01	0.42	0.12
90	20	-1	-0.73	-0.97	-0.51	0.14	-0.62	-0.87	-0.36	0.17	-0.5	-0.71	-0.27	0.13
		-0.5	-0.33	-0.55	-0.11	0.13	-0.29	-0.47	-0.08	0.12	-0.26	-0.49	0	0.14
		0	0.27	0.03	0.49	0.14	0.38	0.13	0.64	0.17	0.5	0.29	0.73	0.13
		0.5	0.48	0.25	0.68	0.14	0.31	0.05	0.57	0.16	0.14	-0.09	0.37	0.15
		1	0.87	0.68	1.08	0.13	0.6	0.44	0.77	0.11	0.3	0.04	0.53	0.15
	40	-1	-0.69	-0.89	-0.53	0.11	-0.54	-0.71	-0.36	0.1	-0.43	-0.58	-0.26	0.1
		-0.5	-0.29	-0.45	-0.11	0.11	-0.27	-0.49	-0.1	0.12	-0.27	-0.45	-0.08	0.11
		0	0.31	0.11	0.47	0.11	0.46	0.29	0.64	0.1	0.57	0.42	0.74	0.1
		0.5	0.49	0.32	0.62	0.09	0.27	0.06	0.45	0.12	0.07	-0.13	0.24	0.11
		1	0.87	0.69	1.03	0.1	0.57	0.39	0.74	0.11	0.24	0.05	0.44	0.11

Appendix A.17: Bias and Error of group estimates: Mixture Proportion 3 (MP3) and Cluster Effect 1 (CE1) for Misspecified model

Cluster Number	Cluster Size	Cluster Effect	Cluster Type 1 Bias			Error	Cluster Type 2 Bias			Error	Cluster Type 3 Bias			Error
			Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.73	-0.97	-0.51	0.14	-0.62	-0.87	-0.36	0.17	-0.5	-0.71	-0.27	0.13
		-0.5	-0.33	-0.55	-0.11	0.13	-0.29	-0.47	-0.08	0.12	-0.26	-0.49	0	0.14
		0	0.27	0.03	0.49	0.14	0.38	0.13	0.64	0.17	0.5	0.29	0.73	0.13
		0.5	0.48	0.25	0.68	0.14	0.31	0.05	0.57	0.16	0.14	-0.09	0.37	0.15
		1	0.87	0.68	1.08	0.13	0.6	0.44	0.77	0.11	0.3	0.04	0.53	0.15
	40	-1	-0.69	-0.89	-0.53	0.11	-0.54	-0.71	-0.36	0.1	-0.43	-0.58	-0.26	0.1
		-0.5	-0.29	-0.45	-0.11	0.11	-0.27	-0.49	-0.1	0.12	-0.27	-0.45	-0.08	0.11
		0	0.31	0.11	0.47	0.11	0.46	0.29	0.64	0.1	0.57	0.42	0.74	0.1
		0.5	0.49	0.32	0.62	0.09	0.27	0.06	0.45	0.12	0.07	-0.13	0.24	0.11
		1	0.87	0.69	1.03	0.1	0.57	0.39	0.74	0.11	0.24	0.05	0.44	0.11
60	20	-1	-0.72	-0.93	-0.51	0.13	-0.59	-0.8	-0.32	0.14	-0.49	-0.69	-0.24	0.14
		-0.5	-0.32	-0.53	-0.1	0.13	-0.28	-0.48	-0.09	0.13	-0.26	-0.48	-0.02	0.14
		0	0.28	0.07	0.49	0.13	0.41	0.2	0.68	0.14	0.51	0.31	0.76	0.14
		0.5	0.48	0.25	0.69	0.13	0.29	0.1	0.47	0.12	0.14	-0.1	0.31	0.11
		1	0.87	0.66	1.08	0.14	0.61	0.35	0.84	0.15	0.3	0.06	0.53	0.14
	40	-1	-0.67	-0.88	-0.5	0.11	-0.57	-0.74	-0.4	0.09	-0.43	-0.59	-0.29	0.1
		-0.5	-0.3	-0.5	-0.12	0.11	-0.26	-0.47	-0.09	0.12	-0.28	-0.47	-0.13	0.11
		0	0.33	0.12	0.5	0.11	0.43	0.27	0.6	0.09	0.57	0.41	0.71	0.1
		0.5	0.5	0.32	0.68	0.11	0.3	0.13	0.49	0.11	0.06	-0.12	0.24	0.11
		1	0.87	0.69	1.06	0.11	0.55	0.4	0.74	0.1	0.24	0.01	0.43	0.11
90	20	-1	-0.64	-0.84	-0.4	0.13	-0.64	-0.86	-0.43	0.13	-0.63	-0.82	-0.43	0.13
		-0.5	-0.32	-0.51	-0.1	0.13	-0.32	-0.5	-0.12	0.12	-0.32	-0.54	-0.08	0.14
		0	0.36	0.16	0.6	0.13	0.36	0.14	0.57	0.13	0.37	0.18	0.57	0.13
		0.5	0.33	0.12	0.53	0.12	0.35	0.14	0.57	0.13	0.31	0.12	0.51	0.13
		1	0.63	0.38	0.87	0.14	0.62	0.39	0.85	0.14	0.62	0.37	0.82	0.13
	40	-1	-0.57	-0.75	-0.37	0.11	-0.56	-0.72	-0.4	0.09	-0.55	-0.75	-0.39	0.11
		-0.5	-0.28	-0.43	-0.15	0.1	-0.29	-0.48	-0.1	0.12	-0.28	-0.46	-0.11	0.1
		0	0.43	0.25	0.63	0.11	0.44	0.28	0.6	0.09	0.45	0.25	0.61	0.11
		0.5	0.28	0.11	0.46	0.1	0.28	0.11	0.45	0.11	0.29	0.13	0.46	0.1
		1	0.57	0.39	0.76	0.11	0.54	0.34	0.72	0.12	0.56	0.37	0.79	0.11

Appendix A.18: Bias and Error of group estimates: Mixture Proportion 4 (MP4) and Cluster Effect 1 (CE1) for Misspecified model

			Cluster Type										
			1				2				3		
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.64	-0.84	-0.4	0.13	-0.64	-0.86	-0.43	0.13	-0.63	-0.82	-0.43	0.13
		-0.5	-0.32	-0.51	-0.1	0.13	-0.32	-0.5	-0.12	0.12	-0.32	-0.54	-0.08	0.14
		0	0.36	0.16	0.6	0.13	0.36	0.14	0.57	0.13	0.37	0.18	0.57	0.13
		0.5	0.33	0.12	0.53	0.12	0.35	0.14	0.57	0.13	0.31	0.12	0.51	0.13
		1	0.63	0.38	0.87	0.14	0.62	0.39	0.85	0.14	0.62	0.37	0.82	0.13
	40	-1	-0.57	-0.75	-0.37	0.11	-0.56	-0.72	-0.4	0.09	-0.55	-0.75	-0.39	0.11
		-0.5	-0.28	-0.43	-0.15	0.1	-0.29	-0.48	-0.1	0.12	-0.28	-0.46	-0.11	0.1
		0	0.43	0.25	0.63	0.11	0.44	0.28	0.6	0.09	0.45	0.25	0.61	0.11
		0.5	0.28	0.11	0.46	0.1	0.28	0.11	0.45	0.11	0.29	0.13	0.46	0.1
		1	0.57	0.39	0.76	0.11	0.54	0.34	0.72	0.12	0.56	0.37	0.79	0.11
60	20	-1	-0.61	-0.84	-0.41	0.13	-0.63	-0.85	-0.4	0.13	-0.61	-0.8	-0.4	0.13
		-0.5	-0.28	-0.49	-0.05	0.14	-0.34	-0.58	-0.09	0.15	-0.32	-0.54	-0.03	0.15
		0	0.39	0.16	0.59	0.13	0.37	0.15	0.61	0.13	0.39	0.2	0.6	0.13
		0.5	0.31	0.09	0.53	0.13	0.32	0.04	0.52	0.15	0.3	0.1	0.54	0.13
		1	0.6	0.36	0.83	0.14	0.62	0.39	0.86	0.14	0.63	0.38	0.84	0.14
	40	-1	-0.57	-0.72	-0.41	0.1	-0.56	-0.72	-0.41	0.1	-0.55	-0.74	-0.39	0.11
		-0.5	-0.26	-0.48	-0.11	0.12	-0.29	-0.46	-0.11	0.11	-0.28	-0.46	-0.11	0.11
		0	0.43	0.28	0.59	0.1	0.44	0.28	0.59	0.1	0.45	0.26	0.61	0.11
		0.5	0.3	0.15	0.52	0.11	0.3	0.12	0.46	0.1	0.26	0.11	0.46	0.11
		1	0.59	0.42	0.74	0.1	0.57	0.4	0.77	0.12	0.56	0.4	0.74	0.1
90	20	-1	-0.61	-0.82	-0.39	0.13	-0.61	-0.86	-0.38	0.14	-0.63	-0.84	-0.38	0.15
		-0.5	-0.31	-0.54	-0.12	0.13	-0.29	-0.52	-0.09	0.14	-0.31	-0.53	-0.08	0.14
		0	0.39	0.18	0.61	0.13	0.39	0.14	0.62	0.14	0.37	0.16	0.62	0.15
		0.5	0.29	0.07	0.47	0.13	0.33	0.09	0.53	0.12	0.31	0.1	0.52	0.13
		1	0.64	0.4	0.91	0.14	0.64	0.42	0.86	0.13	0.6	0.41	0.83	0.13
	40	-1	-0.55	-0.7	-0.4	0.09	-0.55	-0.7	-0.37	0.1	-0.55	-0.74	-0.35	0.12
		-0.5	-0.28	-0.43	-0.1	0.1	-0.29	-0.45	-0.12	0.11	-0.26	-0.44	-0.07	0.11
		0	0.45	0.3	0.6	0.09	0.45	0.3	0.63	0.1	0.45	0.26	0.65	0.12
		0.5	0.28	0.12	0.49	0.11	0.3	0.14	0.48	0.11	0.29	0.13	0.43	0.1
		1	0.55	0.34	0.72	0.11	0.56	0.41	0.71	0.09	0.58	0.44	0.79	0.11

Appendix A.19: Bias and Error of group estimates: Mixture Proportion 1 (MP1) and Cluster Effect 2 (CE2) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.34	-0.63	-0.09	0.17	0	0	0	0	0	0	0	0
		-0.5	-0.05	-0.26	0.19	0.14	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.14	-0.11	0.4	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.04	-0.25	0.29	0.17
		1	0	0	0	0	0	0	0	0	0.08	-0.15	0.32	0.15
	40	-1	-0.3	-0.49	-0.1	0.13	0	0	0	0	0	0	0	0
		-0.5	-0.03	-0.28	0.19	0.13	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.11	-0.08	0.29	0.11	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.04	-0.11	0.21	0.1
		1	0	0	0	0	0	0	0	0	0.05	-0.11	0.2	0.1
60	20	-1	-0.31	-0.63	-0.07	0.17	0	0	0	0	0	0	0	0
		-0.5	-0.05	-0.32	0.22	0.16	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.12	-0.14	0.37	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.06	-0.19	0.36	0.18
		1	0	0	0	0	0	0	0	0	0.08	-0.22	0.33	0.17
	40	-1	-0.29	-0.5	-0.14	0.11	0	0	0	0	0	0	0	0
		-0.5	-0.02	-0.22	0.16	0.12	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.13	-0.06	0.31	0.12	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.02	-0.17	0.19	0.11
		1	0	0	0	0	0	0	0	0	0.04	-0.13	0.21	0.11
90	20	-1	-0.33	-0.55	-0.05	0.15	0	0	0	0	0	0	0	0
		-0.5	-0.06	-0.29	0.18	0.14	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.11	-0.12	0.38	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.04	-0.22	0.25	0.13
		1	0	0	0	0	0	0	0	0	0.06	-0.2	0.37	0.17
	40	-1	-0.28	-0.45	-0.09	0.12	0	0	0	0	0	0	0	0
		-0.5	-0.01	-0.18	0.16	0.1	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.13	-0.08	0.34	0.12	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.03	-0.14	0.22	0.11
		1	0	0	0	0	0	0	0	0	0.04	-0.16	0.22	0.12

Appendix A.20: Bias and Error of group estimates: Mixture Proportion 1 (MP2) and Cluster Effect 2 (CE2) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.33	-0.55	-0.05	0.15	0	0	0	0	0	0	0	0
		-0.5	-0.06	-0.29	0.18	0.14	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.11	-0.12	0.38	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.04	-0.22	0.25	0.13
		1	0	0	0	0	0	0	0	0	0.06	-0.2	0.37	0.17
	40	-1	-0.28	-0.45	-0.09	0.12	0	0	0	0	0	0	0	0
		-0.5	-0.01	-0.18	0.16	0.1	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.13	-0.08	0.34	0.12	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.03	-0.14	0.22	0.11
		1	0	0	0	0	0	0	0	0	0.04	-0.16	0.22	0.12
60	20	-1	-0.62	-0.84	-0.38	0.14	0	0	0	0	0	0	0	0
		-0.5	-0.21	-0.48	0.05	0.16	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.1	-0.11	0.3	0.13	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.21	0	0.46	0.15
		1	0	0	0	0	0	0	0	0	0.43	0.17	0.68	0.15
	40	-1	-0.57	-0.73	-0.43	0.1	0	0	0	0	0	0	0	0
		-0.5	-0.15	-0.33	0.02	0.11	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.13	-0.05	0.3	0.1	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.17	-0.01	0.37	0.12
		1	0	0	0	0	0	0	0	0	0.32	0.15	0.5	0.11
90	20	-1	-0.6	-0.87	-0.3	0.17	0	0	0	0	0	0	0	0
		-0.5	-0.23	-0.52	0.01	0.16	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.1	-0.17	0.35	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.24	-0.02	0.46	0.14
		1	0	0	0	0	0	0	0	0	0.43	0.21	0.64	0.14
	40	-1	-0.57	-0.77	-0.34	0.12	0	0	0	0	0	0	0	0
		-0.5	-0.17	-0.36	0.01	0.11	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.13	-0.03	0.27	0.1	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.18	0.02	0.37	0.11
		1	0	0	0	0	0	0	0	0	0.35	0.18	0.52	0.11

Appendix A.21: Bias and Error of group estimates: Mixture Proportion 1 (MP3) and Cluster Effect 2 (CE2) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.6	-0.87	-0.3	0.17	0	0	0	0	0	0	0	0
		-0.5	-0.23	-0.52	0.01	0.16	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.1	-0.17	0.35	0.16	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.24	-0.02	0.46	0.14
		1	0	0	0	0	0	0	0	0	0.43	0.21	0.64	0.14
	40	-1	-0.57	-0.77	-0.34	0.12	0	0	0	0	0	0	0	0
		-0.5	-0.17	-0.36	0.01	0.11	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.13	-0.03	0.27	0.1	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.18	0.02	0.37	0.11
		1	0	0	0	0	0	0	0	0	0.35	0.18	0.52	0.11
60	20	-1	-0.63	-0.9	-0.42	0.14	0	0	0	0	0	0	0	0
		-0.5	-0.22	-0.44	0.06	0.15	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.11	-0.1	0.36	0.14	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.22	-0.04	0.46	0.14
		1	0	0	0	0	0	0	0	0	0.42	0.17	0.67	0.15
	40	-1	-0.57	-0.75	-0.4	0.11	0	0	0	0	0	0	0	0
		-0.5	-0.17	-0.35	-0.01	0.11	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.11	-0.08	0.32	0.11	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.19	0.03	0.36	0.1
		1	0	0	0	0	0	0	0	0	0.34	0.15	0.55	0.12
90	20	-1	-0.63	-0.85	-0.4	0.14	0	0	0	0	0	0	0	0
		-0.5	-0.32	-0.48	-0.09	0.12	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	-0.2	0.19	0.12	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.32	0.06	0.56	0.14
		1	0	0	0	0	0	0	0	0	0.64	0.43	0.84	0.13
	40	-1	-0.59	-0.78	-0.43	0.11	0	0	0	0	0	0	0	0
		-0.5	-0.28	-0.44	-0.1	0.09	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.03	-0.19	0.14	0.1	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.3	0.14	0.46	0.1
		1	0	0	0	0	0	0	0	0	0.57	0.41	0.78	0.11

Appendix A.22: Bias and Error of group estimates: Mixture Proportion 1 (MP4) and Cluster Effect 2 (CE2) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	-0.63	-0.85	-0.4	0.14	0	0	0	0	0	0	0	0
		-0.5	-0.32	-0.48	-0.09	0.12	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	-0.2	0.19	0.12	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.32	0.06	0.56	0.14
		1	0	0	0	0	0	0	0	0	0.64	0.43	0.84	0.13
	40	-1	-0.59	-0.78	-0.43	0.11	0	0	0	0	0	0	0	0
		-0.5	-0.28	-0.44	-0.1	0.09	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.03	-0.19	0.14	0.1	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.3	0.14	0.46	0.1
		1	0	0	0	0	0	0	0	0	0.57	0.41	0.78	0.11
60	20	-1	-0.62	-0.8	-0.37	0.13	0	0	0	0	0	0	0	0
		-0.5	-0.33	-0.57	-0.1	0.14	0	0	0	0	0	0	0	0
		0	0	0	0	0	-0.01	-0.25	0.23	0.14	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.3	0.1	0.53	0.13
		1	0	0	0	0	0	0	0	0	0.61	0.41	0.82	0.13
	40	-1	-0.58	-0.74	-0.42	0.1	0	0	0	0	0	0	0	0
		-0.5	-0.27	-0.46	-0.07	0.11	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.01	-0.16	0.16	0.1	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.29	0.14	0.47	0.1
		1	0	0	0	0	0	0	0	0	0.56	0.34	0.74	0.12
90	20	-1	-0.62	-0.85	-0.38	0.14	0	0	0	0	0	0	0	0
		-0.5	-0.33	-0.54	-0.11	0.13	0	0	0	0	0	0	0	0
		0	0	0	0	0	0.04	-0.19	0.24	0.14	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.34	0.19	0.49	0.1
		1	0	0	0	0	0	0	0	0	0.63	0.44	0.83	0.12
	40	-1	-0.59	-0.75	-0.42	0.1	0	0	0	0	0	0	0	0
		-0.5	-0.28	-0.44	-0.06	0.11	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	-0.15	0.15	0.1	0	0	0	0
		0.5	0	0	0	0	0	0	0	0	0.29	0.1	0.48	0.11
		1	0	0	0	0	0	0	0	0	0.57	0.4	0.74	0.1

Appendix A.23: Bias and Error of group estimates: Mixture Proportion 1 (MP1) and Cluster Effect 3 (CE2) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	0	0	0	0	0	0	0	0	-0.37	-0.59	-0.13	0.14
		-0.5	0	0	0	0	0	0	0	0	-0.27	-0.45	-0.08	0.12
		0	0	0	0	0	-0.08	-0.34	0.13	0.14	0	0	0	0
		0.5	0.25	0.04	0.49	0.14	0	0	0	0	0	0	0	0
		1	0.55	0.28	0.8	0.15	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.32	-0.49	-0.16	0.11
		-0.5	0	0	0	0	0	0	0	0	-0.24	-0.42	-0.04	0.11
		0	0	0	0	0	-0.11	-0.28	0.07	0.1	0	0	0	0
		0.5	0.25	0.09	0.46	0.1	0	0	0	0	0	0	0	0
		1	0.52	0.36	0.68	0.1	0	0	0	0	0	0	0	0
60	20	-1	0	0	0	0	0	0	0	0	-0.36	-0.56	-0.13	0.13
		-0.5	0	0	0	0	0	0	0	0	-0.28	-0.49	-0.07	0.13
		0	0	0	0	0	-0.08	-0.3	0.13	0.13	0	0	0	0
		0.5	0.28	0.06	0.51	0.14	0	0	0	0	0	0	0	0
		1	0.55	0.34	0.78	0.14	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.31	-0.5	-0.13	0.12
		-0.5	0	0	0	0	0	0	0	0	-0.25	-0.46	-0.08	0.12
		0	0	0	0	0	-0.12	-0.28	0.06	0.12	0	0	0	0
		0.5	0.26	0.06	0.44	0.11	0	0	0	0	0	0	0	0
		1	0.53	0.39	0.69	0.1	0	0	0	0	0	0	0	0
90	20	-1	0	0	0	0	0	0	0	0	-0.38	-0.67	-0.08	0.15
		-0.5	0	0	0	0	0	0	0	0	-0.25	-0.45	0.01	0.14
		0	0	0	0	0	-0.12	-0.39	0.18	0.15	0	0	0	0
		0.5	0.25	0.03	0.5	0.14	0	0	0	0	0	0	0	0
		1	0.56	0.28	0.79	0.15	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.32	-0.53	-0.11	0.13
		-0.5	0	0	0	0	0	0	0	0	-0.25	-0.46	-0.03	0.12
		0	0	0	0	0	-0.12	-0.3	0.06	0.11	0	0	0	0
		0.5	0.27	0.06	0.46	0.12	0	0	0	0	0	0	0	0
		1	0.55	0.38	0.72	0.12	0	0	0	0	0	0	0	0

Appendix A.24: Bias and Error of group estimates: Mixture Proportion 1 (MP2) and Cluster Effect 3 (CE2) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	0	0	0	0	0	0	0	0	-0.38	-0.67	-0.08	0.15
		-0.5	0	0	0	0	0	0	0	0	-0.25	-0.45	0.01	0.14
		0	0	0	0	0	-0.12	-0.39	0.18	0.15	0	0	0	0
		0.5	0.25	0.03	0.5	0.14	0	0	0	0	0	0	0	0
		1	0.56	0.28	0.79	0.15	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.32	-0.53	-0.11	0.13
		-0.5	0	0	0	0	0	0	0	0	-0.25	-0.46	-0.03	0.12
		0	0	0	0	0	-0.12	-0.3	0.06	0.11	0	0	0	0
		0.5	0.27	0.06	0.46	0.12	0	0	0	0	0	0	0	0
		1	0.55	0.38	0.72	0.12	0	0	0	0	0	0	0	0
60	20	-1	0	0	0	0	0	0	0	0	-0.65	-0.82	-0.45	0.12
		-0.5	0	0	0	0	0	0	0	0	-0.41	-0.59	-0.25	0.12
		0	0	0	0	0	-0.09	-0.28	0.11	0.12	0	0	0	0
		0.5	0.4	0.2	0.61	0.12	0	0	0	0	0	0	0	0
		1	0.81	0.64	0.99	0.11	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.58	-0.74	-0.42	0.1
		-0.5	0	0	0	0	0	0	0	0	-0.39	-0.55	-0.23	0.1
		0	0	0	0	0	-0.1	-0.23	0.02	0.08	0	0	0	0
		0.5	0.4	0.18	0.57	0.11	0	0	0	0	0	0	0	0
		1	0.78	0.6	0.94	0.09	0	0	0	0	0	0	0	0
90	20	-1	0	0	0	0	0	0	0	0	-0.66	-0.85	-0.46	0.12
		-0.5	0	0	0	0	0	0	0	0	-0.41	-0.58	-0.21	0.11
		0	0	0	0	0	-0.07	-0.26	0.12	0.11	0	0	0	0
		0.5	0.4	0.2	0.59	0.12	0	0	0	0	0	0	0	0
		1	0.83	0.63	1.03	0.13	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.59	-0.75	-0.44	0.09
		-0.5	0	0	0	0	0	0	0	0	-0.39	-0.52	-0.26	0.08
		0	0	0	0	0	-0.09	-0.24	0.08	0.1	0	0	0	0
		0.5	0.39	0.24	0.53	0.09	0	0	0	0	0	0	0	0
		1	0.8	0.64	0.93	0.09	0	0	0	0	0	0	0	0

Appendix A.25: Bias and Error of group estimates: Mixture Proportion 1 (MP3) and Cluster Effect 3 (CE2) for Misspecified model

Cluster Number	Cluster Size	Cluster Effect	Cluster Type 1 Bias			Cluster Type 1 Error	Cluster Type 2 Bias			Cluster Type 2 Error	Cluster Type 3 Bias			Cluster Type 3 Error
			Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	0	0	0	0	0	0	0	0	-0.66	-0.85	-0.46	0.12
		-0.5	0	0	0	0	0	0	0	0	-0.41	-0.58	-0.21	0.11
		0	0	0	0	0	-0.07	-0.26	0.12	0.11	0	0	0	0
		0.5	0.4	0.2	0.59	0.12	0	0	0	0	0	0	0	0
		1	0.83	0.63	1.03	0.13	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.59	-0.75	-0.44	0.09
		-0.5	0	0	0	0	0	0	0	0	-0.39	-0.52	-0.26	0.08
		0	0	0	0	0	-0.09	-0.24	0.08	0.1	0	0	0	0
		0.5	0.39	0.24	0.53	0.09	0	0	0	0	0	0	0	0
		1	0.8	0.64	0.93	0.09	0	0	0	0	0	0	0	0
60	20	-1	0	0	0	0	0	0	0	0	-0.67	-0.84	-0.51	0.11
		-0.5	0	0	0	0	0	0	0	0	-0.4	-0.59	-0.19	0.12
		0	0	0	0	0	-0.06	-0.25	0.15	0.11	0	0	0	0
		0.5	0.43	0.25	0.59	0.1	0	0	0	0	0	0	0	0
		1	0.83	0.64	1.04	0.12	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.61	-0.76	-0.45	0.1
		-0.5	0	0	0	0	0	0	0	0	-0.4	-0.54	-0.24	0.09
		0	0	0	0	0	-0.1	-0.26	0.07	0.09	0	0	0	0
		0.5	0.38	0.24	0.51	0.09	0	0	0	0	0	0	0	0
		1	0.8	0.65	0.98	0.1	0	0	0	0	0	0	0	0
90	20	-1	0	0	0	0	0	0	0	0	-0.64	-0.84	-0.42	0.14
		-0.5	0	0	0	0	0	0	0	0	-0.31	-0.51	-0.09	0.13
		0	0	0	0	0	0.01	-0.24	0.2	0.12	0	0	0	0
		0.5	0.31	0.12	0.51	0.12	0	0	0	0	0	0	0	0
		1	0.65	0.44	0.91	0.14	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.57	-0.73	-0.39	0.11
		-0.5	0	0	0	0	0	0	0	0	-0.3	-0.48	-0.14	0.1
		0	0	0	0	0	0.02	-0.13	0.18	0.09	0	0	0	0
		0.5	0.27	0.08	0.45	0.11	0	0	0	0	0	0	0	0
		1	0.57	0.4	0.72	0.11	0	0	0	0	0	0	0	0

Appendix A.26: Bias and Error of group estimates: Mixture Proportion 1 (MP4) and Cluster Effect 3 (CE2) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Cluster Number	Cluster Size	Cluster Effect	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
30	20	-1	0	0	0	0	0	0	0	0	-0.64	-0.84	-0.42	0.14
		-0.5	0	0	0	0	0	0	0	0	-0.31	-0.51	-0.09	0.13
		0	0	0	0	0	0.01	-0.24	0.2	0.12	0	0	0	0
		0.5	0.31	0.12	0.51	0.12	0	0	0	0	0	0	0	0
		1	0.65	0.44	0.91	0.14	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.57	-0.73	-0.39	0.11
		-0.5	0	0	0	0	0	0	0	0	-0.3	-0.48	-0.14	0.1
		0	0	0	0	0	0.02	-0.13	0.18	0.09	0	0	0	0
		0.5	0.27	0.08	0.45	0.11	0	0	0	0	0	0	0	0
		1	0.57	0.4	0.72	0.11	0	0	0	0	0	0	0	0
60	20	-1	0	0	0	0	0	0	0	0	-0.63	-0.88	-0.42	0.13
		-0.5	0	0	0	0	0	0	0	0	-0.33	-0.54	-0.1	0.14
		0	0	0	0	0	0.02	-0.18	0.28	0.14	0	0	0	0
		0.5	0.33	0.11	0.54	0.13	0	0	0	0	0	0	0	0
		1	0.67	0.47	0.89	0.13	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.58	-0.8	-0.41	0.11
		-0.5	0	0	0	0	0	0	0	0	-0.29	-0.45	-0.12	0.1
		0	0	0	0	0	0	-0.13	0.16	0.09	0	0	0	0
		0.5	0.3	0.15	0.48	0.11	0	0	0	0	0	0	0	0
		1	0.56	0.4	0.7	0.09	0	0	0	0	0	0	0	0
90	20	-1	0	0	0	0	0	0	0	0	-0.65	-0.88	-0.41	0.14
		-0.5	0	0	0	0	0	0	0	0	-0.32	-0.55	-0.11	0.14
		0	0	0	0	0	-0.02	-0.24	0.19	0.13	0	0	0	0
		0.5	0.31	0.15	0.51	0.11	0	0	0	0	0	0	0	0
		1	0.64	0.41	0.84	0.12	0	0	0	0	0	0	0	0
	40	-1	0	0	0	0	0	0	0	0	-0.58	-0.75	-0.4	0.1
		-0.5	0	0	0	0	0	0	0	0	-0.29	-0.46	-0.09	0.11
		0	0	0	0	0	0	-0.19	0.15	0.11	0	0	0	0
		0.5	0.28	0.13	0.46	0.1	0	0	0	0	0	0	0	0
		1	0.58	0.4	0.75	0.1	0	0	0	0	0	0	0	0

Appendix A.27: Bias and Error of group estimates: Cluster Effect 4 (CE4) for Misspecified model

			Cluster Type										
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Mixture Proportion	Cluster Number	Cluster Size	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
1	30	20	0.09	-0.01	0.21	0.22	0.02	-0.09	0.14	0.36	-0.08	-0.21	0.05	0.44
		40	0.09	-0.03	0.2	0.23	-0.01	-0.13	0.12	0.39	-0.12	-0.25	0.03	0.47
	60	20	0.09	0.01	0.17	0.26	0.02	-0.07	0.09	0.27	-0.08	-0.17	0.01	0.44
		40	0.11	0.04	0.18	0.23	0.01	-0.08	0.08	0.38	-0.09	-0.18	-0.01	0.47
	90	20	0.08	0.02	0.15	0.25	0	-0.07	0.07	0.39	-0.08	-0.17	0.01	0.41
		40	0.1	0.03	0.18	0.26	0	-0.06	0.08	0.4	-0.11	-0.18	-0.02	0.41
2	30	20	0.05	-0.03	0.13	0.15	0	-0.11	0.11	0.21	-0.06	-0.21	0.06	0.3
		40	0.09	0	0.17	0.15	0	-0.1	0.09	0.24	-0.09	-0.21	0.01	0.35
	60	20	0.06	0	0.14	0.13	0	-0.09	0.09	0.21	-0.06	-0.15	0.02	0.28
		40	0.08	0.02	0.13	0.15	0	-0.06	0.06	0.22	-0.09	-0.16	-0.02	0.3
	90	20	0.06	0.02	0.1	0.14	0	-0.06	0.05	0.2	-0.07	-0.13	0	0.3
		40	0.09	0.04	0.14	0.15	0	-0.05	0.05	0.23	-0.09	-0.16	-0.03	0.32
3	30	20	0	-0.09	0.08	0.18	0.01	-0.08	0.1	0.2	0	-0.11	0.09	0.19
		40	0	-0.09	0.09	0.22	0	-0.08	0.08	0.22	0	-0.09	0.09	0.22
	60	20	0	-0.06	0.07	0.2	0	-0.06	0.07	0.19	0.01	-0.06	0.08	0.21
		40	0	-0.07	0.07	0.23	0	-0.07	0.06	0.19	-0.01	-0.07	0.06	0.2
	90	20	0	-0.06	0.04	0.17	0	-0.05	0.06	0.18	0	-0.07	0.05	0.17
		40	0	-0.04	0.04	0.23	0	-0.07	0.06	0.2	0	-0.05	0.05	0.25
4	30	20	0	-0.15	0.12	0.34	-0.01	-0.15	0.11	0.31	0.01	-0.11	0.12	0.34
		40	-0.01	-0.14	0.09	0.4	-0.01	-0.14	0.11	0.37	-0.01	-0.15	0.12	0.39
	60	20	-0.01	-0.1	0.09	0.36	-0.01	-0.1	0.1	0.32	0	-0.1	0.11	0.38
		40	0.01	-0.08	0.1	0.33	0.01	-0.09	0.1	0.35	0	-0.07	0.1	0.33
	90	20	0	-0.07	0.08	0.32	0	-0.07	0.08	0.33	0	-0.08	0.08	0.35
		40	0	-0.08	0.07	0.36	0	-0.09	0.06	0.37	0	-0.07	0.07	0.32

Appendix A.28: Bias and Error of group estimates: Cluster Effect 5 (CE5) for Misspecified model

			Cluster Type											
			1				2				3			
			Bias			Error	Bias			Error	Bias			Error
Mixture Proportion	Cluster Number	Cluster Size	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean	Mean	2.5%	97.5%	Mean
1	30	20	0.09	-0.14	0.39	0.52	-0.01	-0.23	0.3	0.6	-0.12	-0.34	0.2	1.01
		40	0.11	-0.12	0.4	0.47	0.01	-0.23	0.3	0.63	-0.1	-0.34	0.18	1.04
	60	20	0.09	-0.05	0.2	0.48	0	-0.14	0.13	0.78	-0.1	-0.26	0.02	0.98
		40	0.11	-0.07	0.32	0.53	0	-0.18	0.19	0.74	-0.11	-0.29	0.08	1.05
	90	20	0.1	-0.04	0.21	0.45	0.01	-0.13	0.14	0.71	-0.08	-0.24	0.04	0.96
		40	0.11	-0.01	0.23	0.44	0	-0.13	0.15	0.73	-0.1	-0.26	0.05	1.02
2	30	20	0.09	-0.06	0.24	0.27	0.03	-0.14	0.2	0.44	-0.05	-0.24	0.17	0.62
		40	0.09	-0.05	0.25	0.24	0	-0.15	0.17	0.5	-0.1	-0.3	0.08	0.68
	60	20	0.09	-0.02	0.18	0.27	0	-0.12	0.11	0.45	-0.09	-0.22	0.06	0.62
		40	0.09	-0.04	0.2	0.23	-0.01	-0.14	0.11	0.53	-0.11	-0.25	0.01	0.68
	90	20	0.08	0	0.17	0.29	0	-0.11	0.1	0.5	-0.09	-0.2	0.02	0.67
		40	0.09	0.01	0.16	0.25	0	-0.08	0.07	0.46	-0.11	-0.21	-0.03	0.66
3	30	20	0	-0.14	0.15	0.4	-0.01	-0.17	0.14	0.38	-0.01	-0.16	0.15	0.39
		40	0	-0.13	0.12	0.48	0	-0.12	0.16	0.46	0	-0.12	0.16	0.51
	60	20	-0.01	-0.12	0.13	0.42	0	-0.13	0.13	0.48	0	-0.12	0.13	0.45
		40	0	-0.12	0.1	0.4	0	-0.11	0.12	0.5	0	-0.11	0.12	0.4
	90	20	0	-0.09	0.09	0.44	0.01	-0.09	0.1	0.45	0.01	-0.1	0.1	0.41
		40	0	-0.08	0.1	0.44	-0.01	-0.08	0.09	0.51	-0.01	-0.08	0.08	0.47
4	30	20	0	-0.26	0.25	0.69	0	-0.2	0.26	0.68	0	-0.23	0.24	0.69
		40	0.01	-0.23	0.23	0.68	0.02	-0.22	0.24	0.67	0.02	-0.24	0.23	0.82
	60	20	0	-0.19	0.16	0.68	0	-0.21	0.18	0.63	0	-0.2	0.19	0.69
		40	0	-0.16	0.14	0.8	0	-0.14	0.15	0.67	-0.01	-0.14	0.14	0.63
	90	20	0	-0.13	0.12	0.73	0	-0.13	0.16	0.8	0	-0.13	0.14	0.7
		40	0	-0.14	0.13	0.77	0	-0.13	0.13	0.74	0	-0.14	0.13	0.72

References

Anderson, D. R. (2008). *Model based inference in the life sciences: A primer on evidence*. Springer: New York, NY.

Asparouhov, T. & Muthén, B. (2008). Multilevel mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 27-51). Charlotte, NC: Information Age Publishing, Inc.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis*, London: Arnold.

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

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. New York, NY: John Wiley.

Boscardin, C., Muthén, B., Francis, D., & Baker, E. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100,* 192-208.

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

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research, 33*, 261-304.

Burstein, L., Linn, R. L., & Capell, F.J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics, 3*, 547-385.

Chudowsky, N., Chudowsky, V., & Kober, N. (2007). *Answering the question that matters most: Has student achievement increased since No Child Left Behind?* Washington DC: Center on Educational Policy.

Clogg, C.C., & Goodman, L.A. (1985). Simultaneous latent structural analysis in several groups. In N. B. Tuma (Ed.), *Sociological Methodology* (pp. 81-110). San Francisco: Jossey-Bass Publishers.

Croudace, T. J., Jarvelin, M. R., Wadsworth, M. E., & Jones, P. B. (2003). Developmental typology of trajectories to nighttime bladder control: Epidemiologic application of longitudinal latent class analysis. *American Journal of Epidemiology, 157*, 834-842.

Dayton, C. M (1991). Educational applications of latent class analysis. *Measurement and Evaluation in Counseling and Development*, *24*, 131-141.

Doran, H. C., & Lockwood, J. R. (2006). Fitting value-added models in R. *Journal of Educational and Behavioral Statistics, 31*, 205-230.

Duncan, T.E., Duncan, S.C., & Strycker, L.A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications (2nd ed.).* Mahwah, NJ: Lawrence Erlbaum Associates.

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

Feldman, B. J., Masyn, K. E.. & Conger, R. D. (2009). New approaches to studying

   problem behaviors: A comparison of methods for modeling longitudinal,

   categorical adolescent drinking data. *Developmental Psychology, 45*, 3, 652-676.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and

   unidentifiable models. *Biometrika, 61*, 215-231.

Hancock, G. R., & Lawrence, F. R. (2006). Using latent growth models to evaluate

   longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural

   equation modeling: A second course* (pp. 171-196). Greenwood, CT: Information

   Age Publishing, Inc.

Henry, K., & Muthén, B. (2010). Multilevel latent class analysis: An application of

   adolescent smoking typologies with individual and contextual predictors.

   Structural Equation Modeling, 17, 193-215.

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model

   specifications. *Journal of Educational and Behavioral Statistics, 27*, 385-409.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London, UK:

   Sage Publications.

Kreft, I. G. G., de Leeuw, J, & Aiken, L. (1995). The effect of different forms of

   centering in hierarchical linear models. *Multivariate Behavioral Research, 30,* 1-

   22.

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.

Kreuter, F., Yan, T., & Tourangeau, R. (2008). Good item or bad – can latent class

analysis tell?: The utility of latent class analysis for the evaluation of survey

questions. *Journal of the Royal Statistical Society*, *Series A*, *171*, 723-738.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure

analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A.

Star, & J. A. Clausen (Eds.), *Measurement and prediction*. Princeton: Princeton

University Press.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton

Mifflin.

Lazarus, S. S., Wu, Y., Altman, J., & Thurlow, M. L. (2010). The characteristics of low

performing students on large-scale assessments. *NCEO brief*. Minneapolis:

National Center on Educational Outcomes, University of Minnesota.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal

of the Royal Statistical Society*, *Series B, 34,* 1-41.

Lockwood, J. R., & McCaffrey, D. F. (2007). Controlling for individual heterogeneity in

longitudinal models, with applications to student achievement. *Electronic Journal

of Statistics, 1*, 223-252.

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous

factors: Resolution by maximum likelihood? *Multivariate Behavioral Research,

41*, 499-532.

Marsh, H. W., Ludtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., &

Nagengast, B. (2009). Doubly-latent models of school contextual effects:

Integrating multilevel and structural equation approaches to control measurement
and sampling errors. *Multivariate Behavioral Research, 44*, 764-802.

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

McQuarrie, A. D .R., & Tsai, C. L. (1998). *Regression and time series model selection*.
World Scientific, London, UK.

Miles, J., & Shevlin, M. (2000). *Applying regression and correlation: A guide for
students and researchers.* Thousand Oaks, CA: Sage.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations.
Psychometrika, 54, 557-585.

Muthén, B. (1991). Analysis of longitudinal data using latent variable models with
varying parameters. In L. Collins & J. Horn (eds.), *Best methods for the analysis
of change. Recent advances, unanswered questions, future directions* (pp. 1-17).
Washington DC: American Psychological Association.

Muthén, B. (2000). Methodological issues in random coefficient growth modeling using a
latent variable framework: Applications to the development of heavy drinking. In
J. Rose, L. Chassin, C. Presson & J. Sherman (Eds.). *Multivariate applications in
substance use research* (pp. 113-140). Hillsdale, NJ: Erlbaum.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E.
Schumacker (Eds.), *New developments and techniques in structural equation
modeling* (pp. 1-33). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related
techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative*

methodology for the social sciences (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, B. (2006). The potential of growth mixture modeling. *Infant and Child Development, 15*, 623-625.

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

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S.T ., Yang, C. C., Wang, C. P., Kellam, S., Carlin, J. & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3*, 459-475.

Muthén, B., Khoo, S.T., Francis, D. & Kim Boscardin, C. (2003). Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. In S. R. Reise & N. Duan (Eds). *Multilevel Modeling: Methodological Advances, Issues, and Applications* (pp.71-89). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B. O., & Muthén, L. K. (2000). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U.S. national sample. *Journal of Studies on Alcohol, 61*, 290-300.

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

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide (V6.1).* Los Angeles: Muthén & Muthén.

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

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology, 31*, 327-362.

Nezlek, J. B., & Zyzniewski, L. E. (1998). Using hierarchical linear modeling to analyze group data. *Group Dynamics: Theory, Research, and Practice, 2*, 313-320.

No Child Left Behind Act of 2001, 20 U.S.C. § 6161.

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

Pollack, B. N. (1998). Hierarchical linear modeling and the "unit of analysis" problem: A solution for analyzing responses of intact group members. *Group Dynamics, 2*, 299-312.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth models. Quantitative applications in the social sciences*, Thousand Oaks, CA: Sage.

Preacher, K., Zyphur, M. & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209-233.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association, 53*, 873-880.

Quandt, R. E., & Ramsey J. B. (1972). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association, 73*, 730-752.

Raudenbush, S. W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.).* Newbury Park, CA: Sage Publications.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville, TN: University of Tennessee Value-Added Research Center.

SAS Institute. (2008-2010). SAS, release 9.2 [Computer software]. Cary, NC: Author.

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

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

Singer, J. D. (1999). Using SAS Proc Mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *23*, 323-355.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal and structural equation models*. London, England: Chapman & Hall/CRC.

Springer, M.G., Ballou, D., Hamilton, L., Le, V., Lockwood, J.R., McCaffrey, D., Pepper, M., & Stecher, B. (2010). *Teacher pay for performance: Experimental evidence from the project on incentives in teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University

155

Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1985). *Statistical analysis of finite mixture distributions.* Chichester, U.K.: John Wiley & Sons.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in a growth mixture model. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Greenwich, CT: Information Age.

U.S. Department of Education (2006). Letter to Chief State School Officers: Assessment Requirements of NCLB and the Growth Model Pilot Project. http://www2.ed.gov/policy/elsec/guid/secletter/060221.html

U.S. Department of Education (2010). Race to the Top Program Executive Summary. http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association, 91*, 217-221.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J.A. Hagenaars & A.L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-106). Cambridge, UK: Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2008). *Latent GOLD 4.5 user's manual*. Belmont, MA: Statistical Innovations.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67.

Wright, S. P., White, J. T., Snaders, W. L., & Rivers, J. C. (2010). *SAS EVAAS statistical models.* SAS Institute Inc. http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf