A LATENT STATE TRAIT MODEL FOR MULTILEVEL MEDIATION ANALYSIS WITH MULTIPLE TIMEPOINTS

By

Lydia Bradford

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Measurement and Quantitative methods – Doctor of Philosophy

2024

ABSTRACT

In randomized control trials (RCT), the recent focus has shifted to *how* an intervention yields positive results on its intended outcome. This aligns with the recent push of implementation science in healthcare (Bauer et al., 2015) but goes beyond this. RCTs have moved to evaluating the theoretical framing of the intervention as well as differing implementation effects. One example of a typical mediation design in education research is the 3-2-1 mediation design (Pituch et al., 2009) with random assignment at the school level, the mediator at the teacher level, and the outcome at the student level. In such situations, it is not uncommon for the mediator to be measured at multiple time points across the intervention period. However, the current mediation models are not equipped for a longitudinal mediator that does not measure growth and where the outcome measure is only measured once in a 3-2-1 model. This dissertation has three primary goals. The first is to provide a framework to answer research questions, such as the mediating effect of teacher practices on an intervention's impact on student achievement. In this situation, the mediator is measured at multiple time points in a 3-2-1 design, which current methods are not equipped to answer. The second goal is to provide potential estimation methods for the framework and to evaluate their bias and power. The final goal is to understand how these estimation methods perform in an actual study.

Again, this study combines multilevel mediation with a latent state-trait (LST) framework to provide a model that can answer mediation questions when the mediator is at level 2 (teacher level) in a 3-level design and is measured at multiple time points. It also provides four different estimation methods (averages of the summed mediator, averages of factor scores, factor scores from the LST model, and the fully specified model) and the assumptions required for the four methods. These assumptions include assumptions on restrictions of the latent structure, measurement error, and the presence of state vs trait variances. It then uses a simulation study to evaluate the four different methods under varying design conditions: sample size, factor loadings, and effect sizes. Finally, this study investigates these methods in a project-based learning (PBL) science intervention study (Crafting Engaging Science Interventions [CESE]; Schneider et al., 2022).

The results of the simulation study show that the choice of measure for the mediator is critical in reducing bias and increasing power in the estimation of the multilevel LST mediation model. Mediators with low construct validity will lead to bias across estimation methods. These might be mediating measures that are not truly measuring the mediator, have small factor loadings, or otherwise, the variance in the mediator is not explained by the proposed underlying factors. Additionally, mediators with more time-specific variance than trait-specific variance also lead to more bias across the estimation methods. These are situations where the time point explains more variance than the level 2 (teachers) general trait. If the time points are teacher practices in a given class period, this would be the situation where the teacher's practices vary widely from day to day and not from teacher to teacher. The simulation also indicates that the sample sizes required for such research questions are large (>200).

Following the simulation study, the methods were evaluated in the CESE study, a cluster randomized control trial with 61 schools, 102 teachers, and 4238 students. The CESE intervention included professional learning for the teachers, 3 NGSS-aligned units (in either chemistry or physics) with driving questions and hands-on experiences for the students, and NGSS-aligned end-of-unit assessments. During the intervention, as part of data collection, a random sample of teachers was observed 1 to 5 times, and their PBL practices in the classroom were scored. Mediation in this study aimed to understand how the intervention affected teacher PBL practices in the classroom and how those practices directly affected student science achievement at the end of the study. The results of the estimations of these mediation effects indicate that the models can converge and provide results; however, this empirical study reiterates the findings from the simulation study of issues with small sample sizes and low trait-specific variances. Investigating these mediation effects for the CESE intervention also raises several additional design considerations for mediation research questions, such as the effect of using raters, confounders, and the design of the mediating measure (in this case, the observation protocol).

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor and committee chair, Dr. Barbara Schneider. She has always been my biggest supporter, constantly pushing me to be my best. I would not be where I am without her. She has guided me and taught me so much during my PhD. I am prepared to advance in my career because of her. I also want to thank her more specifically for her guidance on this dissertation and all the time and energy she has put into advising me on it.

I also want to thank my committee members, Dr. Spyros Konstatopoulos, Dr. Kenneth Frank, and Dr. Joseph Krajcik. Each of them has guided and advised me in different ways during my time as a doctoral student and helped to make me the academic I am today. I also want to thank them for their advice on this specific study.

I also owe a lot of thanks to my partner in crime in the measurement and quantitative methods program, Kayla Bartz. I want to thank her for all our conversations and discussions about methods, measurement, the study, and any other musings. I also want to thank the rest of the team in the office of the Hannah Chair for always being my biggest supporters and encouraging me to put my best into this study and my work in the program. I also need to thank all my friends in the MQM program, Jiachen Liu, Shimeng Dai, and Jiawei Li, for all the critical conversations and discussions that would eventually help to conceptualize this study.

Finally, I would like to thank my friends and family for all their support: to Colin DeWitt for being a rock during the writing of this dissertation; to my mom, dad, and all my siblings for bringing joy and surprises and support to me whenever I need it; to my salsa community in East Lansing who always cheer me on; and finally, to my church and choir community who always seem to be able to remember my research topic. I would not have been able to finish without the support of all those around me.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

RCT	Randomized Control Trial
LST	Latent State-Trait
CESE	Crafting Engaging Science Environments

CHAPTER 1

INTRODUCTION

Over the past 20 years, governmental agencies have increasingly required robust mediation analysis in randomized control trials (RCT) education research to understand the pathway by which the intervention is implemented and the magnitude of that effect on the outcome. Within the context of RCTs, mediation analysis explores how the intervention works, which can be driven by components of the intervention (the implementation of the intervention) or an underlying theoretical mediational pathway of an intervention. The effects of mediation are investigated across many fields, including health, psychology, and education. One example in healthcare is an investigation into factors that mediate the impact of cognitive functional therapy (CFT) on chronic low back pain (O'Neill et al., 2020). This study randomized participants into CFT vs. group exercise and education. It measured the mediators (pain self-efficacy, stress, fear of physical activity, coping, depression, and anxiety) at the halfway point for the intervention (6 months) and then measured the outcomes (disability and back pain) at the end (12 months). Another example in psychology is an investigation on the mediating effects of mindfulness-based self-efficacy on the treatment effect of meditation or exercise and stress (Goldstein et al., 2020). Here, the study randomized individuals into mindfulness training, exercise training, or the control group. Randomization occurred before the intervention; some of the mediators (mindfulness and self-reported physical activity) were measured at 4 months, while others (mindfulness self-efficacy and exercise self-efficacy) were measured at 6 months, and the outcome (health and mental health) was measured after 8 months. Finally, within education, an example of a mediation study is Herman et al. (2022)'s investigation into whether student time-on-task mediates the effects of the CHAMPS classroom management intervention on student outcomes (student social behavior and academics, standardized academic achievement, and classroom and homework completion). This study theorized that teachers' use of CHAMPS would increase student time on task, leading to the intended behavior and academic outcomes. In this investigation, baseline measures were collected prior to randomization. In contrast, the mediator (student time-on-task) was measured at the baseline and then at the end of the year, and the outcomes were collected at the baseline and end of the intervention. Then, one (standardized academic achievement) was collected as a follow-up the following year.

Depending on the research design, mediation may happen at a different level than the level of the effects (student outcomes, an individual's health outcomes, or an employee), such as through a teacher, a parent, an organizational leader, or a doctor. One example of this is in education, where the effect of higher school SES composition (level 2) affects the outcome of college choice (level 1), which is partially mediated through school practices towards preparing students for college (level 2; Palardy, 2015).

A typical mediation design in education research is the 3-2-1 mediation design (Pituch et al., 2009), where the treatment is assigned at the highest level (the school), implemented at the middle level (the teacher), and evaluated at the individual level (the student). Two examples are a multilevel mediation analysis on the Building Blocks mathematics curriculum (Schenke et al., 2017) and on the Content-Focused literacy Coaching intervention (Matsumura et al., 2013). In both these studies, the treatment (i.e., the Building Blocks mathematics curriculum and Content-Focused Coaching) was assigned at the school level, the mediator was measured at the second level (classroom level) through classroom observations on classroom quality, and the outcome was measured at the student level through achievement in mathematics and reading comprehension respectively (Note, Schenke et al. (2017) use a 2-2-1 analysis, but the true effect of estimating the mediation effects of the intervention would be estimated by a 3-2-1 design whereby the treatment is as the school level, classroom quality was measured at the classroom level and mathematics achievement was measured at the individual level).

Another example is a multilevel mediation analysis on a study on reciprocal teaching and student self-regulated learning (Schünemann et al., 2017). In this study, the treatment was assigned at the classroom level, and the outcome was at the student level. However, for the mediator, the students were working in groups (2nd level), and these groups were observed using videos (Note again: these authors also use a 2-2-1 analysis, but the more rigorous design would again have been to estimate this with a 3-2-1 design). This study had multiple mediators, with two mediators at level two and one mediator at level one. Conceptually, the level 2 mediators (student classroom group observations) influenced the level one mediator (strategy-related task performance), which influenced the outcome (reading comprehension). The 3-2-1 mediation design is not only applicable to educational research but can be applied to the health/medical field where a hospital is at the highest level, the mediation is at the second level, which is the provider, and the outcome is at the patient level (Williams et al., 2022), demonstrating the usefulness and versatility of multilevel mediation analysis. One example of this would be a cluster randomized control trial where the intervention of training on trauma-informed care (Reeves, 2015) to nurses/providers. The treatment would be randomized at the hospital level; the mediator would be nurses' use of trauma-informed care at the nurse level, and the outcome would be patient-reported satisfaction with their care.

However, several problems have arisen from this transition from simple intent-to-treat analyses to including the intervention's mediation effects, including questions of the measurement, data collection, and estimation methods to use for these mediation effects. The remainder of Chapter 1 will investigate the history and different developments of mediation analysis, followed by the current gap in the literature and how this dissertation is addressing this gap.

1.1 History and Development of Mediation Analysis

Single level mediation models

Mediation analysis dates back to prior the 1980s (Ritchie and Miles, 1970; Walberg, 1969). However, it was only in1986 when Baron and Kenny (1986) clearly defined the differences between mediation and moderation, explaining mediation with a path diagram still commonly used today as well as summarizing the estimation methods and sources of issues of mediation analysis at the time. They define moderation as partitioning the effects into subgroups and the mediator as the how/why the effects occur. In this article, they argue that a variable is a mediator under three conditions: 1) the predictor significantly accounts for variation in the mediating variable; 2) the mediator significantly accounts for the variation in the outcome; and 3) the inclusion of the mediator significantly decreases the variation in the outcome that the main predictor explains. They recommend testing this using three regressions; however, they note that this assumes no measurement error and that the outcome variable does not cause the mediator (temporal distinction between the two). If there is measurement error, they recommend using latent variable modeling to model the mediation (Baron and Kenny, 1986).

Following Baron and Kenny (1986)'s work, methods for estimating mediation effects continued to expand, and in 2002, MacKinnon et al. (2002) compared three significant inference methods for mediation effects: casual steps methods, the difference in coefficients methods, and product of coefficients methods using a simulation study. The authors' goal was to guide researchers who were using these methods to estimate the indirect effects in mediation analysis. The indirect effects in mediation are the combined effects of the relationship between the mediator and the predictor and then the outcome and the mediator (Bollen, 1987). They found that the causal steps methods tend to have too low Type I error rates and too low statistical power. Meanwhile, different methods performed better if both the a and b paths were equal to zero or if one of the two paths was equal to zero, making it difficult to choose a method to test the inference. MacKinnon et al. (2002) recommend two different methods for researchers who want to investigate indirect effects to first establish whether the indirect effect ($a \times b$) is zero (empirical distribution or distribution of the z scores of a and b multiplied) and then establish whether both paths are zero (joint significance)—by the early 2010s, bootstrapped and Monte Carlo confidence intervals had been recommended for testing mediation inference (MacKinnon et al., 2004; Preacher and Hayes, 2008; Preacher and Selig, 2012). With new methods and again needing to guide researchers using these methods, Hayes and Scharkow (2013) analyze the accuracy and confidence interval coverage for these new methods along with the Sobel test and distribution of the product across different sample sizes (n < 100, n <200, and n>500). They found that the Sobel test was too conservative in general; the two bootstrap confidence intervals performed the best when there was a true indirect effect, and the Monte Carlo confidence intervals and distribution of the product performed the best when there was no effect. They thus recommend the Monte Carlo confidence intervals and distribution of the product as the best but also conservative tests for mediation inference (Hayes and Scharkow, 2013).

Multilevel mediation models

Moving from mediation in single-level design into multilevel designs in randomized control trials, Bauer et al. (2006) expands mediation that does not require two-step estimation into multilevel models in 1-1-1 and 2-1-1 mediation designs. The 1-1-1 multilevel mediation design indicates that the random assignment, mediation, and outcome are all at the first level. Still, there may be clustering and random effects at a level higher than the first level. This type of study may occur in a health intervention where the treatment is assigned to the patient, the implementation is at the patient level, and the health outcomes are at the patient level; however, the patients are clustered by their providers. The 2-1-1 mediation design is when random assignment is at the second level while the mediator and outcome are at the first level. Bauer et al. (2006) estimate the mediation using a multivariate system of equations and provide the expected value and variance of $a \times b$. The 2-1-1 design may be the case in education where a curriculum intervention is assigned at the school level, student engagement, the mediator, is at the student level, and academic achievement, the outcome, is also at the student level.

Pituch et al. (2009) broaden this work by providing methods for estimating five different multilevel mediation designs: the 3-1-1 design, the 3-2-1 design, the 3-3-1 design, the 2-1-1 design, and the 2-2-1 design. In each of these designs, the first number indicates the level where random assignment occurs, the second number where the mediator is, and the third the outcome level. Thus, the 3-1-1 design indicates randomization at the highest level (such as the school level) and both the mediator and outcome at level one (such as the student level). Following, the 3-2-1 design is randomization at the highest level, the mediator at the middle level, and the outcome at the first level; the 3-3-1 design is randomization and mediation at the highest level and outcome at the lowest level. The 2-1-1 and 2-2-1 are designs with only two levels such as a classroom randomization or a school randomization with teacher outcomes. All of these methods are estimated using multiple multilevel models to estimate the a and b paths, followed by multiplying the paths and testing the inference with Sobel's standard errors (Pituch et al., 2009). Following Bauer et al. (2006) and Pituch et al. (2009)'s expansion of mediation analysis into multilevel designs, Preacher et al. (2010) points out issues in estimation mediation with multilevel models. These issues include bias in the estimation of the indirect effect, the fact that multilevel models are unable to treat upper-level variables as outcomes, such that multilevel models could not be used in the following models: 1-1-2, 1-2-1, 1-2-2, and 2-1-2 models. Additionally, the 2-2-1 mediation models require a two-step process. The authors note that multilevel structural equation modeling faces none of these limitations. Thus, they recommended using general multilevel structural equation modeling in place of multilevel modeling with the one limitation of required sample size (Preacher et al., 2010).

Longitudinal mediation models

Around the same time that mediation was expanding into multilevel models, methods for longitudinal mediation also began emerging. Selig and Preacher (2009) argue for using longitudinal mediation and provide three methods in longitudinal mediation analysis. They note causality concerns when using contemporaneous data for mediation, such as whether the predictor affects the mediator and whether the mediator actually affects the outcome. Jose (2016) specifically shows how concurrent mediation often does not generalize to the proposed longitudinal mediation. The three longitudinal methods provided by Selig and Preacher (2009) are for continuous predictor, mediator, and outcome variables and include a cross-lagged model, a latent growth model, and a latent change model. Jose (2016) expands longitudinal models into experimental and quasi-experimental designs where the treatment is given at T1, the mediator measured at T2, and the outcome measured at T3. Then, Goldsmith et al. (2018) provides applications for the cross-lagged model, latent growth model, and latent change model in experimental settings where the treatment is randomly assigned and the mediator and outcome measures are longitudinal but collected at the same time at every time point, including baseline.

All of these longitudinal data methods assume that all variables are collected at all time points. The cross-lagged model estimates the mediation so that the predictor is one-time point lagged from the mediator and the mediator is one-time point lagged from the outcome, ensuring correct temporal sequencing for mediation causality (Selig and Preacher, 2009; Goldsmith et al., 2018). For example, in a science curriculum intervention where the assumed mediation path is that the intervention increases student interest in science, which would then increase student interest in science careers, if the intervention is assigned at T1 and the students' interest in science and science careers is measured at T1, T2, T3, T4, and T5, then the cross-lagged model would estimate the mediation through the treatment effect on interest in science at T2 accounting for interest in science at T1 and then the relationship between interest in science at T2 and interest in science careers at T3 accounting for interest in science careers at T2. Then, these would be estimated across all time points, consistently accounting for at least one lag. This requires at least three time points for the mediator and outcome to be measured at multiple time points at the same time. The latent growth model allows for the change in the mediator and outcome over time to differ by individuals through latent intercepts and slopes. It also provides for the investigation of the mediator on the rate of change (the slope) of the outcome (Selig and Preacher, 2009; Goldsmith et al., 2018). In an example of a mathematical modeling intervention where the intervention teaches students about mathematical modeling and the assumed mediator is an increased understanding of mathematical modeling, which then increases students' general mathematical ability, both the mediator and outcome variable can be modeled with a growth model where it is expected that students will get better in mathematical modeling and general mathematics over time but allows this growth to be different by students (with the latent intercept and slopes). In this case, the student rate of growth in mathematical modeling can mediate their rate of growth in general mathematics. In these situations, however, the temporal relationship of the mediator occurring before the outcome cannot be established. The latent difference scores allow the relationships between different time points to be different; it also allows for more temporal investigations similar to the cross-lagged model. In this case, respectively, the mediator and outcomes at time t are subtracted from the mediator and outcomes at time t+1, allowing for differences to change across time points if there are more than 3-time points. In the case of 4-time points, the mediation would be the treatment effect on the difference in the mediator between time points 3 and 2 and then the relationship between the mediator difference between time points 3 and 2 and the outcome difference between time points 4 and 3 (Selig and Preacher, 2009; Goldsmith et al., 2018). In the above example, with an intervention addressing students' science career interests, this mediation would be the treatment effect on the change in science interest between time points 2 and 3 and then the relationship between the change in science interest and the change in science career interest between timepoints 3 and 4.

Longitudinal and multilevel mediation models

Combining longitudinal mediation with multilevel mediation, Zhang and Phillips (2018) proposes both a 2-2-1 cross-lagged mediation and a 2-1-1 cross-lagged mediation, including random effects for the level 2 cluster. Finally, McNeish and MacKinnon (2022) applies dynamic structural equation modeling for intensive longitudinal mediation. It expands upon the cross-lagged model but combines time-series and multilevel modeling. Using this dynamic structural equation model allows for the estimation of stationary mediation, person-specific mediation where the mediation varies by person, dynamic mediation where the indirect effect can change over time, and cross-classified mediation which allows the mediation to vary by both person and time (McNeish and MacKinnon, 2022).

Recent measurement concerns in mediation models

Finally, recent years have led to an increase in research on measurement questions regarding mediation. Olivera-Aguilar et al. (2018) investigate the effects of measurement noninvariance in the mediator on the mediation effects. They note that mediation analysis assumes measurement invariance but find that mediation estimation is robust to violations of factorial invariance where the loadings are not equivalent across groups but not to violations of metric invariance where the latent intercepts are not equivalent across groups. However, Olivera-Aguilar et al. (2018) note that even though mediation analysis is not robust to metric invariance, this may be a result of the treatment (or group) effect being tested with the mediation. On a different measurement question, Gonzalez and MacKinnon (2021) investigates different misspecification of models when the true mediator is a bifactor model. They compare ten different mediator models across the various factors, including whether the general factor or a specific factor is the mediator in the estimation. They find that the probability of finding a mediation effect decreases with measurement error and that increasing complexity decreases the power of finding an effect (Gonzalez and MacKinnon, 2021). However, previous work done by these authors had already shown that using a unidimensional model instead of a bifactor model when investigating a specific latent factor can lead to bias and lower power (Gonzalez and MacKinnon, 2018), so there may also be a tradeoff between bias and efficiency in mediation with complex latent structures.

1.2 Current Problem

Researchers interested in answering research questions that include mediators must consider the choice, measurement, and data collection design for the mediator variables. These choices are driven by their assumptions about what could impact the implementation of the outcome, their theoretical mediation models, and measurement considerations for the specific mediators. Within education, it is not uncommon to have longitudinal mediators, such as teacher practices measured through teacher observations at multiple time points, or student emotionality measures, such as engagement, measured through longitudinal data collection methods, such as the experience sampling method (ESM). However, current longitudinal mediation assumes either that the mediator and outcome variables are both being measured longitudinally or that there is growth in the mediator, which then affects the outcome measure.

However, measures can differ from those used in the current longitudinal mediation literature in several ways. First, it is possible that only the mediator is longitudinal as opposed to having both the mediator and outcome longitudinal and measured simultaneously. In essence, these are specific situations where there is a start and an end to an intervention where the treatment is assigned at the beginning; the intervention lasts a specified amount of time (for example, one academic year); the outcome is measured at the end of the intervention; and between the beginning and end, the mediator is measured multiple times. In these cases, neither the cross-lagged model nor the more expansive dynamic structural equation model mediation fits the data. Additionally, in these mediation measures, there may not be an expected growth; instead, the measures assume a person-level trait with also time-specific variance that does not necessarily follow a trend. These measures do not fit with the latent growth model or the latent difference models in mediation analysis.

1.3 Latent Trait-State Theory

To address the gap in estimating mediation with measures that assume a personal level trait but also time-specific variance, this study combines multilevel mediation with latent trait-state theory (LTS). The basis for LTS is to address the longitudinal question of how to consider the variability of a construct between time points. Numerous latent variables measured across time have contextual influences, particularly in education research. One such construct is student engagement (Vongkulluksn and Xie, 2022). Student engagement may change over time, but it may also be that the most significant influence on student engagement is differences in classroom activities, as well as other factors within a student's life, such as their general emotions that day or their interactions with friends and family at that time. Latent Trait-State theory models the separation of the trait (time-invariant), which in the above example would be the student's propensity toward engagement, and the state (time-dependent), which in the above example would be the day-to-day differences in classroom activities or the effects of a students' general mood on any given day, aspects of a latent construct. Although LTS theory has been around since the 1980s under relatively strict assumptions, which assumed that trait constructs did not change over time (Geiser, 2020), Steyer et al. (2015) revised the theory under less stringent assumptions to allow the trait constructs to change over time and provided models for single trait single indicator to single trait multiple indicators to multiple states single indicator to multiple states multiple indicators. The single vs. multiple traits indicate the number of latent constructs at the trait level. The above example of student engagement would be a single trait; however, if the researcher were interested in both student engagement and student science ability simultaneously, those would be multiple traits. The number of indicators is the number of items that load onto the trait constructs. For example, if, at any given time point, there was only one question measuring student engagement, this would be a single indicator. In contrast, if there were multiple questions measuring student engagement, that would be a multiple indicator example.

1.4 Goals and Structure of Study

To deal with longitudinal measures of mediation that have more occasion specificity as opposed to growth, this current study expands upon the current literature on longitudinal mediators by applying a latent trait-state model to multilevel SEM mediation analysis. The goals of this study were:

1. Provide a model to address mediation measures with multiple time points through a latent state-trait perspective as opposed to through a growth perspective

2. Provide different methods for the estimation of these latent state-trait mediation effects

3. Compare the different estimation methods through bias, power, and convergence

4. Apply these methods to an empirical example to explore how they perform in real-world situations

This study achieves these goals by explaining the theoretical basis of latent trait-state theory and statistical multilevel mediation and deriving the combined model, followed by four different ways to potentially estimate this mediation, three of which do not use the fully specified model but instead use estimates of the trait mediator. At the same time, the fourth is the estimation of the full model using structural equation modeling. These estimates of the trait include an average sum, average factor scores from a factor analysis, and general factor scores from the latent state-trait model estimated separately. Then, the dissertation tests the bias and power of the four different estimation methods using a simulation study, followed by an empirical example to show how these estimation methods perform with actual data through the analysis of the mediation effects of teacher practices from a project-based learning (PBL) science intervention in high school chemistry and physics classes.

In the next chapter, the latent trait-state theory is briefly explained, along with how it can be considered in a bifactor model when there are a limited number of time points. The latent trait state model is then expanded into a multilevel model, allowing for random effects. Following this explanation of the latent trait-state theory, a summary of statistical multilevel mediation is given along with considerations for when using an outcome variable that is measured with binary items (which could be expanded to other types of items, such as unordered or ordered categorical items). Then, the multilevel latent trait-state model is combined with the multilevel mediation to provide a mediation model using latent trait-state theory. Following this, the four estimation methods are provided, as described above. Then, in Chapter 3, a simulation study explores how these four different estimation methods perform across 3rd-level cluster sample sizes, general and specific factor loadings in the latent trait state model, and differing effect sizes for the a and b pathways in the mediation analysis. Finally, chapter 4 explores these estimation methods in a real-world science curriculum intervention (Schneider et al., 2022) where teacher observations measured at multiple times are used as a mediator with an assumed latent trait state model for teacher practices. Finally, chapter 5 wraps up a discussion on this proposed model and how it performed, followed by current limitations and future research.

CHAPTER 2

LATENT TRAIT-STATE THEORY AND STATISTICAL MULTILEVEL MEDIATION

This chapter gives the theoretical basis for latent trait-state theory, followed by advances in multilevel latent trait-state theory. Then, after delving into statistical mediation and, more specifically, multilevel mediation, it combines the two theoretical models into the latent trait-state multilevel mediation model. Finally, it will provide estimation methods for the latent trait-state multilevel mediation model.

2.1 Latent Trait-State Theory

The proposed mediation model of this study considers an LTS model where there are multiple indicators for the mediation construct so that there are multiple questions measuring the mediating construct, which is measured over time (at multiple time points) and does not restrict time invariance of the trait. Not restricting time invariance means that how each indicator measures the mediator may not be the same from time to time. Although time invariance of the trait would be ideal, as many researchers hope that the measure always measures the same construct with the same scale, there are several instances where this may be violated in an experimental study. For example, a mediator with an observer or scorer with a different person observing or scoring at other times may change how the items load onto the construct. Another example would be a self-administered survey. As time changes, the individual's perception of the question changes as they gain a better understanding of what the question may be asking. The proposed model does not assume time invariance; however, a simple constraining of parameters would shift this assumption so that the proposed model is plausible in the situation of time invariance or time non-invariance.

Starting with understanding the LTS model to be incorporated into the mediation model, this study proposes an LTS model through a bifactor model for a single trait multistate model with multiple indicators. The bifactor model is used in a broad set of measurement contexts, such as personality (Chen et al., 2012), assessments (DeMars, 2006), and in the single trait multistate LTS (Geiser, 2020). In the more general context, the bifactor model is a multidimensional model where each item loads (or is related) to a general factor and then another more specific factor.

An example of this would be in a science assessment where the general factor would be general science ability, but the specific factors may be content-specific factors (such as physical science, biological science, or environmental science) or skill-specific factors, such as different science practices (modeling, mathematical representations, or planning an investigation). These models often fit the data better and give more information about students' general and specific abilities. More specifically, using the bifactor model in the LTS context, each item loads onto the trait factor and then each time-specific factor, allowing for the distinguishing of trait factors from the state (time-specific) variance. In the example of student engagement, the general/trait factor would again be the students' propensity to be engaged, and the specific/state factor would be the time-specific variations in student engagement.

Statistically, this LTS bifactor model is:

$$Y_{itj} = \alpha_{it0} + \lambda_{it1}\xi_j + \delta_{it2}\zeta_{tj} + \epsilon_{itj}$$

Where Y_{itj} represents item i at time point t for individual j, α_{it0} is the intercept for item i at time point t, λ_{it1} is the time specific trait loading for item i at time t, ξ_j is the common trait factor for individual j, δ_{it2} is the factor loading for item i at time point t on the state residual factor, ζ_{tj} is the state residual factor at time point t for individual j, and ϵ_{itj} is the unique factor for item i at time t for individual j (or otherwise, the measurement error; Geiser, 2020, p. 179).

In the above example of student engagement measured at random timepoints, if the indicators for engagement are student interest, skill, and challenge (Schneider et al., 2016), then Y_{itj} is item i (either interest, skill, or challenge) at the specific time point t for student j, α_{it0} is the intercept of the indicator (interest, skill, or challenge) at the specific time point t, λ_{it1} is the loading of the indicator (interest, skill, or challenge) at time point t on the students' propensity to be engaged, ξ_j (trait factor). δ_{it2} is the loading of the indicator (interest, skill, or challenge) onto the state residual factor (time-specific factor of the student's engagement, ζ_{tj}), ϵ_{itj} is the remaining unexplained variance in the student's response to the indicator of interest, skill, or challenge.

In a situation with three indicators (such as the engagement example above) at three-time points, this model is visually represented by Figure 2.1.

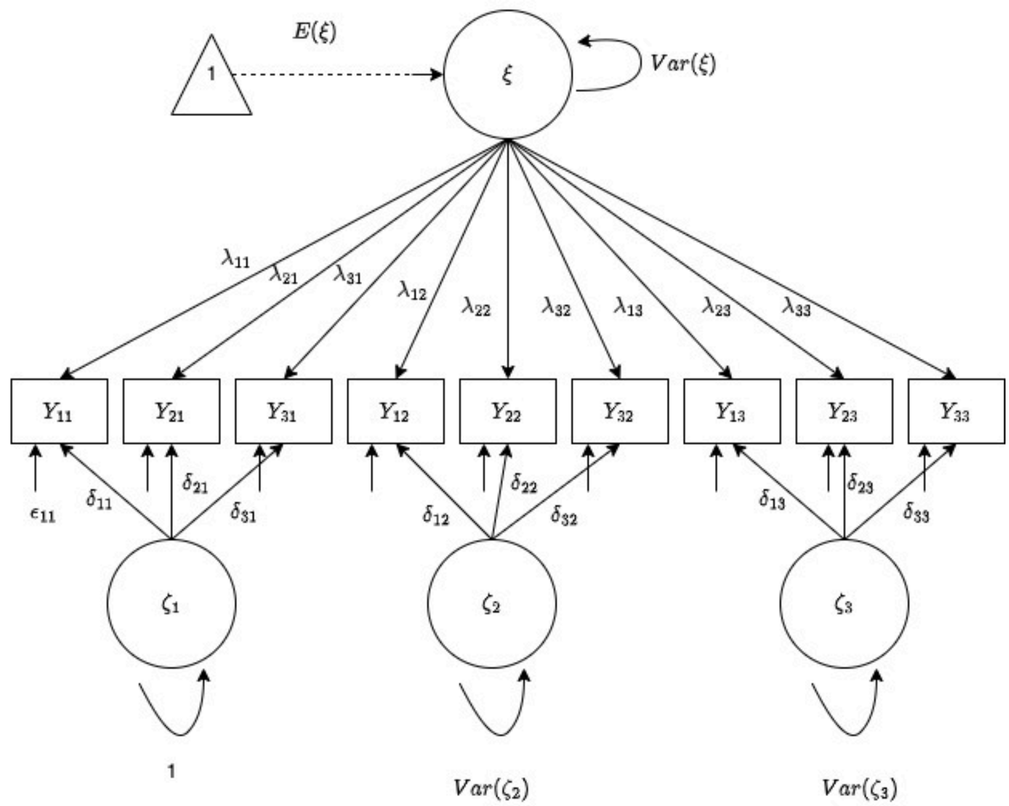

Figure 2.1 Bifactor Latent State-Trait Model

This bifactor latent state-trait model can be expanded into a multilevel component in cases where the observations are clustered, such as teachers or students within schools (Wang et al., 2018). When extending the bifactor model to two levels, the trait and state factors now have within and between-level components. For example, if level 1 is students and level 2 is schools, there are within-school (student level) trait and state factors that represent these factors for the individual students, but then there are between-school (school level) trait and state factors. These indicate that variance in these trait and state factors is explained by the students belonging to the same school.

Level 1:

$$Y_{itjk} = \alpha_{it0k} + \lambda_{it1k}\xi^{w}_{jk} + \delta_{it2k}\zeta^{w}_{tjk} + \epsilon_{itjk}$$

Level 2:

$$\lambda_{it1k} = \lambda_{it10}$$

$$\delta_{it2k} = \delta_{it20}$$

$$\alpha_{it0k} = \alpha_{it00} + \lambda_{it01}\xi_k^b + \delta_{it02}\zeta_{tk}^b + \upsilon_{it0k}$$

Where ξ_k^b is the between level common trait factor for cluster k; ζ_{tk}^b is the between level state factor for cluster k at time point t; and υ_{it0k} is the random intercept for item i at time point j for cluster k. In the engagement example with students nested within schools, ξ_k^b is the school level propensity for students to be engaged in school k (do these propensities vary by school), ζ_{tk}^b is the school level time specific engagement factor for school k at time point t(factors such as classroom practices that affect engagement from day to day), and υ_{it0k} is the random variance for school k on the responses to the indicators (challenge, interest, and skill).

2.2 Statistical Multilevel Mediation

The 3-2-1 statistical multilevel mediation model is exemplified in a standard education RCT sampling and data collection framework. Schools are randomly sampled and assigned treatment. All teachers within the school who are applicable to the treatment are included in the study (or randomly chosen if the treatment applies to all teachers). Then, all students within those teachers are included. Treatment is assigned before the start of the school year. Teachers are randomly observed once throughout the school year to measure teacher practices as a mediator. Finally, the outcome is measured at the student level at the end of the year (after all observations and at the end of the treatment).

Pituch et al. (2009) provide the following mediation model for the 3-2-1 mediation design:

Level 1:

$$Y_{jkl} = \pi_{0kl} + e_{jkl}$$

Level 2:

$$M_{kl} = \beta_{0l}^M + r_{kl}^M$$

$$\pi_{0kl} = \beta_{00l} + \beta_{01l}M_{kl} + r_{0kl}$$

Level 3:

$$\beta_{0l}^M = \gamma_{00}^M + \gamma_{01}^M T_l + u_{0l}^M$$

$$\beta_{00l} = \gamma_{000} + \gamma_{001}T_l + u_{00l}$$

$$\beta_{01k} = \gamma_{010}$$

Where M_{kl} is the mediator at level two for the level two cluster k within the level three cluster l; T_l is the treatment at the level 3 cluster l; Y_{jkl} is the outcome at level 1 for individual j in the level 2 cluster k in the level 3 cluster l. e_{jkl} is the level 1 residual on the outcome for individual j in the level 2 cluster k in the level 3 cluster l; r^M_{0kl} is the level 2 residual on the mediator for the level two cluster k within the level three cluster l; r_{0kl} is the level 2 random intercept on the outcome for the level two cluster k within the level three cluster l; u^M_{0l} is random intercept for the level 3 cluster l on the mediator; u_{00l} is random intercept for the level 3 cluster l on the outcome. γ^M_{00} is mean of the mediator; γ_{000} is the mean of the outcome;γ^M_{01} is pathway a in the mediation estimation; γ_{010} is the pathway b where the product of a and b estimates the indirect effect; and $\gamma_{001}T_l$ is the c' pathway or the treatment effect not explained by the mediator. In the example where the mediator is teacher practice and outcome, student achievement, M_{kl} is a single indicator for teacher practice; Y_{jkl} is a single indicator for student achievement (such as a standardized test score); γ^M_{01} is the effect of the treatment on teacher practices; γ_{010} is the relationship between student achievement and teacher practices; and $\gamma_{001}T_l$ is the treatment effect on student achievement not explained by the mediator.

The 3-2-1 mediation can be applied in a situation where the mediator and outcome are latent variables (have multiple indicators that load onto the latent variable) instead of observed variables in a multilevel SEM framework (Silva et al., 2019). This may be a situation where the mediator, teacher practices, are measured with multiple items in the observation, and the outcome, student achievement, is measured with a test with multiple items. In this case, the mediator and outcome are latent constructs that may also have measurement error (these measures are not perfectly reliable instruments). This expansion into multilevel SEM is essential, especially in education research trials where the outcome variable is often a student-level construct that includes some measurement error.

Level 1:

$$Y_{ijkl} = \alpha_{i0kl} + \lambda_{i1kl}\theta^w_{jkl} + \epsilon_{ijkl}$$

Level 2:

$$M_{ikl} = \alpha^M_{i0l} + \lambda^M_{i1l}\theta^{Mw}_{kl} + \epsilon^M_{ikl}$$

$$\alpha_{i0kl} = \alpha_{i00l} + \lambda_{i01l}\theta^{bk}_{kl} + \upsilon_{i0kl}$$

$$\lambda_{i1kl} = \lambda_{i10l}$$

Level 3:

$$\alpha^M_{i0l} = \alpha^M_{i00} + \lambda^M_{i01}\theta^{Mb}_l + \upsilon^M_{i0l}$$

$$\lambda^M_{i1l} = \lambda^M_{i10}$$

$$\theta^{Mb}_l = \gamma_0 + \gamma_1 T_l + u^M_l$$

$$\alpha_{i00l} = \alpha_{i000} + \lambda_{i001}\theta^{bl}_l + \nu_{i00l}$$

$$\lambda_{i01l} = \lambda_{i010}$$

$$\lambda_{i10l} = \lambda_{i100}$$

$$\theta^{bl}_l = \beta_0 + \beta_1 T_l + \beta_2\theta^{Mb}_l + u$$

Where latent mediator variable is split into a within and a between latent construct (θ^{Mw}_{kl} and θ^{Mb}_l) and the latent outcome variable is split into a within, a between 2nd level, and a between 3rd level construct (θ^w_{jkl}, θ^{bk}_{kl}, and θ^{bl}_l respectively; such as within students, between teachers, and between schools). Like the multilevel LTS model, the latent constructs are allowed to vary by each level. In the example above, the student's academic ability would have a within-level (student level) ability score, a teacher-level score, and a school-level score, allowing for the measured ability to vary across teachers and schools. Similarly, the latent variable of teacher practices is allowed to vary at the school level, so there is a difference in the average teacher practice factor scores between schools.

In this multilevel SEM, the treatment and mediation effects are estimated at the third level. γ_{01} is the pathway a in the mediation estimation and β_2 is pathway b where the product of a and b estimates the indirect effect. Similarly to the single indicator model above, in the example, now, γ_{01} is the effect of the treatment on the between school teacher practices, and β_2 is the relationship between the school level teacher practices and the between school level student academic ability. This procedure allows for estimating the mediation effects in the presence of measurement error within the mediator and the outcome variable. This model is represented in Figure 2.2.

2.3 Latent Trait-State Theory in Multilevel Mediation

Combining Latent Trait-State Theory into the mediation framework, where the mediator is a longitudinal measure that has both trait and state latent variables, such as in the case where teacher practices or social and emotional well-being are measured at multiple time points during the intervention, the model becomes:

Level 1:

$$Y_{ijkl} = \alpha_{i0kl} + \lambda_{i1kl}\theta^{w}_{jkl} + \epsilon_{ijkl}$$

Level 2:

$$M_{itkl} = \alpha_{it0l} + \lambda_{it1l}\xi^{w}_{kl} + \delta_{it2l}\zeta^{w}_{tkl} + \epsilon^{M}_{itkl}$$

$$\alpha_{i0kl} = \alpha_{i00l} + \lambda_{i01l}\theta^{bk}_{kl} + \upsilon_{i0kl}$$

$$\lambda_{i1kl} = \lambda_{i10l}$$

Level 3:

$$\alpha_{it0l} = \alpha_{it00} + \lambda_{it01}\xi^{b}_{l} + \delta_{it02}\zeta^{b}_{tl} + \upsilon^{M}_{it0l}$$

$$\lambda_{it1l} = \lambda_{it10}$$

$$\delta_{it2l} = \delta_{it20}$$

$$\alpha_{i00l} = \alpha_{i000} + \lambda_{i001}\theta^{bl}_{l} + \nu_{i00l}$$

$$\lambda_{i01l} = \lambda_{i010}$$

$$\lambda_{i10l} = \lambda_{i100}$$

$$\xi^{b}_{l} = \gamma_0 + \gamma_1 T_l + \epsilon$$

$$\theta^{bl}_{l} = \beta_0 + \beta_1\xi^{b}_{l} + \beta_2 T_l + \upsilon$$

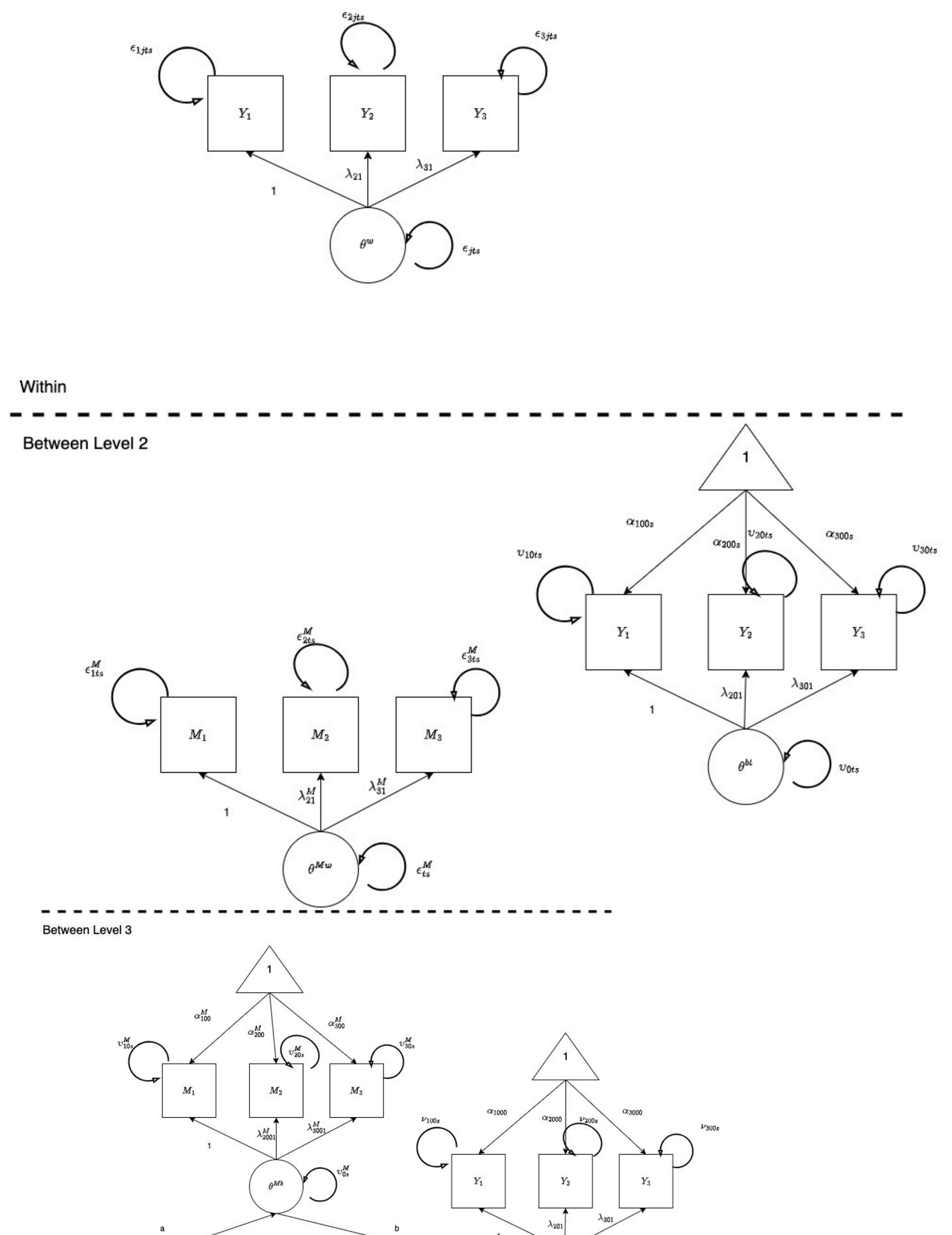

Figure 2.2 Multilevel Structural Equation Mediation Model

Where Y_{ijkl} is item i for individual j in the level 2 cluster k within the level 3 cluster l on the outcome measure and M_{itkl} is the mediator item i at time point t for the level 2 cluster k within the level 3 cluster l. Where a latent mediator is modeled by the mediator trait factor which is split into a within and a between construct (ξ^{w}_{kl} and ξ^{b}_{l}), the mediator state factor which is split into a within and between level construct (ζ^{w}_{tkl} and ζ^{b}_{tl}) and the latent outcome variable is split into a within, a between 2nd level, and a between 3rd level construct (θ^{w}_{jkl}, θ^{bk}_{kl}, and θ^{bl}_{l} respectively). α_{it00} is the mean of the mediator item i, λ_{it01} is the mediator item i at time t loadings onto the between level 3 trait factor, δ_{it02} is the mediator item i at time t loadings onto the between level 3 state factor, υ^{M}_{it0l} is the level 3 random intercept on the mediator item i, λ_{it10} is the mediator item at time t loadings onto the level 2 trait factor, δ_{it20} is the mediator item i loadings at time t onto the level 2 state factor, α_{i000} is the mean of the outcome item i, λ_{i001} is the outcome item i loading onto the level 3 outcome factor, ν_{i00l} is the level 3 random intercept for item i. λ_{i010} is the outcome item i loading for the level 2 outcome factor, υ_{i0kl} is the level 2 random intercept for outcome item i, λ_{i100} is the outcome item i loading for the level 1 outcome factor. In this multilevel SEM, the treatment and mediation effects are estimated at the third level. γ_1 is the pathway a in the mediation estimation and β_1 is pathway b where the product of a and b estimates the indirect effect. Finally, β_2 is the c' pathway or the remaining treatment effect not explained by the mediator.

Returning to the healthcare example from Chapter 1 (see page 3), if level 1 were patients, level 2, nurses, and level 3, hospitals where the treatment is training on trauma-informed care and assigned at the hospital level, the mediator of nurse (level 2) use of trauma-informed care is measured longitudinally, and the outcome is patient-reported satisfaction (measured with multiple items), then Y_{ijkl} represents item i on the patient reported satisfaction survey for patent j under nurse k within hospital l and M_{itkl} represents item i on the nurse's measure of trauma-informed care use at time point t for nurse k within hospital l. Then, ξ^{w}_{kl} is the use of trauma informed care trait for nurse k in hospital l and ξ^{b}_{l} is the hospital level use of trauma informed care trait for hospital l (allowing for this trait to vary by different hospitals); ζ^{w}_{tkl} is trauma informed care time point variance at time point t for nurse k in hospital l; and ζ^{b}_{tl} is the hospital level trauma informed care time point variance at time point t for hospital l (allowing hospital wide time point factors, such as how busy the hospital is, to number of staff out, to general stress levels of staff, to effect the time point variance of trauma informed care use); θ^{w}_{jkl} is the satisfaction factor for patient j under nurse k in hospital l, θ^{bk}_{kl} is the between nurses patient satisfaction factor for nurse k in hospital l (so that the patient satisfaction factor varies by nurses), and θ^{bl}_{l} is the hospital level patient satisfaction factor for hospital l (allowing for variance in the patient satisfaction factor at the hospital level as well). Finally, γ_1 is the hospital-level effect of the training on nurse's trauma-informed care use trait, β_1 is the relationship between the hospital-level nurse's trauma-informed care use trait, and the hospital-level patient satisfaction, and β_2 is the remaining effect of the training directly on hospital level patient satisfaction not explained by the nurse's trauma-informed care use trait.

Next, suppose the logit function is applied to the outcome variable, such as in situations where the outcome measure is a student achievement test (where the items are binary). In that case, the model becomes as is shown below. The only difference here compared to above is how the outcome items are modeled.

Level 1:

$$Y_{ijkl} = \frac{1}{1 + e^{-(u_{i0kl} + a_{i1kl}(\theta^{w}_{jkl} - b_{i2kl}))}}$$

Level 2:

$$M_{itkl} = \alpha_{it0l} + \lambda_{it1l}\xi^{w}_{kl} + \delta_{it2l}\zeta_{tkl} + \epsilon^{M}_{itkl}$$

$$u_{i0kl} = u_{i00l} + a_{i01l}(\theta^{bk}_{kl} - b_{i02l})$$

$$a_{i1kl} = a_{i10l}$$

$$b_{i2kl} = b_{i20l}$$

Level 3:

$$\alpha_{it0l} = \alpha_{it00} + \lambda_{it01}\xi^{b}_{l} + \delta_{it02}\zeta^{b}_{tl} + \upsilon^{M}_{it0l}$$

$$\lambda_{it1l} = \lambda_{it10}$$

$$\delta_{it2l} = \delta_{it20}$$

$$u_{i00l} = u_{i000} + a_{i001}(\theta_l^{bl} - b_{i002})$$

$$a_{i01l} = a_{i010}$$

$$b_{i02l} = b_{i020}$$

$$a_{i10l} = a_{i100}$$

$$b_{i20l} = b_{i200}$$

$$\xi_l^b = \gamma_0 + \gamma_1 T_l + \epsilon$$

$$\theta_l^{bl} = \beta_0 + \beta_1 \xi_l^{bl} + \beta_2 T_l + \upsilon$$

The parameters for the mediator at level 2 and then level 3, as well as the mediation and treatment effects, remain the same as those above; however, the parameters that are different are in the outcome model. Here, a_{i100} is the level 1 item discrimination parameter for outcome item i, b_{i200} is the level 1 item difficulty parameter for outcome item i, a_{i010} is the level 2 item discrimination parameter for outcome item i, b_{i020} is the level 2 item difficulty parameter for outcome item i, a_{i001} is the level 3 item discrimination parameter for outcome item i, and b_{i002} is the level 3 item difficulty parameter for outcome item i. In many cases, it is not expected that the item difficulty will vary by level 2 or level 3, in which case these would be constrained to zero. Item difficulty is related to how "easy" the item is or how likely someone will get it correct, with larger difficulty parameters indicating easier items. Item discrimination is the ability of the item to differentiate individuals based on their true ability.

In the education curriculum example where level 1 is the student level, level 2 is the teacher level, and level 3 is the school level and where the curriculum is randomly assigned at the school level, teachers implement it in their classrooms (at the second level), and the effect is measured through a multiple choice test (binary items) administered at the student level, then Y_{ijkl} represents item i on the multiple choice test for student j in teacher k's classroom in school l where each item has its own student level, teacher level, and school level discrimination (a_{i100},a_{i010}, and a_{i001}) and difficulty parameters (b_{i200},b_{i020}, and b_{i002}) and M_{itkl} represents item i on the implementation measure at time t for teacher k in school l. Then, ξ_{kl}^w is implementation trait for teacher k in school l and ξ_l^b is the school level implementation trait for school l (allowing for this trait to vary by different schools); ζ_{tkl}^w is the teacher's implementation at time point t variance for teacher k in school l and ζ_{tl}^b is the school level implementation at time point t variance for school l (allowing school wide time point factors); θ_{jkl}^w is the student academic ability for student j in teacher j's class in school l, θ_{kl}^{bk} is the between teacher student academic ability for teacher k in school l (so that student academic ability varies by teachers), and θ_l^{bs} is the school level student academic ability for school l (allowing for variance in the student academic ability at the school level as well). Finally, γ_1 is the school-level effect of the curriculum on the teacher's implementation trait, β_1 is the relationship between the school-level teacher implementation trait and the school-level student academic ability, and β_2 is the remaining effect of the curriculum directly on school level student academic ability not explained by the teacher's implementation.

The above model is represented in Figure 2.3.

2.4 Estimation of the LST Mediation

In estimating the mediating effects of a longitudinal variable that follows a latent state-trait model, four estimation methods are explored in this dissertation, along with the simulation study to examine the bias and power of each method. The four estimation methods are the averages of the mediator across time points using a 3-2-1 mediation, using factor scores for each time point and averaging them across the time points as the mediator used in a 3-2-1 mediation model; factor scores from the latent state-trait model as the mediator in the 3-2-1 mediation; and finally, estimation of the fully specified multilevel LST mediation. The four estimation methods have different assumptions that are required for them to be unbiased and recommended. In the following sections, each estimation method will be defined and incorporated into an estimation with a single indicator outcome (not a latent variable) and a multiple indicator outcome (latent variable). Then, the assumptions for each of the estimation methods will be described.

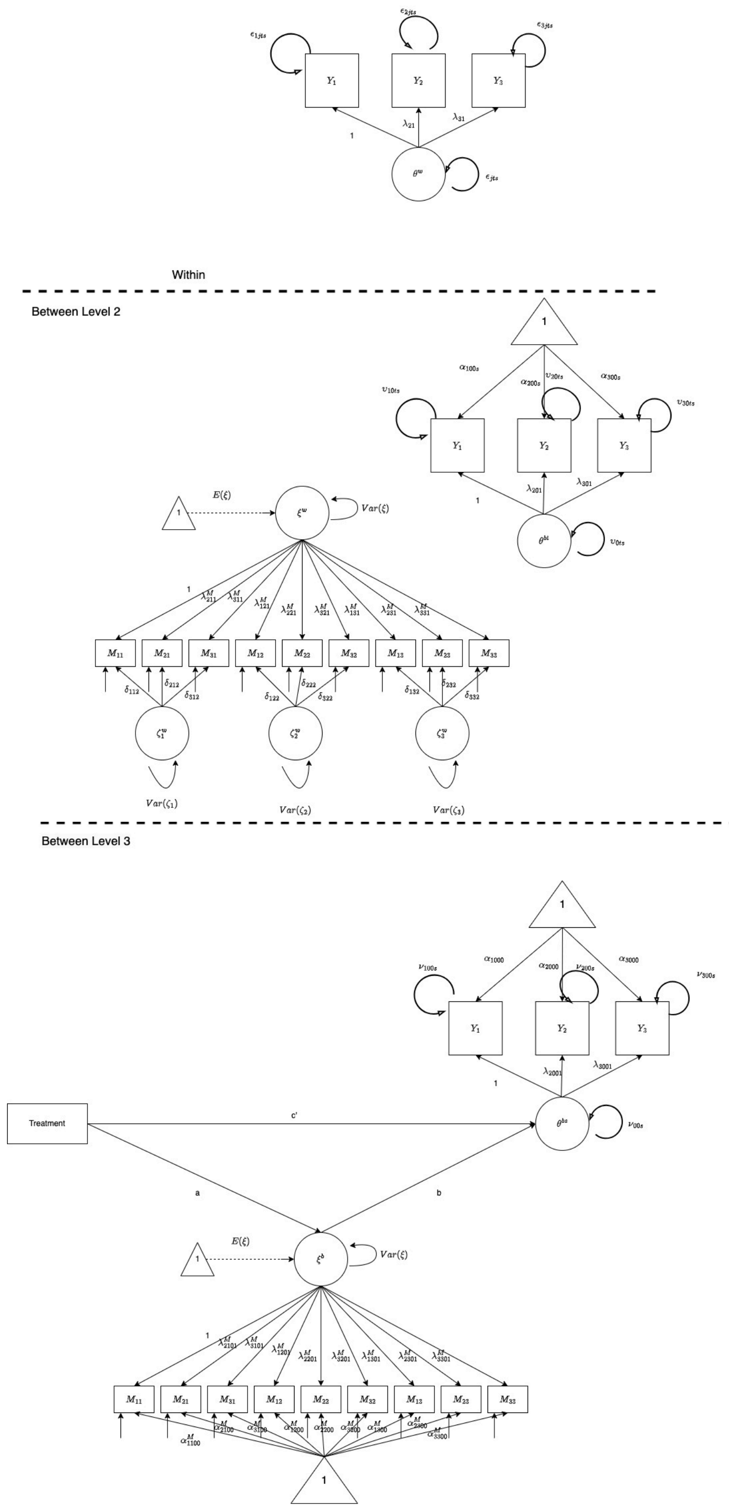

Figure 2.3 LST in Multilevel Mediation Model

First Estimation: averages of the summed mediator

For the first estimation method, the averages of the mediator across the time points will be defined as:

$$\bar{M}_{kl} = \frac{1}{T \times I} \Sigma_{t=1}^{t=T} \Sigma_{i=1}^{i=I} M_{itkl}$$

T is the total number of time points, and I is the total number of items in the mediator measure. $\bar{M}$ is the average of the items averaged across time points. In the situation where the mediator is teacher practices, this would be the averages of the teacher practice scores at each time that the teacher is observed. Then, these are averaged across the observation time points.

Using $\bar{M}_{kl}$ in the 3-2-1 mediation analysis, the following is estimated:

Single Indicator Outcome

Level 1:

$$Y_{jkl} = \pi_{0kl} + e_{jkl}$$

Level 2:

$$\bar{M}_{kl} = \beta_{0l}^{M} + r_{kl}^{M}$$

$$\pi_{0kl} = \beta_{00l} + \beta_{01l} \bar{M}_{kl} + r_{0kl}$$

Level 3:

$$\beta_{0l}^{M} = \hat{\gamma}_{00}^{M} + \hat{\gamma}_{01}^{M} T_l + u_{0l}^{M}$$

$$\beta_{00l} = \hat{\gamma}_{000} + \hat{\gamma}_{001} T_l + u_{00l}$$

$$\beta_{01l} = \hat{\gamma}_{010}$$

Then, $\hat{\gamma}_{01}^{M}$ is the estimated a pathway, $\hat{\gamma}_{010}$ is the estimated b pathway, and $\hat{\gamma}_{001}$ is the estimated c' pathway for the mediation analysis. With the inclusion of the average time points, these can be estimated with either a multilevel model or a multilevel structural equation model with either robust maximum likelihood or Bayesian estimation.

Multiple Indicator Outcome

Level 1:

$$Y_{ijkl} = \alpha_{i0kl} + \lambda_{i1kl} \theta_{jkl}^{w} + \epsilon_{ijkl}$$

Level 2:

$$\bar{M}_{kl} = \beta_{0l}^{M} + r_{kl}^{M}$$

$$\alpha_{i0kl} = \alpha_{i00l} + \lambda_{i01l}\theta_{kl}^{bk} + \upsilon_{i0kl}$$

$$\theta_{kl}^{bk} = \beta_{0l} + \beta_{1l}\bar{M}_{kl} + r_{kl}$$

$$\lambda_{i1kl} = \lambda_{i10l}$$

Level 3:

$$\beta_{0l}^{M} = \hat{\gamma}_{00}^{M} + \hat{\gamma}_{01}^{M}T_l + u_{0l}^{M}$$

$$\alpha_{i00l} = \hat{\alpha}_{i000} + \hat{\lambda}_{i001}\theta_{l}^{bl} + \nu_{i00l}$$

$$\lambda_{i01l} = \hat{\lambda}_{i010}$$

$$\lambda_{i10l} = \hat{\lambda}_{i100}$$

$$\beta_{0l} = \hat{\beta}_{00}$$

$$\beta_{1l} = \hat{\beta}_{10}$$

$$\theta_{l}^{bl} = \hat{\beta}_0 + \hat{\beta}_1 T_l + u$$

Then, $\hat{\gamma}_{01}^{M}$ is the estimated a pathway, $\hat{\beta}_{00}$ is the estimated b pathway, and $\hat{\beta}_1$ is the estimated c' pathway for the mediation analysis. With the inclusion of the average time points and using multiple indicator outcomes, these can be estimated with a multilevel structural equation model with either robust maximum likelihood or Bayesian estimation.

In the situation where the mediator is teacher practices and the outcome is student achievement, the $\hat{\gamma}_{01}^{M}$ is the effect of the treatment on teacher practices, $\hat{\gamma}_{010}$ is the relationship between teacher practices and student achievement, and $\hat{\gamma}_{001}$ is the remaining effect of the treatment on student achievement not due to teacher practices.

This estimation method of using the summed averages of the mediator assumes no time-varying variance. It also assumes that all items are equally loading onto the construct and that there is no measurement error. In the teacher practices example, the mediation measure perfectly measures the teacher practices at each time point with no other outside influences affecting the estimation of the teacher practices, and each item is weighted equally with regards to estimating the teacher practices construct. If these assumptions are met, this estimation method will be unbiased, tend to converge more quickly than the other methods, be easy to interpret and implement, and be the most parsimonious.

Second Estimation: averages of the factor scores

For the second estimation method, the factor for each time point will be defined and estimated by:

$$M_{itkl} = \alpha_{it00} + \hat{\delta}_{it10}\hat{\zeta}_{tkl} + \epsilon^{M}_{itkl}$$

$$\bar{\zeta}_{kl} = \frac{1}{T}\Sigma_{t=1}^{t=T}\hat{\zeta}_{tkl}$$

Where, once again, T is the total number of time points. $\hat{\zeta}_{kl}$ is the average factor scores across the timepoints. In the teacher practices example, at each time point, the observation scores for each item would be included in a factor analysis to estimate the teacher practice factor. Then, these estimated factor scores for teacher practice would be averaged across each time point for the teacher.

Using $\hat{\zeta}_{kl}$ in the 3-2-1 mediation analysis, the following is estimated:

Single Indicator outcome

Level 1:

$$Y_{jkl} = \pi_{0kl} + e_{jkl}$$

Level 2:

$$\bar{\zeta}_{kl} = \beta^{M}_{0l} + r^{M}_{0kl}$$

$$\pi_{0kl} = \beta_{00l} + \beta_{01l}\bar{\zeta}_{kl} + r_{kl}$$

Level 3:

$$\beta^{M}_{0l} = \hat{\gamma}^{M}_{00} + \hat{\gamma}^{M}_{01}T_l + u^{M}_{0l}$$

$$\beta_{00l} = \hat{\gamma}_{000} + \hat{\gamma}_{001}T_l + u_{00l}$$

$$\beta_{01l} = \hat{\gamma}_{010}$$

Then, $\hat{\gamma}_{01}^{M}$ is the estimated a pathway, $\hat{\gamma}_{010}$ is the estimated b pathway, and $\hat{\gamma}_{001}$ is the estimated c' pathway for the mediation analysis. With the inclusion of the factor scores, these can be estimated with either a multilevel model or a multilevel structural equation model with either robust maximum likelihood or Bayesian estimation.

Multiple indicator outcome

Level 1:

$$Y_{ijkl} = \alpha_{i0kl} + \lambda_{i1kl}\theta_{jkl}^{w} + \epsilon_{ijkl}$$

Level 2:

$$\bar{\zeta}_{kl} = \beta_{0l}^{M} + r_{kl}^{M}$$

$$\alpha_{i0kl} = \alpha_{i00l} + \lambda_{i01l}\theta_{kl}^{bk} + \upsilon_{i0kl}$$

$$\theta_{kl}^{bk} = \beta_{0l} + \beta_{1l}\bar{\zeta}_{kl} + r_{kl}$$

$$\lambda_{i1kl} = \lambda_{i10l}$$

Level 3:

$$\beta_{0l}^{M} = \hat{\gamma}_{00}^{M} + \hat{\gamma}_{01}^{M}T_{l} + u_{0l}^{M}$$

$$\alpha_{i00l} = \hat{\alpha}_{i000} + \hat{\lambda}_{i001}\theta_{l}^{bl} + \nu_{i00l}$$

$$\lambda_{i01l} = \hat{\lambda}_{i010}$$

$$\lambda_{i10l} = \hat{\lambda}_{i100}$$

$$\beta_{0l} = \hat{\beta}_{00}$$

$$\beta_{1l} = \hat{\beta}_{10}$$

$$\theta_{l}^{bl} = \hat{\beta}_{0} + \hat{\beta}_{1}T_{l} + u$$

Then, $\hat{\gamma}_{01}^{M}$ is the estimated a pathway, $\hat{\beta}_{00}$ is the estimated b pathway, and $\hat{\beta}_{1}$ is the estimated c' pathway for the mediation analysis. With the inclusion of these factor scores and using a multiple indicator outcome, these can be estimated with a multilevel structural equation model with either robust maximum likelihood or Bayesian estimation.

Similarly in the situation where the mediator is teacher practices and the outcome is student achievement, the $\hat{\gamma}_{01}^{M}$ is the effect of the treatment on teacher practices, $\hat{\gamma}_{010}$ is the relationship between teacher practices and student achievement, and $\hat{\gamma}_{001}$ is the remaining effect of the treatment on student achievement remaining not due to teacher practices.

This estimation method assumes no general trait outside of the average of each time point but does allow for time-varying variance. Similarly to the estimation method above, this estimation also assumes no measurement error in the estimation of the time-specific factor scores. Again, in the example with teacher practices, this assumes that at each time point, the measure perfectly estimates the teacher practice with the given item loadings estimated in the factor analysis at that specific time point. Under these assumptions, this estimation method is unbiased, will have fewer convergence issues, and will be the more parsimonious method in estimating the factor scores than the following estimation method.

Third estimation: factor scores from the LST model

For the third estimation method, the factor from the LST model is estimated by:

$$M_{itkl} = \hat{\alpha}_{it00} + \hat{\lambda}_{it10}\hat{\xi}_{kl} + \hat{\delta}_{it20}\zeta_{tkl} + \epsilon_{itkl}$$

Using $\hat{\xi}_{kl}$ in the 3-2-1 mediation analysis, the following is then estimated:

Single indicator outcome

Level 1:

$$Y_{jkl} = \pi_{0kl} + e_{jkl}$$

Level 2:

$$\hat{\xi}_{kl} = \beta_{0l}^{M} + r_{kl}^{M}$$

$$\pi_{0kl} = \beta_{00l} + \beta_{01l}\hat{\xi}_{kl} + r_{0kl}$$

Level 3:

$$\beta_{0l}^{M} = \hat{\gamma}_{00}^{M} + \hat{\gamma}_{01}^{M}T_l + u_{0l}^{M}$$

$$\beta_{00l} = \hat{\gamma}_{000} + \hat{\gamma}_{001}T_l + u_{00l}$$

$$\beta_{01l} = \hat{\gamma}_{010}$$

Then, $\hat{\gamma}_{01}^{M}$ is the estimated a pathway, $\hat{\gamma}_{010}$ is the estimated b pathway, and $\hat{\gamma}_{001}$ is the estimated c' pathway for the mediation analysis. With the inclusion of the estimated trait factor scores, these can be estimated with either a multilevel model or a multilevel structural equation model with either robust maximum likelihood or Bayesian estimation.

Multiple indicator outcome

Level 1:

$$Y_{ijkl} = \alpha_{i0kl} + \lambda_{i1kl}\theta_{jkl}^{w} + \epsilon_{ijkl}$$

Level 2:

$$\hat{\xi}_{kl} = \beta_{0l}^{M} + r_{kl}^{M}$$

$$\alpha_{i0kl} = \alpha_{i00l} + \lambda_{i01l}\theta_{kl}^{bk} + \upsilon_{i0kl}$$

$$\theta_{kl}^{bk} = \beta_{0l} + \beta_{1l}\hat{\xi}_{kl} + r_{kl}$$

$$\lambda_{i1kl} = \lambda_{i10l}$$

Level 3:

$$\beta_{0l}^{M} = \hat{\gamma}_{00}^{M} + \hat{\gamma}_{01}^{M} T_l + u_{0l}^{M}$$

$$\alpha_{i00l} = \hat{\alpha}_{i000} + \hat{\lambda}_{i001}\theta_{l}^{bl} + \nu_{i00l}$$

$$\lambda_{i01l} = \hat{\lambda}_{i010}$$

$$\lambda_{i10l} = \hat{\lambda}_{i100}$$

$$\beta_{0l} = \hat{\beta}_{00}$$

$$\beta_{1l} = \hat{\beta}_{10}$$

$$\theta_{l}^{bl} = \hat{\beta}_{0} + \hat{\beta}_{1} T_l + u$$

Then, $\hat{\gamma}_{01}^{M}$ is the estimated a pathway, $\hat{\beta}_{00}$ is the estimated b pathway, and $\hat{\beta}_{1}$ is the estimated c' pathway for the mediation analysis. With the inclusion of these estimated trait factor scores and using a multiple indicator outcome, these can be estimated with a multilevel structural equation model with either robust maximum likelihood or Bayesian estimation.

Again, in the situation where the mediator is teacher practices and the outcome is student achievement, the $\hat{\gamma}_{01}^{M}$ is the effect of the treatment on teacher practices, $\hat{\gamma}_{010}$ is the relationship between teacher practices and student achievement, and $\hat{\gamma}_{001}$ is the remaining effect of the treatment on student achievement remaining not due to teacher practices.

This estimation method allows for a general trait and time-specific state constructs; however, it assumes no measurement error for the estimated general trait. Essentially, the LST model estimation perfectly predicts the mediator trait. The benefit of this estimation method is that the assumptions are not as constrictive as the above two methods; however, it will converge at much higher rates than the fourth estimation method below.

Fourth estimation: fully specified model

The final estimation is the fully specified multilevel model proposed on pages 7-9:

Single indicator outcome

Level 1:

$$Y_{jkl} = \pi_{0kl} + e_{jkl}$$

Level 2:

$$M_{itkl} = \alpha_{it0l} + \lambda_{it1l}\xi_{kl}^{w} + \delta_{it2l}\zeta_{tkl}^{w} + \epsilon_{itkl}^{M}$$

$$\pi_{0kl} = \beta_{00l} + \beta_{01l}\xi_{kl}^{w} + r_{0kl}$$

Level 3:

$$\alpha_{it0l} = \hat{\alpha}_{it00} + \hat{\lambda}_{it01}\xi_{l}^{b} + \hat{\delta}_{it02}\zeta_{tl}^{b} + \upsilon_{it0l}^{M}$$

$$\xi_{l}^{b} = \hat{\gamma}_{0} + \hat{\gamma}_{1}T_{l} + \epsilon$$

$$\lambda_{it1l} = \hat{\lambda}_{it10}$$

$$\delta_{it2l} = \hat{\delta}_{it20}$$

$$\beta_{00l} = \hat{\gamma}_{000} + \hat{\gamma}_{001}T_{l} + u_{00l}$$

$$\beta_{01l} = \hat{\gamma}_{010}$$

Then, $\hat{\gamma}_1$ is the estimated a pathway, $\hat{\gamma}_{010}$ is the estimated b pathway, and $\hat{\gamma}_{001}$ is the estimated c' pathway for the mediation analysis. This can be estimated with a multilevel structural equation model with either robust maximum likelihood or Bayesian estimation.

Multiple indicator outcome

Level 1:

$$Y_{ijkl} = \alpha_{i0kl} + \lambda_{i1kl}\theta^{w}_{jkl} + \epsilon_{ijkl}$$

Level 2:

$$M_{itkl} = \alpha_{it0l} + \lambda_{it1l}\xi^{w}_{kl} + \delta_{it2l}\zeta^{w}_{tkl} + \epsilon^{M}_{itkl}$$

$$\alpha_{i0kl} = \alpha_{i00l} + \lambda_{i01l}\theta^{bk}_{kl} + \upsilon_{i0kl}$$

$$\lambda_{i1ts} = \lambda_{i10s}$$

Level 3:

$$\alpha_{it0l} = \hat{\alpha}_{it00} + \hat{\lambda}_{it01}\xi^{b}_{l} + \hat{\delta}_{it02}\zeta^{b}_{tl} + \upsilon^{M}_{it0l}$$

$$\lambda_{it1l} = \hat{\lambda}_{it10}$$

$$\delta_{it2l} = \hat{\delta}_{it20}$$

$$\alpha_{i00l} = \hat{\alpha}_{i000} + \hat{\lambda}_{i001}\theta^{bl}_{l} + \nu_{i00l}$$

$$\lambda_{i01l} = \hat{\lambda}_{i010}$$

$$\lambda_{i10l} = \hat{\lambda}_{i100}$$

$$\xi^{b}_{l} = \hat{\gamma}_0 + \hat{\gamma}_1 T_l + \epsilon$$

$$\theta^{bl}_{l} = \hat{\beta}_0 + \hat{\beta}_1\xi^{bl}_{l} + \hat{\beta}_2 T_l + \upsilon$$

Then, $\hat{\gamma}_1$ is the estimated a pathway, $\hat{\beta}_1$ is the estimated b pathway, and $\hat{\beta}_2$ is the estimated c' pathway for the mediation analysis. This can be estimated with a multilevel structural equation model with either robust maximum likelihood or Bayesian estimation.

Comparably, in the situation where the mediator is teacher practices and the outcome is student achievement, the $\hat{\gamma}_1$ is the effect of the treatment on teacher practices, $\hat{\gamma}_{010}$ is the relationship between teacher practices and student achievement, and $\hat{\gamma}_{001}$ is the remaining effect of the treatment on student achievement remaining not due to teacher practices.

This model allows measurement error to be modeled directly into the mediation effects; however, it does assume that the latent state-trait model is correctly specified for the mediator. The following simulation study investigates how these four methods perform across several specifications, including effect sizes, sample sizes, and loading sizes.

CHAPTER 3

SIMULATION STUDY

This chapter investigates the bias and power of the above four different estimations of the mediation variable under the assumption of measurement error and various conditions. Using Paxton et al. (2001)'s framework for Monte Carlo simulation studies, this chapter will estimate the four different estimations of the mediation variable: Standardized averages of the mediator across timepoints (3-2-1 mediation); factor scores for each time point averaged across timepoints (ignoring the correlation between timepoints; 3-2-1 mediation); factor scores from the LST as the mediator (3-2-1 mediation); full specified model (Multilevel SEM). The method of data generation, followed by the method for evaluating the simulation study, and then the results of the simulation study are described below.

3.1 Method

The simulations will be drawn from a multilevel structural equation model of a standard cluster randomized control trial in education research where the school is assigned the treatment, implemented at the classroom level, and the outcome is measured at the student level. This structural equation model will include a treatment indicator, items that are measured at multiple times for the mediator, and items that load onto the outcome variable.

The conditions for this simulation study for the four estimation methods varied across school-level sample sizes, mediation effect sizes, and loadings for the LTS mediation model. These varying conditions are reported in table 3.1.

The school-level sample size included 30, 60, and 200, plausible school sample sizes in educa-

Table 3.1 Simulation Varying Conditions

Parameter	Conditions
school sample size	30, 60, 200
a path (γ_1)	0.15, 0.25, 0.45
b path (β_1)	0.05, 0.15, 0.25
trait specific loadings (λ_{it10})	0.3, 0.6, 0.9
time specific loadings (δ_{it20})	0.3, 0.6, 0.9

Table 3.2 Simulation Constant Conditions

Parameter	Condition
number of teachers per school	2
number of students per teacher	30
student level variance of outcome factor ($Var(\theta^{w})$)	1
teacher random effects on outcome ($Var(\theta^{bt})$)	0.15
school random effects on outcome ($Var(\theta^{bs})$)	0.2
teacher level variance of mediator trait factor ($Var(\xi^{w})$)	1
teacher level variance of mediator time-specific factors ($Var(\zeta_t^{w})$)	1
correlation between trait and time-specific factors ($Corr(\xi^{w}, \zeta_t^{w}), Corr(\zeta_t^{w}, \zeta_t^{w})$)	0
school random effects on mediator ($Var(\xi^{b})$)	0.8
outcome loadings (λ_{i100})	1, 0.5, 0.8, 0.3
total treatment effect ($\gamma_1 \times \beta_1 + \beta_2$)	0.2
treatment/control school split	0.5
random seed	2024
number of simulation reps	100

tion efficacy research. The a path (treatment on teacher practices) of the mediation effects will vary by small effect (0.15), medium effect (0.25), and large effect (0.45; Smith and Sheridan, 2019). These are the effect sizes of the treatment on teacher practices. The b path (practices on student achievement) of the mediation effects will vary by small effect (0.05), medium effect (0.15), and large effect (0.25; Kraft, 2020). The general and time-specific latent variables loadings will vary by small, medium, and large (0.3, 0.6, 0.9). Overall, this is 324 different conditions per model estimation. With 100 replications for each condition, 32,400 different data sets were produced.

The fixed conditions of the simulation include the number of teachers per school, number of students per teacher, teacher random effects on outcome, school random effects on outcome, school random effects on the mediator, outcome loadings, total treatment effect, treatment/control schools split, random seed, number of simulation reps, outcome variance, teacher level mediator variance of trait and time-specific factors, and correlation between trait and time-specific factors. These are reported in table 3.2.

The school random effect will be fixed at 0.2, and the outcome loadings will be fixed at the items varying between 0.1 and 0.8. The total treatment effect on the outcome will be fixed at 0.2, while the percentage of treatment explained by the mediator will vary depending on the mediation effect sizes.

Data Generation

The data will be drawn from a population model that assumes the 3-2-1 LST mediation model (see pages 23-24) is the true population parameter where there are three-time points with three items each for the observed mediator and four dichotomous items for the outcome with the above varying and fixed conditions. The data generation and simulation were completed in the R package, "MplusAutomation," (Hallquist and Wiley, 2018) in combination with Mplus (Muthen and Muthen, 2017), which provided a method for the automation of data generation and estimation. The R code can be found in Appendix A.

Estimation Methods

Once the data has been generated, the four estimation methods presented in Chapter Two will be estimated on each simulated dataset. As the focus of interest, the a and b paths will be estimated separately for each method (standardized averages of mediator, pages 27-28; averaged factor scores, pages 29-30; factor scores from LST, pages 31-32; fully specified model, pages 33-34).

For the first estimation method (standardized averages of the mediator), these values are calculated by the average of the mediator at each point, averaged across the time points, and then standardized. For the second estimation method (averaged factor scores, pages 29-30), the factor scores are estimated using maximum likelihood with the variances of the factors at each time point constrained to 1 for identification. For the third estimation method (factor scores from LST, pages 31-32), the factor scores are estimated using maximum likelihood with the variances of the trait and time-specific factors constrained to 1 and the correlation between the trait and time-specific factors constrained to zero.

The pathways a and b (effect of treatment on mediator and relationship between mediator and outcome) are estimated with non-informative multilevel Bayesian methods with the Gibbs sampler (Depaoli and Clifton, 2015) with either the estimated mediators in the model or the fully specified measurement model. In the models that used estimated values of the mediator, the variance of the student-level outcome is fixed to 1 to ensure identification of the model. In the fully specified models, variances of the student-level outcome and teacher-level mediators were fixed to 1, and the correlations between the trait and time-specific mediator factors were fixed to 0 to ensure model identification. For all models, the convergence was determined by using the Mplus default settings (Asparouhouv and Muthen, 2010) where two MCMC chains are used, with the first half being discarded and the second half being used for estimating the posterior distribution and determining convergence. Convergence in this instance is determined by the Potential Scale Reduction (PSR) convergence criteria, with the default PSR for convergence set at 1.05 (Asparouhouv and Muthen, 2010). The PSR is a comparison of the within and between iteration variance of the parameters where the smaller PSR indicates smaller between iteration variance, indicating convergence.

After each method estimates each condition over 100 repetitions, it will be evaluated on the relative bias, convergence rates, and power of the a and b paths.

Evaluation Criteria

The evaluation criteria for this simulation study are relative bias, convergence rates, and power. Note that for this simulation study, neither the bias of the standard error nor the mean square error is explored. These are available upon request. Each model is evaluated on both the and b paths. The relative bias is defined as:

$$relativebias = \frac{\hat{\theta} - \theta}{\theta}$$

Where $\hat{\theta}$ is the estimated parameter of either path a or path b, and θ is the true parameter of either path a or path b.

The convergence rate is the percentage of times the model converges so that out of every 100 replications if 90 replications converge, the convergence rate would be 90%. Finally, power is the percentage of replications that estimated significant effects for a and b paths, so that similarly, if out of every 100 replications, 90 had significant a pathways, then the power for pathway a would be 0.90.

Table 3.3 Bias for 30 schools

		Pathway a																	
		S: 0.15						M:0.25						L:0.45					
		Pathway b																	
General	Time Specific	S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25	
Loading	Loading	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
Standardized Averages of the Mediator																			
0.3	0.3	-0.427	-0.926	-0.4287	-0.135	-0.431	-0.737	-0.536	-0.926	-0.537	-0.136	-0.538	-0.739	-0.609	-0.932	-0.609	-0.135	-0.588	-0.528
	0.6	-0.379	-0.074	-0.376	-0.450	-0.391	-0.539	-0.507	-0.074	-0.506	-0.450	-0.516	-0.539	-0.593	-0.080	-0.592	-0.452	-0.597	-0.523
	0.9	-0.340	-0.090	-0.387	-0.041	-0.337	-0.555	-0.484	-0.090	-0.512	-0.040	-0.482	-0.555	-0.580	-0.092	-0.596	-0.039	-0.579	-0.528
0.6	0.3	-0.063	0.392	-0.089	0.959	-0.056	-0.208	-0.197	0.392	-0.213	0.961	-0.194	-0.208	-0.288	0.382	-0.296	0.962	-0.285	-0.208
	0.6	0.023	0.496	-0.001	-0.467	-0.004	-0.241	-0.146	0.498	-0.161	-0.461	-0.162	-0.241	-0.259	0.496	-0.267	-0.460	-0.268	-0.244
	0.9	0.057	0.918	0.050	-0.533	0.027	-0.246	-0.126	0.918	-0.130	-0.537	-0.144	-0.247	-0.248	0.918	-0.250	-0.531	-0.258	-0.245
0.9	0.3	0.325	0.686	0.338	0.401	0.321	-0.160	0.154	0.698	0.163	0.401	0.154	-0.208	0.042	0.688	0.046	0.401	0.041	-0.210
	0.6	0.395	0.676	0.391	0.334	0.381	-0.175	0.197	0.676	0.195	0.331	0.188	-0.177	0.065	0.676	0.064	0.331	0.060	-0.177
	0.9	0.443	0.980	0.455	-0.231	0.398	-0.148	0.226	0.980	0.233	-0.231	0.199	-0.150	0.081	0.980	0.085	-0.231	0.066	-0.152
Averaged factor scores																			
0.3	0.3	-0.423	1.782	-0.405	0.127	-0.411	-0.150	-0.539	-0.086	-0.528	0.192	-0.530	-0.116	-0.618	2.290	-0.611	0.299	-0.582	-0.072
	0.6	-0.383	1.886	-0.372	0.157	-0.393	-0.194	-0.515	2.126	-0.509	0.216	-0.515	-0.125	-0.606	2.482	-0.600	0.325	-0.603	-0.055
	0.9	-0.360	1.820	-0.343	0.063	-0.363	-0.235	-0.501	1.996	-0.496	0.119	-0.502	-0.194	-0.598	2.346	-0.594	0.233	-0.594	-0.120
0.6	0.3	-0.095	2.260	-0.088	0.406	-0.065	0.141	-0.228	2.588	-0.222	0.509	-0.210	0.207	-0.311	3.250	-0.307	0.805	-0.312	0.256
	0.6	-0.059	2.424	-0.044	0.443	-0.047	0.064	-0.204	2.720	-0.194	0.523	-0.194	0.136	-0.302	3.298	-0.297	0.699	-0.294	0.244
	0.9	-0.029	2.426	-0.011	0.374	-0.010	0.019	-0.183	2.564	-0.172	0.432	-0.177	0.075	-0.291	3.078	-0.283	0.615	-0.283	0.181
0.9	0.3	0.265	2.456	0.248	0.433	0.251	0.104	0.106	2.730	0.096	0.510	0.098	0.170	-0.008	2.970	-0.001	0.672	-0.004	0.258
	0.6	0.274	2.318	0.296	0.449	0.295	0.104	0.113	2.592	0.127	0.538	0.122	0.146	0.007	3.266	0.016	0.711	0.013	0.246
	0.9	0.313	2.250	0.333	0.404	0.327	0.051	0.134	2.608	0.152	0.517	0.147	0.118	0.022	3.204	0.030	0.677	0.027	0.208
Factor scores of the LST Model																			
0.3	0.3	-0.511	-0.184	-0.528	-0.460	-0.523	-0.583	-0.601	-0.314	-0.604	-0.458	-0.631	-0.588	-0.652	-0.406	-0.655	-0.493	-0.627	-0.376
	0.6	-0.463	-0.054	-0.461	-0.398	-0.465	-0.546	-0.561	**-0.050**	-0.560	-0.395	-0.562	-0.554	-0.625	0.094	-0.626	-0.349	-0.630	-0.561
	0.9	-0.391	0.334	-0.394	-0.329	-0.477	-0.590	-0.526	0.150	-0.526	-0.391	-0.526	-0.586	-0.603	0.342	-0.602	-0.335	-0.608	-0.589
0.6	0.3	-0.187	0.340	-0.197	-0.161	-0.195	-0.335	-0.272	0.416	-0.277	-0.110	-0.290	-0.359	-0.346	0.234	-0.349	-0.296	-0.346	-0.365
	0.6	-0.149	0.318	-0.150	-0.175	-0.154	-0.359	-0.248	0.546	-0.250	-0.113	-0.261	-0.348	-0.330	0.304	-0.331	-0.166	-0.334	-0.300
	0.9	-0.077	0.606	-0.087	-0.167	-0.109	-0.418	-0.212	0.540	-0.218	-0.189	-0.229	-0.429	-0.310	0.370	-0.307	-0.212	-0.313	-0.454
0.9	0.3	0.158	0.604	0.159	**0.008**	0.136	-0.226	**0.044**	0.528	**0.032**	**-0.049**	**0.030**	-0.258	**-0.031**	0.132	**-0.031**	-0.132	**-0.035**	-0.216
	0.6	0.202	0.620	0.186	-0.076	0.163	-0.284	0.070	0.562	0.052	-0.106	**0.048**	-0.304	**-0.014**	0.202	**-0.018**	-0.163	**-0.025**	-0.278
	0.9	0.241	0.510	0.229	-0.152	0.204	-0.367	0.084	0.418	0.078	-0.181	0.076	-0.380	**-0.006**	0.376	**-0.002**	-0.263	**-0.011**	-0.427
Fully specified model																			
0.3	0.3	-0.920	3.546	-0.981	0.513	-1.009	-0.129	-0.964	3.592	-0.961	0.569	-0.986	-0.254	-0.976	3.150	-0.990	0.599	-0.992	**0.013**
	0.6	-0.844	2.652	-0.991	0.123	-1.046	-0.100	-0.688	2.984	-0.859	0.940	-1.013	-0.386	-1.041	3.176	-0.964	0.640	-0.999	0.116
	0.9	-0.369	1.258	-0.591	-0.403	-0.333	-0.304	-1.120	4.910	-0.742	-0.232	-0.796	-0.551	-1.854	4.406	-0.534	0.336	-1.000	-1.000
0.6	0.3	-0.880	3.352	-0.853	0.539	-0.918	-0.112	-1.091	2.766	-1.076	0.320	-0.931	-0.177	-1.027	3.378	-1.018	0.477	-1.001	-0.067
	0.6	-0.879	4.276	-1.198	0.380	-1.694	-0.160	-0.922	2.408	-0.995	0.271	-0.789	-0.365	-1.102	2.846	-0.963	0.103	-1.250	0.278
	0.9	-1.000	-1.000	-1.061	-0.061	-1.073	0.061	-0.802	3.782	-0.946	-0.277	-0.923	-0.695	-1.173	3.280	-1.111	-0.105	-0.943	-0.358
0.9	0.3	-1.000	3.166	-0.859	0.564	-1.303	-0.262	-0.931	2.936	-0.916	0.273	-1.248	-0.185	-1.067	2.664	-1.052	0.443	-0.956	-0.136
	0.6	-1.222	4.136	-1.091	0.575	-0.863	0.198	-1.268	4.668	-1.048	0.223	-0.966	-0.114	-0.989	2.578	-0.971	0.492	-1.112	-0.241
	0.9	-1.138	0.432	-2.278	0.495	-0.653	-0.054	-0.824	2.618	-1.155	0.543	-1.229	0.090	-0.921	4.872	-0.918	0.992	-1.000	-1.000

Note: Bias less than 0.05 is bolded

3.2 Results

Bias

For 30 schools, almost none of the models meets the threshold for acceptable bias ($\|bias\| \leq 0.05$). Bias was the smallest when factor scores from the LST model were used when the general loading (trait loading) was high, and the pathways were of medium or strong strength. In general, the a paths tend to be less biased than the b paths, and this is particularly true when pathway b is tiny (0.05), where the b pathways have a lot of bias. These results are reported in Table 3.3.

For 60 schools, there is less bias in general compared to 30 schools. Again, the bias is the slightest with the factor scores from the LST model with high general loadings (trait loadings), and this is true across all of the pathways of different sizes under the highest general loadings. However, there are now some unbiased estimates from both the average factor scores and the standardized

Table 3.4 Bias for 60 schools

		Pathway a																	
		S: 0.15						M:0.25						L:0.45					
		Pathway b																	
General	Time Specific	S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25	
Loading	Loading	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
Standardized Averages of the Mediator																			
0.3	0.3	-0.630	4.000	-0.639	-0.322	-0.653	-0.219	-0.658	**-0.010**	-0.664	-0.322	-0.672	-0.219	-0.677	**-0.012**	-0.680	-0.322	-0.684	-0.220
	0.6	-0.604	-0.934	-0.622	-0.484	-0.627	-0.449	-0.642	-0.934	-0.653	-0.484	-0.656	-0.449	-0.668	-0.936	-0.674	-0.484	-0.676	-0.449
	0.9	-0.565	-1.020	-0.593	-0.688	-0.597	-0.575	-0.619	-3.000	-0.636	-0.688	-0.638	-0.575	-0.655	-1.020	-0.664	-0.688	-0.666	-0.575
0.6	0.3	-0.289	0.830	-0.301	0.229	-0.316	0.159	-0.333	0.830	-0.341	0.229	-0.350	0.159	-0.363	0.830	-0.367	0.229	-0.372	0.158
	0.6	-0.251	0.582	-0.265	0.071	-0.278	**-0.006**	-0.311	0.584	-0.319	0.071	-0.327	**-0.006**	-0.350	0.584	-0.355	**0.071**	-0.359	**-0.006**
	0.9	-0.232	0.266	-0.245	-0.100	-0.245	-0.194	-0.299	0.266	-0.307	-0.100	-0.307	-0.194	-0.344	0.266	-0.348	-0.100	-0.348	-0.194
0.9	0.3	0.056	1.514	**0.046**	0.348	**0.029**	0.141	**-0.006**	1.514	**-0.012**	0.348	**-0.023**	0.141	**-0.048**	1.514	-0.051	0.348	-0.057	0.141
	0.6	0.099	1.028	0.091	0.149	0.076	**0.004**	**0.020**	1.028	**0.015**	0.149	**0.006**	**0.004**	**-0.034**	1.028	**-0.036**	0.149	**-0.041**	**0.004**
	0.9	0.127	0.398	0.112	-0.057	0.107	-0.130	**0.036**	0.398	**0.027**	-0.057	**0.024**	-0.130	**-0.024**	0.400	**-0.029**	-0.057	**-0.031**	-0.130
Averaged factor scores																			
0.3	0.3	-0.625	0.498	-0.631	0.057	-0.628	**-0.028**	-0.657	0.298	-0.662	**0.015**	-0.660	-0.067	-0.680	-0.082	-0.681	-0.146	-0.683	-0.104
	0.6	-0.603	-0.166	-0.612	-0.204	-0.623	-0.230	-0.648	-0.188	-0.650	-0.233	-0.655	-0.258	-0.674	-0.600	-0.675	-0.349	-0.677	-0.314
	0.9	-0.591	-0.416	-0.593	-0.406	-0.602	-0.403	-0.635	-0.602	-0.638	-0.465	-0.640	-0.424	-0.667	-0.804	-0.666	-0.597	-0.670	-0.524
0.6	0.3	-0.321	0.868	-0.332	0.275	-0.342	0.127	-0.358	0.608	-0.360	0.151	-0.369	0.065	-0.378	0.226	-0.381	**0.025**	-0.386	**-0.051**
	0.6	-0.288	0.328	-0.298	**0.041**	-0.308	**-0.025**	-0.337	0.092	-0.339	**-0.045**	-0.348	-0.094	-0.369	-0.318	-0.370	-0.202	-0.374	-0.215
	0.9	-0.269	-0.060	-0.273	-0.166	-0.263	-0.071	-0.324	-0.294	-0.326	-0.222	-0.322	-0.118	-0.360	-0.900	-0.362	-0.412	-0.365	-0.355
0.9	0.3	**-0.010**	0.954	**-0.025**	0.219	**-0.019**	0.182	**-0.048**	0.702	-0.056	0.143	-0.062	**0.047**	-0.074	0.168	-0.076	-0.055	-0.079	-0.091
	0.6	**0.029**	0.636	**0.019**	0.087	**0.028**	0.068	**-0.023**	0.344	**-0.031**	**0.009**	**-0.027**	**0.013**	-0.061	-0.138	-0.063	-0.172	-0.066	-0.224
	0.9	0.057	0.326	**0.049**	-0.063	0.061	**-0.044**	**-0.006**	0.158	**-0.014**	-0.155	**-0.005**	-0.087	-0.051	-0.526	-0.053	-0.334	-0.056	-0.296
Factor scores of the LST Model																			
0.3	0.3	-0.661	2.612	-0.652	0.621	-0.669	0.255	-0.680	2.390	-0.682	0.573	-0.686	0.215	-0.696	1.996	-0.697	0.431	-0.698	0.149
	0.6	-0.617	2.286	-0.603	0.369	-0.618	**0.034**	-0.656	2.326	-0.657	0.421	-0.654	**0.005**	-0.675	1.878	-0.679	0.310	-0.678	-0.058
	0.9	-0.579	1.634	-0.579	0.095	-0.567	-0.220	-0.627	1.438	-0.630	0.053	-0.622	-0.242	-0.657	1.440	-0.658	**0.021**	-0.660	-0.244
0.6	0.3	-0.325	2.586	-0.333	0.755	-0.341	0.398	-0.360	2.284	-0.363	0.667	-0.368	0.346	-0.382	1.094	-0.387	0.317	-0.386	0.225
	0.6	-0.311	2.124	-0.313	0.475	-0.305	0.185	-0.338	1.968	-0.344	0.433	-0.346	0.134	-0.371	0.888	-0.375	0.114	-0.374	**0.047**
	0.9	-0.265	1.456	-0.269	0.165	-0.251	-0.077	-0.311	1.566	-0.329	0.179	-0.317	-0.088	-0.358	0.756	-0.357	0.059	-0.358	-0.162
0.9	0.3	**-0.003**	1.938	**-0.007**	0.514	**0.006**	0.257	**-0.031**	1.684	**-0.038**	0.438	**-0.040**	0.202	-0.068	0.658	-0.072	0.125	-0.070	0.064
	0.6	**0.019**	1.644	**0.018**	0.348	**0.033**	0.116	**-0.017**	1.460	**-0.024**	0.293	**-0.024**	0.065	-0.061	0.578	-0.064	**0.014**	-0.062	**-0.044**
	0.9	**0.049**	1.266	**0.043**	0.137	0.071	-0.083	**-0.012**	1.116	**-0.014**	0.110	**-0.002**	-0.101	**-0.047**	0.810	**-0.049**	**-0.010**	**-0.049**	-0.180
Fully specified model																			
0.3	0.3	-0.948	1.596	-1.000	-0.135	-0.940	-0.459	-0.973	1.724	-0.969	-0.105	-0.991	-0.465	-0.943	1.756	-0.968	-0.123	-0.972	-0.463
	0.6	-1.069	1.830	-0.873	-0.038	-1.101	-0.525	-0.765	1.620	-0.787	-0.189	-1.035	-0.426	-1.081	2.252	-0.930	-0.072	-0.956	-0.460
	0.9	-0.830	1.070	-0.853	-0.561	-1.028	-0.668	-0.930	2.808	-0.738	-0.104	-0.131	-0.441	-0.750	1.840	-1.000	-1.000	-1.000	-1.000
0.6	0.3	-0.883	1.538	-1.133	-0.109	-1.025	-0.442	-0.986	1.624	-0.960	-0.149	-1.009	-0.498	-0.964	1.694	-1.047	-0.052	-0.984	-0.460
	0.6	-1.058	1.734	-1.220	-0.267	-1.279	-0.506	-1.080	1.512	-1.232	-0.157	-0.976	-0.357	-1.040	2.102	-0.927	-0.089	-0.892	-0.419
	0.9	-0.911	1.714	-0.573	0.003	-1.279	-0.358	-0.472	1.532	-0.828	-0.131	-1.171	-0.136	-0.914	1.892	-0.704	0.288	-0.943	-0.588
0.9	0.3	-0.827	1.576	-0.869	-0.080	-0.981	-0.419	-1.064	1.536	-1.062	-0.097	-0.951	-0.436	-0.999	1.660	-1.024	-0.151	-0.990	-0.474
	0.6	-1.286	1.464	-1.024	-0.137	-0.832	-0.515	-0.859	1.834	-0.860	-0.192	-1.022	-0.375	-1.099	1.916	-0.952	-0.115	-1.069	-0.390
	0.9	-0.641	2.790	-0.792	-0.055	-1.385	-0.476	-1.087	0.596	-1.526	**0.023**	-1.302	-0.330	-0.534	1.042	-1.384	-0.093	-1.138	-0.176

Note: Bias less than 0.05 is bolded

average of the mediator with high general loadings. In the standardized averages of the mediator, these are in the medium and high pathways and for the average factor scores, these tend to be under the small to medium pathways. Again, pathway a tends to be less biased than pathway b. In general, the bias of the b pathway decreases as the magnitude of the true b pathway increases, which is not consistently true across the a path. These results are reported in Table 3.4.

For 200 schools, the bias is comparable to the 60 schools. Here, the bias is most negligible in both the averaged factor scores and the factor scores from the latent state-trait model under the highest trait loadings. In the averaged factor scores, only the pathway a is unbiased under certain conditions whereas the factor scores from the LST model also had unbiased pathways b under certain conditions. These results are reported in Table 3.5.

Across all of the sample sizes and conditions, the fully specified model was never unbiased, a

Table 3.5 Bias for 200 schools

		Pathway a																	
		S: 0.15						M:0.25						L:0.45					
		Pathway b																	
General	Time Specific	S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25	
Loading	Loading	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
Standardized Averages of the Mediator																			
0.3	0.3	-0.697	1.310	-0.716	0.834	-0.733	0.726	-0.698	1.362	-0.709	0.883	-0.720	0.726	-0.699	1.310	-0.705	0.834	-0.711	0.725
	0.6	-0.684	0.360	-0.699	0.266	-0.709	0.240	-0.691	0.396	-0.699	0.302	-0.706	0.239	-0.695	0.376	-0.666	1.264	-0.704	0.251
	0.9	-0.673	-0.210	-0.685	-0.167	-0.695	-0.157	-0.684	-0.126	-0.691	-0.141	-0.697	-0.157	-0.691	-0.154	-0.695	-0.149	-0.698	-0.146
0.6	0.3	-0.395	1.068	-0.416	0.581	-0.433	0.484	-0.397	1.128	-0.409	0.621	-0.419	0.488	-0.398	1.086	-0.405	0.589	-0.411	0.490
	0.6	-0.376	0.552	-0.393	0.328	-0.407	0.252	-0.386	0.592	-0.396	0.363	-0.404	0.253	-0.392	0.592	-0.398	0.341	-0.402	0.260
	0.9	-0.359	0.168	-0.375	**0.044**	-0.387	**0.007**	-0.376	0.196	-0.385	0.072	-0.393	0.011	-0.387	0.190	-0.392	0.059	-0.396	**0.016**
0.9	0.3	-0.095	0.626	-0.111	0.210	-0.126	0.132	-0.097	0.656	-0.107	0.239	-0.116	0.132	-0.098	0.592	-0.103	0.192	-0.108	0.121
	0.6	-0.070	0.400	-0.085	0.102	**0.029**	0.062	-0.082	0.428	-0.090	0.121	-0.099	**0.026**	-0.090	0.398	-0.095	0.102	-0.099	**0.026**
	0.9	**-0.050**	0.140	-0.065	-0.047	-0.075	-0.102	-0.070	0.164	-0.079	**-0.047**	-0.086	-0.102	-0.084	0.098	-0.089	**-0.039**	**-0.044**	0.082
Averaged factor scores																			
0.3	0.3	-0.642	2.778	-0.633	1.409	-0.627	1.122	-0.666	3.168	-0.661	1.541	-0.656	1.177	-0.683	3.778	-0.680	1.682	-0.676	1.296
	0.6	-0.633	1.496	-0.626	0.693	-0.620	0.530	-0.660	1.786	-0.657	0.817	-0.654	0.566	-0.680	2.284	-0.653	1.668	-0.675	0.657
	0.9	-0.625	0.646	-0.621	0.119	-0.617	**0.026**	-0.656	0.844	-0.654	0.206	-0.650	0.062	-0.677	1.178	-0.676	0.283	-0.674	0.132
0.6	0.3	-0.336	1.960	-0.329	0.905	-0.323	0.690	-0.364	2.294	-0.360	1.037	-0.356	0.741	-0.383	2.806	-0.380	1.1867	-0.378	0.859
	0.6	-0.317	1.440	-0.313	0.603	-0.307	0.442	-0.353	1.712	-0.350	0.711	-0.348	0.487	-0.377	2.172	-0.374	0.8547	-0.372	0.594
	0.9	-0.308	0.924	-0.303	0.294	-0.299	0.170	-0.346	1.164	-0.345	0.387	-0.342	0.212	-0.372	1.556	-0.372	0.5057	-0.370	0.294
0.9	0.3	**-0.035**	1.230	**-0.031**	0.415	**-0.027**	0.251	-0.063	1.532	-0.060	0.530	-0.059	0.294	-0.082	1.940	-0.081	0.631	-0.080	0.380
	0.6	**-0.012**	1.014	**-0.008**	0.295	0.152	0.304	**-0.050**	1.276	**-0.047**	0.403	**-0.045**	0.192	-0.075	1.660	-0.073	0.501	-0.072	0.271
	0.9	**0.004**	0.728	**0.007**	0.144	**0.011**	**0.020**	**-0.039**	0.960	**-0.038**	0.209	**-0.036**	0.061	-0.070	1.340	-0.068	0.341	**-0.022**	0.328
Factor scores of the LST Model																			
0.3	0.3	-0.642	2.778	-0.633	1.409	-0.627	1.122	-0.767	1.658	-0.771	0.793	-0.777	0.584	-0.740	1.092	-0.742	0.586	-0.745	0.496
	0.6	-0.633	1.496	-0.646	0.759	-0.620	0.530	-0.760	0.634	-0.764	**0.037**	-0.766	-0.102	-0.736	0.326	-0.736	0.816	-0.739	-0.156
	0.9	-0.625	0.646	-0.621	0.119	-0.617	**0.026**	-0.758	**0.014**	-0.759	-0.380	-0.760	-0.471	-0.734	-0.150	-0.734	-0.444	-0.735	-0.502
0.6	0.3	-0.336	1.960	-0.329	0.905	-0.323	0.690	-0.463	1.044	-0.469	0.509	-0.473	0.371	-0.438	0.614	-0.441	0.343	-0.444	0.296
	0.6	-0.317	1.440	-0.313	0.603	-0.307	0.442	-0.460	0.540	-0.464	0.109	-0.467	**-0.000**	-0.436	0.232	-0.437	**-0.007**	-0.440	-0.058
	0.9	-0.308	0.924	-0.303	0.294	-0.299	0.170	-0.458	0.108	-0.461	-0.227	-0.462	-0.308	-0.434	-0.128	-0.435	-0.313	-0.437	-0.350
0.9	0.3	**-0.035**	1.230	**-0.031**	0.415	**-0.027**	0.251	-0.159	0.486	-0.164	0.145	-0.168	**0.050**	-0.135	0.172	-0.137	**0.031**	-0.140	**-0.002**
	0.6	**-0.012**	1.014	**-0.008**	0.295	0.152	0.304	-0.158	0.272	-0.162	**-0.041**	-0.164	-0.123	-0.134	**0.020**	-0.136	-0.135	-0.138	-0.168
	0.9	**0.004**	0.728	**0.007**	0.144	**0.011**	**0.022**	-0.157	**0.026**	-0.160	-0.260	-0.162	-0.310	-0.133	-0.186	-0.135	-0.315	-0.116	-0.077
Fully specified model																			
0.3	0.3	-1.077	1.368	-1.013	-0.209	-1.173	-0.530	-1.020	1.314	-1.043	-0.218	-1.067	-0.543	-0.980	1.324	-0.956	-0.229	-0.968	-0.537
	0.6	-1.359	1.454	-0.081	-0.229	-0.761	-0.518	-0.945	1.318	-1.066	-0.246	-0.739	-0.518	-0.966	1.374	-0.965	**0**	-1.163	-0.528
	0.9	-1.167	1.534	-1.859	-0.223	-1.184	-0.557	-0.433	1.374	-1.169	-0.244	-1.811	-0.526	-1.083	1.494	-0.928	-0.270	-0.342	-0.478
0.6	0.3	-1.084	1.346	-1.041	-0.228	-1.200	-0.531	-1.030	1.348	-1.054	-0.225	-1.068	-0.528	-0.890	1.286	-0.955	-0.221	-0.992	-0.514
	0.6	-1.797	1.304	-0.912	-0.223	-1.065	-0.527	-0.788	1.272	-0.951	-0.190	-1.183	-0.552	-0.675	1.342	-0.954	-0.223	-0.993	-0.528
	0.9	-1.722	1.468	-1.379	-0.088	-1.273	-0.551	-0.382	1.228	0.973	-0.231	-0.168	-0.633	-1.080	1.324	-0.885	-0.255	-1.465	-0.507
0.9	0.3	-1.280	1.288	-1.039	-0.217	-1.227	-0.528	-1.049	1.314	-0.997	-0.220	-0.934	-0.517	-1.025	1.324	-1.042	-0.201	-0.957	-0.530
	0.6	-0.735	1.416	-0.665	-0.218	–	–	-1.244	1.426	-0.878	-0.195	-0.982	-0.539	-0.859	1.358	-0.909	-0.247	-0.818	-0.526
	0.9	-0.747	1.248	-1.177	-0.291	-1.494	-0.566	-0.253	1.446	-0.660	-0.277	-0.671	-0.510	-1.132	1.234	-0.617	-0.081	–	–

surprising finding. However, when investigated, this was due to the between level 3 trait loadings being 0 instead of the population parameter. This may result from the level 2 sample size or the level 2 interclass correlation coefficients. Changes in these parameters should be investigated in future work to evaluate this unforeseen issue. Additionally, Using a weakly informed prior for the random variances at level 2 and level 3 may decrease bias through decreasing issues in the variance-covariance matrix, which should also be investigated in future work.

An important takeaway from the results of the bias across the different school sample sizes is the importance of solid measures for the mediator, measures explicitly that load strongly onto the trait factor of the mediator. When researching questions related to mediators with trait-state properties, investigators should invest a decent amount of time and resources into ensuring that the measure is able to measure the trait factor of the mediator. Then, depending on the number of schools and the strength of the mediation measure, the investigator should consider using the factor scores from the LST model (with a smaller number of schools).

Convergence

Across all school sample sizes, the standardized averages, averaged factor scores, and factor scores from the LST model converged at 100%. This indicates that researchers should not have convergence problems when using these estimation methods. At 30 schools, the fully specified model converged more than 70% of the time only under one condition (small general loading, small specific loadings, and medium a and b pathways). In general, convergence decreased as both the general and specific loadings increased. There is no general trend in convergence dependent on the magnitude of the true a or b pathways. Results are reported in Table 3.6.

At 60 schools, the fully specified model converged at rates higher than 70% when the general and specific loadings were small across all pathway magnitudes, as well as with a medium general loading and small specific loadings with medium and small pathways a. In general, convergence was higher with a larger school sample, and the trend continued to decrease as the general and specific loadings increased. Results are reported in Table 3.7.

At 200 schools, convergence rates were higher than at 30 or 60 schools. Convergence rates met the threshold of 70% under small and medium trait loadings with small state loadings. In general, convergence rates decreased as factor loadings increased with no pattern across the a and b pathways. Results are reported in Table 3.8. In general, when considering using the fully specified model, convergence will occur with larger sample sizes and smaller factor loadings. These convergence issues may be due to similar problems causing bias, such as the lack of variance in the level 3 mediator. In the future, this should be investigated with more conditions.

Under this current simulation study, with 200 schools or fewer and the recommendation to use factor scores from the LST under the bias conditions, factor scores from the LST converged under all conditions, indicating that there would likely be no convergence issues under this estimation method.

Table 3.6 Convergence for 30 schools

		Pathway a								
		S: 0.15			M:0.25			L:0.45		
General	Time Specific	Pathway b								
Loading	Loading	S:0.05	M:0.15	L:0.25	S:0.05	M:0.15	L:0.25	S:0.05	M:0.15	L:0.25
Standardized Averages of the Mediator										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Averaged factor scores										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Factor scores of the LST Model										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Fully specified model										
0.3	0.3	0.66	0.66	0.69	0.6	0.7	0.65	0.65	0.61	0.68
	0.6	0.15	0.12	0.14	0.16	0.12	0.15	0.11	0.15	0.11
	0.9	0.02	0.02	0.02	0.03	0.02	0.01	0.02	0.03	0
0.6	0.3	0.54	0.52	0.55	0.49	0.49	0.41	0.47	0.52	0.54
	0.6	0.15	0.16	0.08	0.27	0.12	0.14	0.13	0.14	0.08
	0.9	0	0.01	0.01	0.01	0.03	0.02	0.02	0.02	0.05
0.9	0.3	0.42	0.39	0.33	0.39	0.42	0.31	0.3	0.32	0.35
	0.6	0.05	0.07	0.07	0.12	0.15	0.09	0.11	0.09	0.09
	0.9	0.01	0.04	0.02	0.04	0.03	0.02	0.02	0.04	0

Table 3.7 Convergence for 60 schools

General Loading	Time Specific Loading	Pathway a: S: 0.15, Pathway b: S:0.05	M:0.15	L:0.25	Pathway a: M:0.25, Pathway b: S:0.05	M:0.15	L:0.25	Pathway a: L:0.45, Pathway b: S:0.05	M:0.15	L:0.25
Standardized Averages of the Mediator										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Averaged factor scores										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Factor scores of the LST Model										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Fully specified model										
0.3	0.3	0.88	0.85	0.86	0.86	0.87	0.89	0.83	0.90	0.93
	0.6	0.22	0.19	0.14	0.21	0.12	0.22	0.21	0.14	0.18
	0.9	0.05	0.03	0.06	0.02	0.03	0.02	0.01	0	0
0.6	0.3	0.66	0.69	0.73	0.75	0.75	0.73	0.68	0.61	0.74
	0.6	0.16	0.14	0.24	0.20	0.18	0.2	0.21	0.19	0.22
	0.9	0.05	0.02	0.04	0.02	0.04	0.04	0.05	0.04	0.03
0.9	0.3	0.47	0.49	0.47	0.45	0.46	0.56	0.53	0.54	0.56
	0.6	0.20	0.13	0.18	0.19	0.12	0.18	0.18	0.15	0.15
	0.9	0.04	0.03	0.02	0.03	0.05	0.04	0.02	0.03	0.02

Table 3.8 Convergence for 200 schools

		Pathway a								
		S: 0.15			M:0.25			L:0.45		
General	Time Specific	Pathway b								
Loading	Loading	S:0.05	M:0.15	L:0.25	S:0.05	M:0.15	L:0.25	S:0.05	M:0.15	L:0.25
Standardized Averages of the Mediator										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Averaged factor scores										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Factor scores of the LST Model										
0.3	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.6	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
0.9	0.3	1	1	1	1	1	1	1	1	1
	0.6	1	1	1	1	1	1	1	1	1
	0.9	1	1	1	1	1	1	1	1	1
Fully specified model										
0.3	0.3	0.90	0.89	0.85	0.83	0.87	0.83	0.92	0.88	0.93
	0.6	0.20	0.18	0.28	0.23	0.21	0.23	0.19	0.20	0.29
	0.9	0.02	0.04	0.05	0.03	0.02	0.04	0.04	0.02	0.01
0.6	0.3	0.72	0.65	0.74	0.68	0.75	0.72	0.63	0.71	0.65
	0.6	0.21	0.22	0.25	0.17	0.24	0.25	0.17	0.23	0.23
	0.9	0.04	0.03	0.02	0.04	0.03	0.01	0.05	0.05	0.02
0.9	0.3	0.4	0.54	0.51	0.42	0.45	0.34	0.43	0.48	0.35
	0.6	0.10	0.18	0	0.20	0.12	0.17	0.15	0.12	0.11
	0.9	0.01	0.01	0.05	0.03	0.02	0.04	0.01	0.01	0

Table 3.9 Power for 30 schools

		Pathway a																	
		S: 0.15						M:0.25						L:0.45					
		Pathway b																	
General	Time Specific	S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25	
Loading	Loading	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
Standardized Averages of the Mediator																			
0.3	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0.6	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0.011	0	0	0	0	0	0.011	0	0	0	0	0	0.011	0	0	0	0
	0.9	0	0.011	0	0.011	0	0	0	0.011	0	0.011	0	0	0.011	0.011	0	0.011	0.011	0
0.9	0.3	0	0.022	0	0.022	0	0.022	0.011	0.022	0	0.022	0	0.022	0.022	0.022	0.022	0.022	0.022	0.022
	0.6	0.011	0.011	0	0.022	0.011	0.033	0.011	0.011	0.011	0.022	0.011	0.033	0.032	0.011	0.033	0.022	0.033	0.033
	0.9	0.011	0.011	0.011	0.011	0.011	0.022	0.011	0.011	0.011	0.011	0.011	0.022	0.043	0.011	0.033	0.011	0.033	0.022
Averaged factor scores																			
0.3	0.3	0	0	0	0	0	0	0	0.021	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.9	0.3	0	0.021	0	0.021	0	0.021	0	0.021	0	0.021	0	0.021	0.031	0.021	0.031	0.010	0.031	0.010
	0.6	0	0.010	0	0.010	0	0.031	0.010	0.010	0.010	0.010	0	0.021	0.021	0.010	0.021	0.021	0.031	0.010
	0.9	0.010	0.010	0.010	0.010	0.010	0.021	0.010	0.010	0.010	0.010	0.010	0.021	0.042	0.010	0.031	0.010	0.031	0.021
Factor scores of the LST Model																			
0.3	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.9	0.3	0	0.010	0	0.010	0	0	0	0.010	0	0	0	0	0.010	0.010	0.010	0	0.010	0.010
	0.6	0	0.010	0	0.010	0	0.010	0	0.010	0	0.010	0	0.010	0.030	0.010	0.030	0	0.030	0.010
	0.9	0	0.010	0	0.020	0	0.020	0.010	0.010	0.010	0.020	0.010	0.020	0.052	0.010	0.051	0.020	0.051	0.010
Fully specified model																			
0.3	0.3	0.00	0.030	0	0.015	0	0.029	0	0.017	0	0.014	0	0	0	0.015	0	0.033	0	0.058
	0.6	0	0	0	0	0	0.071	0	0	0	0	0	0	0	0	0	0	0	0.091
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6	0.3	0	0.019	0	0.019	0	0.382	0	0	0	0	0	0	0	0	0	0.038	0	0.019
	0.6	0	0.067	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.9	0.3	0	0	0	0.026	0	0	0	0	0	0	0	0.032	0	0	0	0	0	0
	0.6	0	0	0	0	0	0.143	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Power

Across all three school sample sizes, no power of 80% was ever achieved for any of the models or estimated pathways. For 200 schools, power increased substantially for the a pathway when the a pathway is 0.45 and the trait loadings were high when using the standardized averages, averaged factor scores, and the factor scores of the LST model. As the state loadings increased, the power also increased within the high trait loadings. Power remained low in the fully specified model. These results are reported in Tables 3.9-3.11.

In general, when researchers hope to answer research questions that may use this LST multilevel mediation model, power will not reach adequate levels with 200 schools or with effect sizes less than or equal to 0.45. To increase power, researchers may consider increasing the sample size, adjusting

Table 3.10 Power for 60 schools

		Pathway a																	
		S: 0.15						M:0.25						L:0.45					
		Pathway b																	
General	Time Specific	S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25	
Loading	Loading	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
Standardized Averages of the Mediator																			
0.3	0.3	0	0.055	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0.011	0	0.011	0	0.011	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0.011	0	0.011	0	0.011	0
0.9	0.3	0.011	0	0	0.011	0	0.032	0.011	0	0.011	0.011	0.011	0.032	0.065	0	0.064	0.011	0.064	0.032
	0.6	0.011	0	0.011	0	0.011	0.043	0.022	0	0.021	0	0.021	0.043	0.065	0	0.064	0	0.064	0.043
	0.9	0.021	0	0.032	0.011	0.021	0.021	0.021	0	0.021	0.011	0.021	0.021	0.064	0	0.064	0.011	0.064	0.021
Averaged factor scores																			
0.3	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0.011	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0.021	0	0.021	0	0.021	0
0.9	0.3	0	0.011	0	0.021	0	0.021	0.021	0.011	0.021	0.010	0.021	0.021	0.053	0.021	0.053	0.021	0.053	0.011
	0.6	0.021	0	0.021	0.010	0.021	0.031	0.032	0	0.031	0.010	0.031	0.010	0.074	0.021	0.074	0.011	0.074	0.011
	0.9	0.021	0	0.021	0.010	0.021	0.021	0.031	0	0.031	0.010	0.031	0.021	0.074	0	0.063	0.011	0.062	0.010
Factor scores of the LST Model																			
0.3	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0.01	0	0.01	0	0.011	0	0.01	0	0.01	0	0.011	0	0.011	0	0.011	0	0.011
0.6	0.3	0	0	0	0	0	0.011	0	0	0	0	0	0.011	0	0.010	0	0.010	0	0.011
	0.6	0	0	0	0	0	0.032	0	0	0	0.011	0	0.032	0	0.01	0	0.010	0	0.021
	0.9	0	0.010	0	0.031	0	0.032	0	0.021	0	0.031	0	0.032	0.031	0.021	0.032	0.032	0.032	0.043
0.9	0.3	0	0.021	0	0.021	0	0.032	0.021	0.021	0.021	0.021	0.021	0.032	0.042	0.031	0.042	0.042	0.043	0.032
	0.6	0.01	0.03	0	0.042	0	0.043	0.021	0.032	0.021	0.032	0.032	0.043	0.062	0.042	0.062	0.052	0.064	0.043
	0.9	0.021	0.042	0.021	0.031	0.021	0.043	0.032	0.042	0.031	0.042	0.032	0.043	0.095	0.053	0.084	0.063	0.085	0.043
Fully specified model																			
0.3	0.3	0	0	0	0	0	0	0	0	0	0	0	0.011	0	0.012	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.9	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

to larger minimum detectable effect sizes for the a and b pathways, or using weakly informative priors in the Bayesian estimation.

Table 3.11 Power for 200 schools

		Pathway a																	
		S: 0.15						M:0.25						L:0.45					
		Pathway b																	
General	Time Specific	S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25		S:0.05		M:0.15		L:0.25	
Loading	Loading	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
Standardized Averages of the Mediator																			
0.3	0.3	0	0	0	0.03	0	0.03	0	0.01	0	0.02	0	0.03	0	0	0	0.03	0	0.03
	0.6	0	0.03	0	0.03	0	0.04	0	0.03	0	0.03	0	0.04	0	0.03	0	0	0	0.04
	0.9	0	0.03	0	0.04	0	0.05	0	0.03	0	0.04	0	0.05	0	0.03	0	0.04	0	0.05
0.6	0.3	0	0.08	0	0.07	0	0.11	0	0.08	0	0.06	0	0.11	0.02	0.08	0.02	0.07	0.02	0.11
	0.6	0	0.05	0	0.06	0	0.1	0	0.05	0	0.06	0	0.1	0.04	0.05	0.03	0.06	0.03	0.10
	0.9	0	0.05	0	0.06	0	0.07	0.01	0.05	0.01	0.06	0	0.07	0.08	0.05	0.08	0.06	0.08	0.07
0.9	0.3	0.01	0.087	0.01	0.11	0.01	0.12	0.03	0.08	0.03	0.11	0.03	0.12	0.46	0.08	0.43	0.11	0.43	0.12
	0.6	0.01	0.06	0.01	0.10	0	0.086	0.04	0.06	0.04	0.10	0.03	0.13	0.49	0.06	0.47	0.11	0.45	0.13
	0.9	0.01	0.06	0.01	0.09	0.01	0.13	0.06	0.06	0.06	0.09	0.07	0.13	0.51	0.06	0.50	0.08	0.63	0.11
Averaged factor scores																			
0.3	0.3	0	0	0	0.01	0	0.02	0	0	0	0.02	0	0.03	0	0.01	0	0.03	0	0.06
	0.6	0	0.01	0	0.03	0	0.04	0	0.02	0	0.03	0	0.04	0	0.03	0	0.053	0	0.05
	0.9	0	0.02	0	0.03	0	0.04	0	0.03	0	0.03	0	0.04	0	0.04	0	0.04	0	0.07
0.6	0.3	0	0.06	0	0.09	0	0.12	0	0.06	0	0.09	0	0.12	0.03	0.09	0.03	0.09	0.02	0.14
	0.6	0	0.05	0	0.08	0	0.11	0	0.05	0	0.08	0	0.12	0.05	0.08	0.05	0.09	0.06	0.11
	0.9	0	0.05	0	0.07	0	0.09	0	0.06	0	0.08	0	0.11	0.08	0.09	0.08	0.08	0.08	0.10
0.9	0.3	0.01	0.07	0.01	0.09	0.01	0.12	0.03	0.09	0.03	0.09	0.03	0.12	0.48	0.11	0.48	0.10	0.48	0.15
	0.6	0.01	0.09	0.01	0.09	0	0.083	0.05	0.11	0.05	0.10	0.05	0.11	0.52	0.12	0.52	0.11	0.052	0.13
	0.9	0.01	0.07	0.01	0.09	0.01	0.120	0.08	0.08	0.08	0.10	0.08	0.13	0.57	0.09	0.57	0.11	0.679	0.179
Factor scores of the LST Model																			
0.3	0.3	0	0	0	0.01	0	0.02	0	0.03	0	0.03	0	0.05	0	0.04	0	0.03	0	0.06
	0.6	0	0.01	0	0.03	0	0.04	0	0.06	0	0.08	0	0.08	0	0.06	0	0.05	0	0.09
	0.9	0	0.02	0	0.03	0	0.04	0	0.08	0	0.08	0	0.09	0	0.07	0	0.08	0	0.09
0.6	0.3	0	0.06	0	0.09	0	0.12	0	0.05	0	0.08	0	0.12	0.02	0.08	0.02	0.10	0.02	0.13
	0.6	0	0.05	0	0.08	0	0.11	0	0.06	0	0.1	0	0.12	0.04	0.08	0.04	0.13	0.04	0.13
	0.9	0	0.05	0	0.07	0	0.09	0.01	0.08	0.01	0.10	0.01	0.13	0.06	0.07	0.05	0.10	0.05	0.13
0.9	0.3	0.01	0.07	0.01	0.09	0.01	0.12	0.03	0.08	0.03	0.13	0.02	0.13	0.42	0.10	0.41	0.13	0.39	0.13
	0.6	0.01	0.09	0.01	0.09	0	0.083	0.04	0.09	0.04	0.12	0.04	0.13	0.43	0.11	0.43	0.13	0.42	0.13
	0.9	0.01	0.07	0.01	0.09	0.01	0.12	0.05	0.10	0.04	0.12	0.05	0.13	0.44	0.11	0.44	0.13	0.429	0.179
Fully specified model																			
0.3	0.3	0	0	0	0	0	0.012	0	0	0	0.011	0	0	0	0	0	0	0	0
	0.6	0	0.05	0	0	0	0.036	0	0.043	0	0	0	0.043	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.6	0.3	0	0.014	0	0	0	0	0	0.015	0	0	0	0	0	0.016	0	0.014	0	0
	0.6	0	0	0	0	0	0.04	0	0	0	0.042	0	0	0	0	0	0	0	0.043
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.9	0.3	0	0	0	0.019	0	0	0	0.024	0	0	0	0.029	0	0	0	0	0	0
	0.6	0	0	0	0	–	–	0	0	0	0	0	0	0	0	0	0	0	0
	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	–	–

CHAPTER 4

EMPIRICAL STUDY

Following the simulation study, data from the Crafting Engaging Science Environments (CESE) intervention is analyzed and compared using the same four estimation models. Data from the CESE intervention include an intervention indicator, items from a summative assessment, and items for the teacher observations measured at multiple time points with different raters during the intervention as a mediating effect for the CESE intervention.

The data for this is from the evaluation of the intervention tested in Michigan and California in 2018-2019. The overall treatment effect and preliminary mediation analyses were reported in Schneider et al. (2022). For the treatment effect and the mediation analysis, hierarchical linear models were used with an equated (between chemistry and physics) standardized score as the outcome and controlling for the pre-test with an estimated factor score. The two mediators explored were a composite score of the teacher's self-reported use of project-based learning measured from their exit survey post-treatment and the students' reporting of modeling from their exit survey. Schneider et al. (2022) found that teachers' incorporation of PBL did not significantly mediate the treatment effect; however, students' use of modeling did mediate the treatment effect (at a significance level of 0.10) and accounted for about 28% of the treatment. This empirical study reanalyzes the teacher use of project-based learning using longitudinal observation scores instead of the post-treatment self-reported scores. This study first explains the methods (including describing the sample, the measures, the assumed mediation model, and the estimation methods used) and then reports and compares the results of the mediation analysis across the four different estimation methods.

4.1 Method

As noted previously, this study is the result of a cluster randomized control trial on the efficacy of the CESE intervention. Before the start of the study in 2018, the 70 schools participating in the study were randomized into the treatment and control conditions (within four different regional blocks). Preceding the start of the intervention, 36 treatment and 34 control schools were used. Table 4.1

gives the final analytic sample of 61 schools. At the beginning of the intervention, the treatment teachers were provided with three days of professional learning on the CESE curriculum, and the control teachers were provided with one day of professional learning on the Next Generation of Science Standards (NGSS). Throughout the academic year, the CESE intervention included three project-based learning, NGSS-aligned units in either chemistry or physics, along with end-of-unit formative assessments. Through the professional learning, the provided units, and the formative assessments as a holistic intervention, it was expected that teacher use of project-based learning would increase along with NGSS-aligned teaching, which would increase student science achievement. Schneider et al. (2022) give evidence that the intervention was effective and increased student achievement by about 0.20 standard deviations. The next question is whether the treatment increased teacher use of project-based learning in their classroom and whether that increase mediates the intent-to-treat treatment effects.

Sample

In its entirety, during the intervention time, the following data were collected: teacher background survey, student background survey, student pretest, student Experience Sampling Method (ESM) surveys, teacher ESM surveys, teacher observations, treatment student unit assessments, teacher exit survey, student exit survey, student summative assessment, and school-linked data from the Common Core of Data. Table 4.1 gives the number of beeps/observations, students, teachers, and schools for all the different data collection points, as well as the analytic sample used to test the main treatment effect.

The final analytic sample included 4,238 students in 102 teachers in 61 schools. Of these 102 teachers, 55 had one or more observations in 38 schools, just over half of the analytic sample. This reduction in sample size may cause the sample to be underpowered for both the treatment and mediating effects. Regardless of whether the sample is powered, these data will still be used to estimate all models that converge.

Table 4.1 Data collection for Students, Teachers, and Schools

Data source	Beep/observation	Student	Teacher	School
Student Background		6694	119	67
Teacher Background			115	66
Student Pretest		6720	118	66
Student ESM	7009	546	27	21
Teacher ESM	273		28	21
Teacher Observations	108		55	38
Teacher Exit Survey			107	63
Student Exit Survey		5435	103	59
Student Summative Assessment		5977	107	62
Linked CCD				69
Analytic Sample		4238	102	61

Measures

The CESE observation protocol

During the CESE intervention, a subsample of teachers was observed 1-5 times throughout the three units by 14 different observers. The instrument used for the observation had 13 items related to teacher actions/behaviors and seven items related to student actions/behaviors. Each item was scored on a scale from 1 to 4, with one indicating that a behavior or practice was not observed and four indicating that the practice was observed entirely. For each item, the observer must give a justification for the scores. The ten teacher items can be divided into four different constructs. The first construct is related to teacher PBL practices and includes four items. The second construct is the teacher's support of social and emotional learning. The third construct is direct measures of implementation fidelity with two items, and the final construct is classroom management with four items. Table 4.2 reports the items and their constructs.

The CESE pretest

The CESE pretest was developed from 12 NAEP 8th-grade physical science items. Nine of these items were multiple-choice, and four were free-response (one item included both a multiple-choice and a free-response portion). This analysis will only use the multiple-choice items, as not all of the free-response items are available. The multiple-choice items are loaded onto one construct of science knowledge.

Table 4.2 CESE Teacher Observation Teacher Items and Constructs

Construct	Item
PBL Practices	Teacher's use of DQ (Driving Question) Teacher's Support for Figuring Out Teacher provides feedback to encourage students in using the SEPs and CCCs to make sense of phenomena Teacher's use of discourse moves to engage in sensemaking
Social and Emotional Learning	Teacher's support of agency Teacher's support of persistence Teachers support for collaboration and small group work
Fidelity of Implementation	Lesson as Written PBL Practices
Classroom Management	Clear Evidence of Norms and Routines Teacher supports Student work-overall monitoring of student work Instructional sequencing and pacing Behavior management

Table 4.3 CESE Pretest 1PL vs 2 PL vs 3PL

Model	AIC	BIC	Likelihood	LRT	df	p-value
1-PL	64165.67	64232.23	-32072.83			
2-PL	63613.77	63733.60	-31788.89	5673.89	8	<0.001
2-PL	63613.77	63733.60	-31788.89			
3-PL	63526.74	63706.48	-31736.37	105.03	9	<0.001

The psychometrician on the project evaluated whether the unidimensional IRT model of 1 PL, 2 PL, or 3 PL was the best fit for these items. The likelihood ratio test results are reported in Table 4.3.

The 2-PL model was better than the 1-PL model, with an LRT of 567.89, df of 8, and a p-value less than 0.001. The 3-PL model was better than the 2-PL model, with an LRT of 105.03, df of 9, and a p-value less than 0.001.

After determining that the 3-PL model was the best fit, the guessing, difficulty, and discrimination parameters were estimated for the nine items, these are reported in Table 4.4.

The pretest has varying levels of difficulty and discrimination. The guessing parameter falls between approximately zero and 0.34, with some items having a close-to-no guessing probability and some having a higher-than-probable guessing probability.

Table 4.4 CESE Pretest 3-PL model

Item	Guessing	Difficulty	Discrimination
Item 1	0.29	1.149	1.168
Item 2	0.336	-0.576	1.544
Item 3	0.332	1.264	2.076
Item 4	0.295	1.711	1.92
Item 5	0.002	-1.007	1.011
Item 6	0.275	0.959	0.895
Item 8	0.002	-0.64	1.675
Item 9	0.289	0.401	1.384
Item 10	0.004	-1.004	1.753

Table 4.5 unidimensionality of the CESE summative assessment

	Chemistry	Physics
1 factor vs 2 factor	$\chi^2 = 437.92$	$\chi^2 = 142$
	df = 1	df = 1
	p-value < 0.001	p-value < 0.001
1 factor vs bifactor	$\chi^2 = 1807.55$	$\chi^2 = 475$
	df=25	df=12
	p-value<0.001	p-value<0.001
2 factor vs bifactor	$\chi^2 = 1343.69$	$\chi^2 = 273.432$
	df=24	df=11
	p-value<0.001	p-value<0.001

The CESE summative assessment

The CESE summative assessment was split between physics and chemistry students, both developed with items from the 11th grade science assessment of the Michigan Department of Education. The chemistry assessment had 13 questions, with some having sub-items, and the physics assessment had six questions, with some also having sub-items. Because the questions with sub-items did not require the student to get the first part correct to get the second part correct, each is treated as its own item. The items were checked for unidimensionality for the chemistry and physics summative assessments. A one-factor confirmatory factor analysis was compared to a two-factor confirmatory factor analysis, then both the one factor and two factor models were compared to a bifactor model. The model comparison results for the one factor vs. two factors vs. bifactor are reported in Table 4.5.

The bifactor model outperforms the 1-factor and 2-factor models for both chemistry and physics.

Table 4.6 Bifactor model fit of the CESE summative assessment

	Chemistry	Physics
RMSEA	0.036	0.018
	[0.034,0.037]	[0.017,0.018]
CFI	0.949	0.993
TLI	0.938	0.926

Note: 95% confidence intervals for the RMSEA are in brackets.

Table 4.7 CESE Summative Assessment 1PL vs 2 PL vs 3PL

Model	AIC	BIC	Likelihood	Scaling factor	Parameters	Difference scaling correction	LRT	df	p-value
Chemistry									
1-PL	124131.836	124310.418	-62037.918	0.983	28.000				
2-PL	120154.937	120633.282	-60002.469	1.090	75.000	1.155	3525.561	47.000	<0.001
2-PL	120154.937	120633.282	-60002.469	1.090	75.000				
3-PL	119696.482	120334.275	-59748.210	1.054	100.000	0.944	538.456	25.000	<0.001
Physics									
1-PL	26290.527	26373.125	-13130.263	1.009	15.000				
2-PL	25949.374	26147.612	-12938.687	1.024	36.000	1.035	370.129	21.000	<0.001
2-PL	25949.374	26147.612	-12938.687	1.024	36.000				
3-PL	25741.935	26006.251	-12822.967	0.839	48.000	0.284	816.079	12.000	<0.001

The model fit indices for the two bifactor models are reported in Table **??**.

The RMSEA values and confidence intervals for chemistry and physics fall less than 0.05, and their CFI and TLI values are greater than 0.90, indicating a good model fit for the bifactor model. Following ensuring that the bifactor model was a good fit for the summative assessment, the 1-PL, 2-PL, and 3-PL bifactor IRT models were compared for chemistry and physics. This model comparison for the chemistry and physics model fit between 1-PL, 2-PL, and 3-PL is reported in Table 4.7.

There were issues in the estimation for the 3-PL models for both physics and chemistry. This led to additional constraints in the physics 3-PL model and likely incorrect estimates in the chemistry 3-PL model. Because of this, even though the 3-PL models were a better model fit than the 2-PL models, the 2-PL models were chosen as the better model for the physics and chemistry items. Table 4.8 provides the item discriminations and difficulties for the summative assessment for chemistry and physics. Using the IRT bifactor model parameters, the reliabilities were estimated (Raykov et al., 2010). The reliability estimate for the chemistry summative assessment general factor was 0.824, and the reliability for the physics general factor was 0.853.

Table 4.8 CESE Summative Assessment 2PL parameters

	Chemistry				Physics			
Item	a_G	a_1	a_2	Difficulty	a_G	a_1	a_2	Difficulty
Item 1	1.0948	0.3672		0.6375	1.9652	-0.0765		0.680
Item 2	4.2228	7.3542		-8.5085	1.8462	-0.5287		-0.2771
Item 3	8.7057	15.6859		-17.3281	3.1433	-0.0629		2.0009
Item 4	0.816	-0.0221		-0.4267	3.2861	-0.1258		-2.159
Item 5	1.6116	0.2482		1.0336	1.6354	-0.2329		-0.3689
Item 6	2.2712	0.3893		0.6834	1.4093	-0.1292		0.4114
Item 7	0.2448	0.1802		0.7939	1.1305	1.7425		1.5538
Item 8	2.0451	0.4403		-2.1981	2.2729	1.3821		2.006
Item 9	0.5219	0.0221		1.6898	1.190	1.5674		1.7867
Item 10	1.8394	0.1921		-0.6987	1.5623		4.318	0.6018
Item 11	3.1501	2.5942		6.6198	1.0812		2.9665	0.8993
Item 12	2.0638	1.5555		4.8178	2.1539		0.6001	3.4578
Item 13	0.5389		0.1275	1.7918				
Item 14	1.1492		0.3502	-1.4331				
Item 15	1.615		0.0782	-0.4199				
Item 16	0.0493		0.3587	1.0064				
Item 17	1.3702		0.3655	-0.1326				
Item 18	1.5402		0.4471	2.7676				
Item 19	0.5389		0.3995	0.2346				
Item 20	-0.2686		0.2227	0.8228				
Item 21	-0.0323		2.9087	1.4144				
Item 22	-0.3128		2.8866	1.7391				
Item 23	1.5096		0.2142	0.0748				
Item 24	2.3324		0.2533	2.8458				
Item 25	3.451		0.2227	6.0367				

Analysis

Assumed Mediation Model

The assumed mediation model for this study is shown in Figures 4.1-4.3 with the following assumed models to be estimated:

Level 1:

$$chem_{ijkl} = \frac{1}{1 + e^{-(u^{c}_{i0kl} + a^{c}_{i1ks}(\theta^{cw}_{fjkl}) - b^{c}_{i2kl})}}$$

$$phy_{ijkl} = \frac{1}{1 + e^{-(u^{p}_{i0kl} + a^{p}_{i1kl}(\theta^{pw}_{fjkl}) - b^{p}_{i2kl})}}$$

$$pretest_{ijkl} = c_i + (1 - c_i)\frac{1}{1 + e^{-(u^{g}_{i0kl} + a^{g}_{i1kl}(\theta^{gw}_{jkl}) - b^{g}_{i2kl})}}$$

$$\theta^{cw}_{Gjkl} = \beta^{cw}_0 + \beta^{cw}_1 \theta^{gw}_{jkl} + \epsilon^{cw}_{jkl}$$

$$\theta^{pw}_{Gjkl} = \beta^{pw}_0 + \beta^{pw}_1 \theta^{gw}_{jkl} + \epsilon^{pw}_{jkl}$$

Level 2:

$$M_{itkl} = \alpha_{it0l} + \lambda_{it1kl}\xi^{w}_{kl} + \delta_{it2s}\zeta_{tkl} + \epsilon^{M}_{itkl}$$

$$u^{c}_{i0kl} = u^{c}_{i00l} + a^{c}_{i01l}(\theta^{cbk}_{fkl}) - b^{c}_{i02l}$$

$$u^{p}_{i0kl} = u^{p}_{i00l} + a^{p}_{i01l}(\theta^{pbk}_{fkl}) - b^{p}_{i02l}$$

$$u^{g}_{i0kl} = u^{g}_{i00l} + a^{g}_{i01l}(\theta^{gbk}_{kl}) - b^{g}_{i02l}$$

$$a^{c}_{i1kl} = a^{c}_{i10l}$$

$$a^{p}_{i1kl} = a^{p}_{i10l}$$

$$a^{g}_{i1kl} = a^{g}_{i10l}$$

$$b^{c}_{i2kl} = b^{c}_{i20l}$$

$$b^{p}_{i2kl} = b^{p}_{i20l}$$

$$b^{g}_{i2kl} = b^{g}_{i20l}$$

$$\theta^{cbk}_{Gkl} = \beta^{cbk}_0 + \beta^{cbk}_1 \xi^{w}_{kl} + \beta^{cbk}_2 \theta^{gbk}_{kl} + \epsilon^{cbk}_{kl}$$

$$\theta^{pbk}_{Gkl} = \beta^{pbk}_0 + \beta^{pbk}_1 \xi^{w}_{kl} + \beta^{pbk}_2 \theta^{gbk}_{kl} + \epsilon^{pbk}_{kl}$$

Level 3:

$$\alpha_{it0l} = \alpha_{it00} + \lambda_{it01}\xi^{b}_{l} + \upsilon^{M}_{it0l}$$

$$\lambda_{it1l} = \lambda_{it10}$$

$$\delta_{it2l} = \delta_{it20}$$

$$u^{c}_{i00l} = u^{c}_{i000} + a^{c}_{i001}(\theta^{cbl}_{fl} - b^{c}_{il})$$

$$u^{p}_{i00l} = u^{p}_{i000} + a^{p}_{i001}(\theta^{pbl}_{fl} - b^{p}_{il})$$

$$u^{g}_{i00l} = u^{g}_{i000} + a^{g}_{i001}(\theta^{gbl}_{l} - b^{g}_{il})$$

$$a^c_{i01l} = a^c_{i010}$$

$$a^p_{i01l} = a^p_{i010}$$

$$a^g_{i01l} = a^g_{i010}$$

$$b^c_{i02l} = b^c_{i020}$$

$$b^p_{i02l} = b^p_{i020}$$

$$b^g_{i02l} = b^g_{i020}$$

$$a^c_{i10l} = a^c_{i100}$$

$$a^p_{i10l} = a^p_{i100}$$

$$a^g_{i10l} = a^g_{i100}$$

$$b^c_{i20l} = b^c_{i200}$$

$$b^p_{i20l} = b^p_{i200}$$

$$b^g_{i20l} = b^g_{i200}$$

$$\xi^b_l = \gamma_0 + \gamma_1 T_l + \epsilon$$

$$\theta^{cbl}_{Gl} = \beta^c_0 + \beta^c_1 \xi^b_l + \beta^c_2 T_l + \beta^c_3 \theta^{gbl}_l + \upsilon^c$$

$$\theta^{pbl}_{Gl} = \beta^p_0 + \beta^p_1 \xi^b_l + \beta^p_2 T_l + \beta^p_3 \theta^{gbl}_l + \upsilon^p$$

At the student level, $chem_{ijkl}$ and phy_{ijkl} are the chemistry and physics summative assessment item i for student j in teacher k in school l. θ^{cw}_{fjkl} and θ^{pw}_{fjkl} are the vector of latent factors for the chemistry and physics abilities within a student. a^c_{i100} and a^p_{i100} are the student level vector of discrimination parameters for the chemistry and physics summative assessment items, and b^c_{i200} and b^p_{i200} are the student level chemistry and physics item difficulties for item i. $pretest_{ijkl}$ is the general science pretest items; θ^{gw}_{jkl} is the general science ability for student j in teacher k in school l; a^g_{i100}, the student level pretest item discriminations; b^g_{i200}, the student level pretest item difficulties; and c_i, the pretest guessing parameter for item i. Finally, at the student level, β^{cw}_1 is the relationship between the general science ability from the pretest and the general chemistry ability, and β_1^{pw} is the relationship between the general science ability from the pretest and the general physics ability.

At the teacher level, M_{itkl} is the teacher observation PBL item i at timepoint t for teacher k in school l. ξ_{kl}^{w} is the teacher PBL trait for teacher k in school l, and ζ_{tkl} is the teacher PBL state at timepoint t for teacher k in school l. λ_{it10} and δ_{it20} are the teacher-level factor loadings for the PBL trait and states, respectively. θ_{fkl}^{cbk}, θ_{fkl}^{pbk}, and θ_{kl}^{gbk} are the between teacher chemistry, physics, and general science latent abilities; a_{i010}^{c}, a_{i010}^{p}, and a_{i010}^{g}, the between teacher item discrimination parameters; b_{i020}^{c}, b_{i020}^{p}, and b_{i020}^{g}, the between teacher item difficulty parameters. β_1^{cbk} and β_1^{pbk} are the relationships between the teacher's PBL trait and the teacher-level general chemistry and physics ability. β_2^{cbk} and β_2^{pbk} are the relationships between teacher general science ability and teacher general chemistry and physics ability respectively.

At the school level, ξ_{l}^{b} is the between school teacher PBL trait for school l, and λ_{it01} is the between school factor loadings for the teacher PBL trait. θ_{fl}^{cbs}, θ_{fl}^{pbs}, and θ_{l}^{gbs} are the between school chemistry, physics, and general science latent abilities; a_{i001}^{c}, a_{i001}^{p}, and a_{i001}^{g}, the between school item discrimination parameters; b_{i002}^{c}, b_{i002}^{p}, and b_{i002}^{g}, the between school item difficulty parameters. γ_1 is the treatment effect on the between-school teacher PBL trait. β_1^{c} and β_1^{p} are the relationships between the between school teacher PBL trait and the between school general chemistry and physics ability. β_2^{c} and β_2^{p} are the treatment effects on the between school general chemistry and physics ability. β_3^{c} and β_3^{p} are the relationships between the between school general science ability and the between school general chemistry and physics ability.

Finally, the estimated mediation of the teacher PBL trait is $\gamma_1 \times \beta_1^{c}$ and $\gamma_1 \times \beta_1^{p}$.

Because of the complexity of estimating a model that includes two different 2-PL bifactor outcomes where a different set of students have different measures (chemistry vs physics) and a 3-PL pretest measure, this study uses equated outcome scores and pretest factor scores which correspond to the estimation conducted in Schneider et al. (2022). The online appendix also describes the equating process in Schneider et al. (2022). This simplifies the final estimated mediation model to:

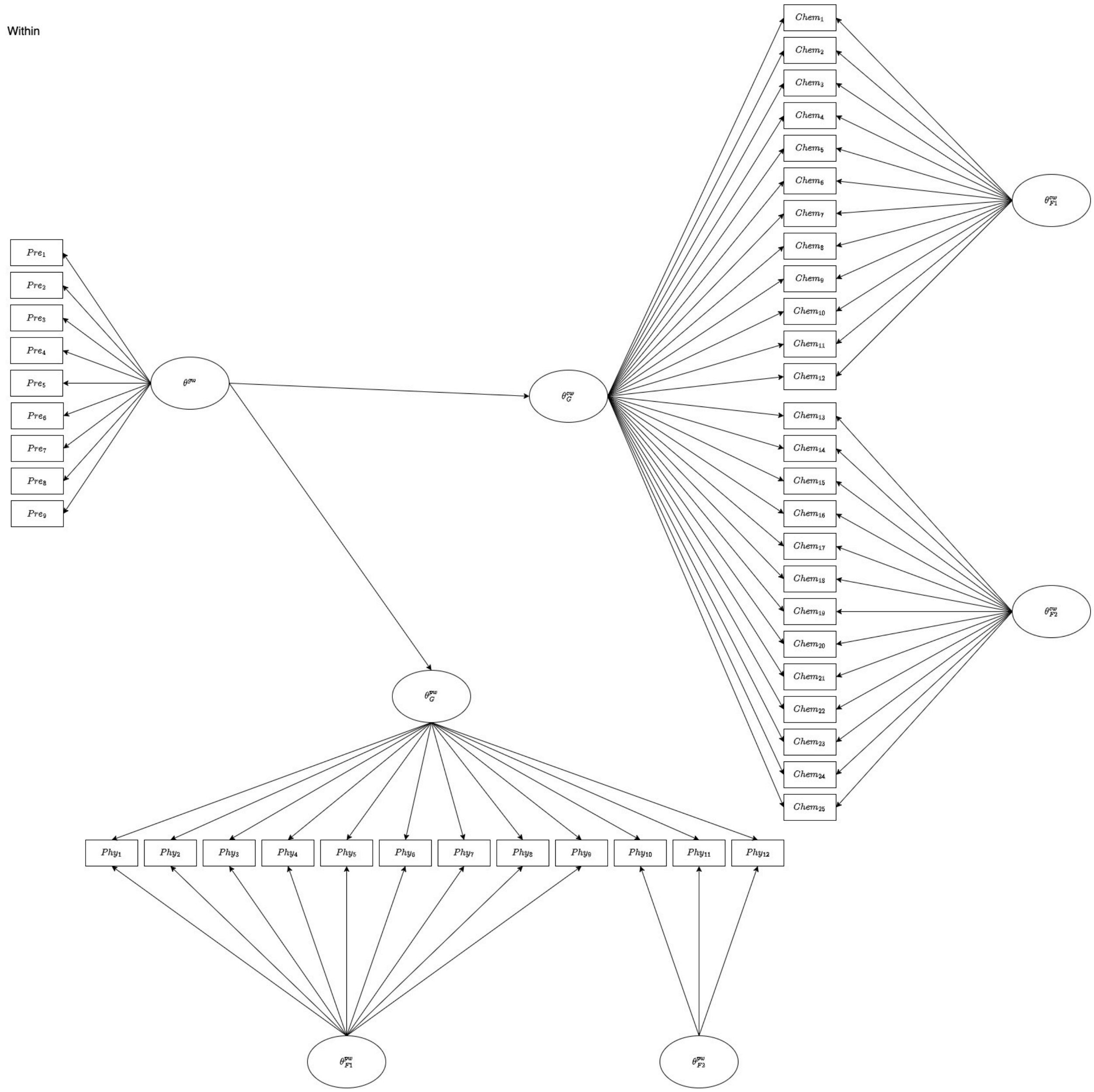

Figure 4.1 CESE multiple timepoint mediation model within

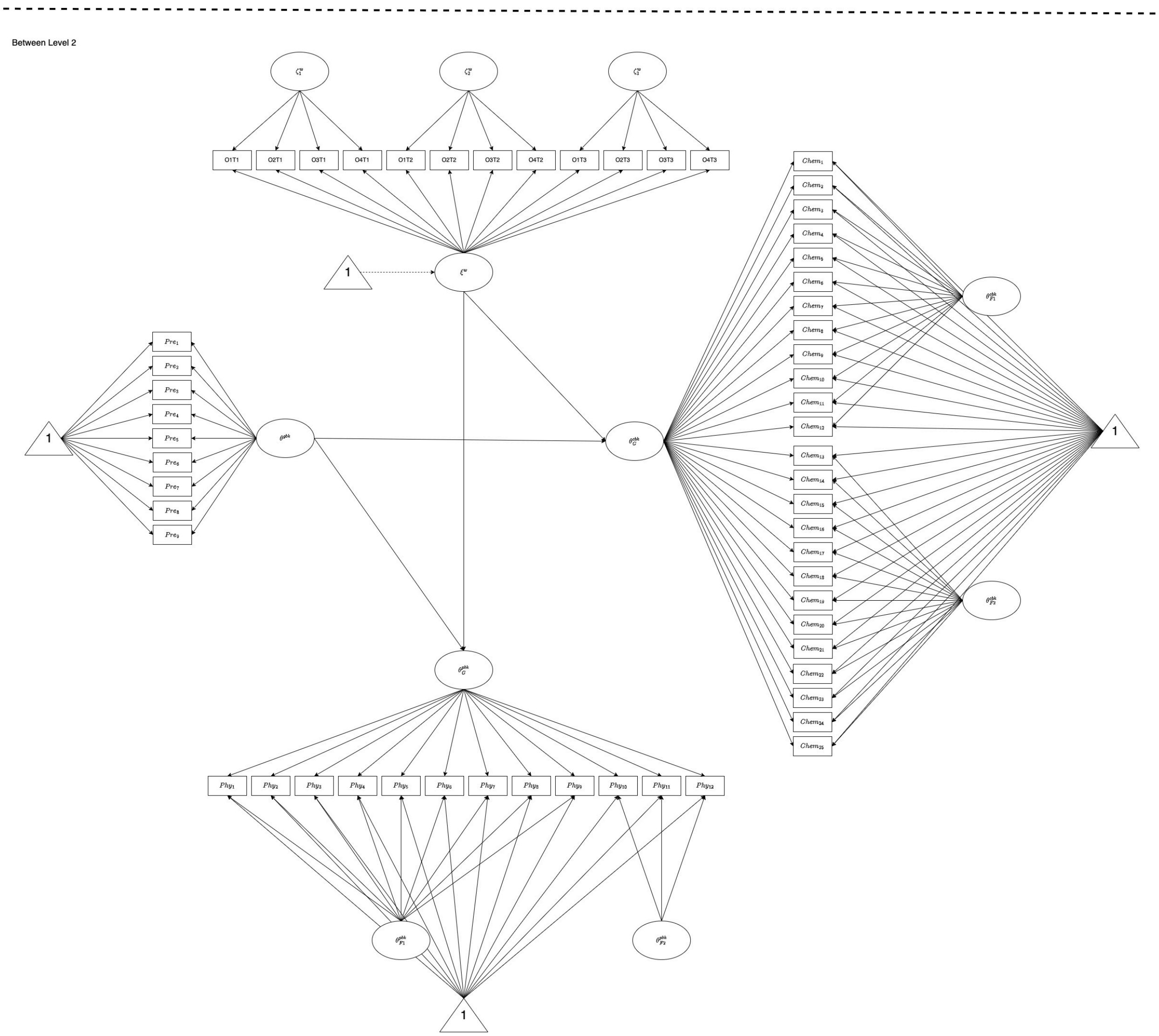

Figure 4.2 CESE multiple timepoint mediation model between teacher level

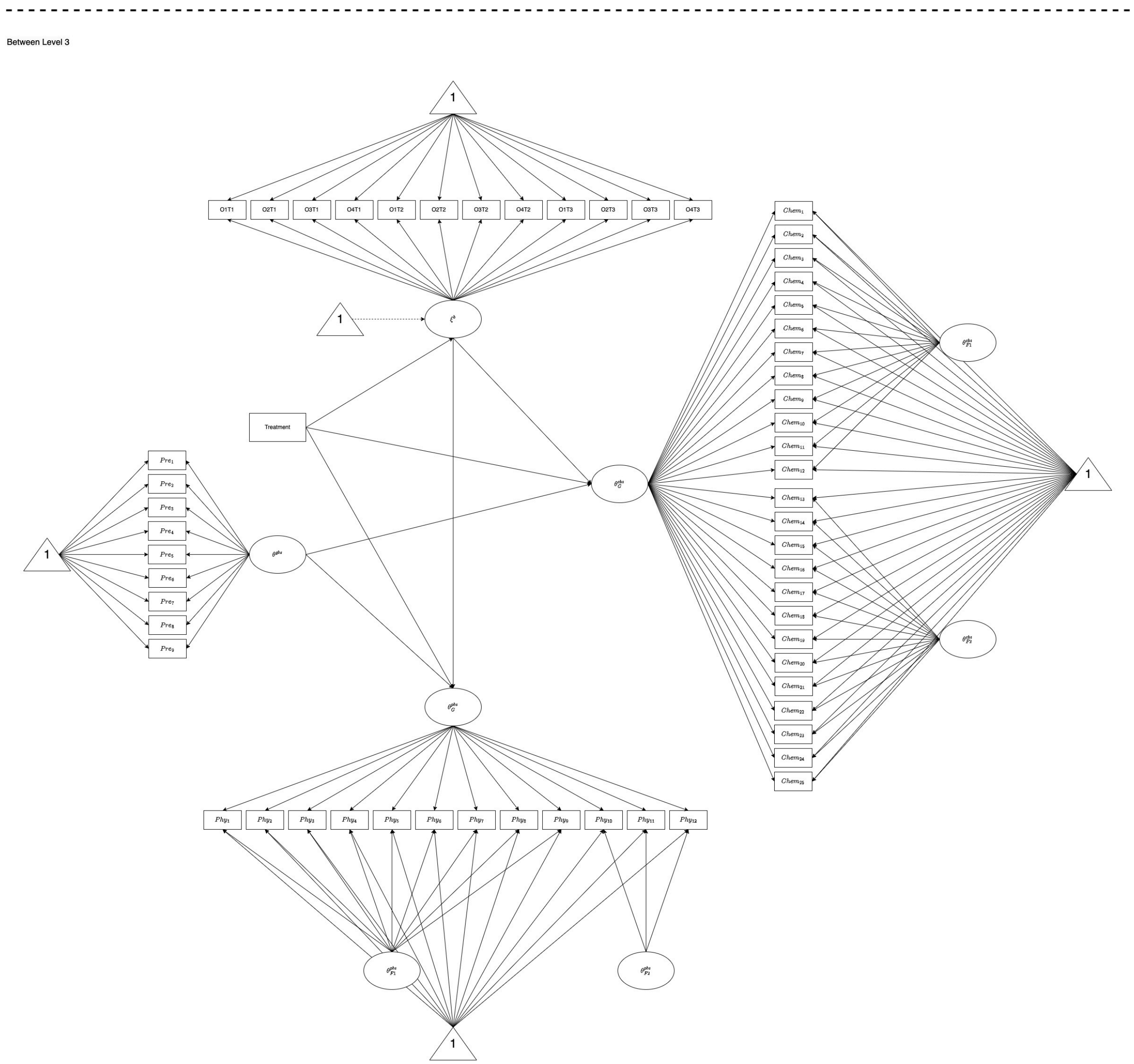

Figure 4.3 CESE multiple timepoint mediation model between school level

Level 1:

$$scienceachievement = \beta_{0kl} + \beta_{1kl}\hat{\theta}^{wg}_{jkl} + \epsilon_{jkl}$$

Level 2:

$$M_{itkl} = \alpha_{it0l} + \lambda_{it1l}\xi^{w}_{kl} + \delta_{it2l}\zeta_{tkl} + \epsilon^{M}_{itkl}$$

$$\beta_{0kl} = \beta_{00l} + \beta_{01l}\xi^{w}_{kl} + \beta_{02l}\hat{\theta}^{bg}_{kl} + r_{0kl}$$

$$\beta_{1kl} = \beta_{10l}$$

Level 3:

$$\alpha_{it0l} = \alpha_{it00} + \lambda_{it01}\xi^{b}_{l} + \upsilon^{M}_{it0l}$$

$$\lambda_{it1l} = \lambda_{it10}$$

$$\delta_{it2l} = \delta_{it20}$$

$$\beta_{00l} = \beta_{000} + \beta_{001}T_{l} + u_{00l}$$

$$\beta_{10l} = \beta_{100}$$

$$\beta_{01l} = \beta_{010}$$

$$\beta_{02l} = \beta_{020}$$

$$\xi^{b}_{l} = \gamma_0 + \gamma_1 T_l + \epsilon$$

Where, at the student level, β_{100} is the relationship between the general science ability factor scores from the pretest ($\hat{\theta}^{wg}_{jkl}$) with the equated physics and chemistry science scores.

At the teacher level, M_{itkl} is the teacher observation PBL item i at timepoint t for teacher k in school l. ξ^{w}_{kl} is the teacher PBL trait for teacher k in school l, and ζ_{tkl} is the teacher PBL state at timepoint t for teacher k in school l. λ_{it10} and δ_{it20} are the factor loadings for the teacher-level PBL trait and states, respectively. β_{010} is the relationship between the teacher PBL trait and the equated physics and chemistry science scores and β_{020} is the relationship between the between teacher general science ability factor scores and the equated physics and chemistry science scores.

At the school level, ξ_l^b is the between school teacher PBL trait for school l, and λ_{it01} is the between school factor loadings for the teacher PBL trait. β_{001} is the treatment effects on the equated physics and chemistry science scores.

Finally, the estimated mediation of the teacher PBL trait is $\gamma_1 \times \beta_{010}$. Figure 4.4 depicts this simplified mediation model.

Model Comparison

The teacher observation measure of teacher PBL practices will be compared across three different models: a unidimensional model, which assumes no state factors; a three-factor model, which assumes no trait factor; and the bifactor model, which allows for both state and trait factors of the teacher mediation. In all of these models, time-invariant factor loadings are not assumed as is often assumed in latent trait state theory. The reason for not assuming time invariant factor loadings is because these measures were collected by observers who may themselves be influenced by different times and thus affect the loadings at various time points. Similarly, the trait factor loadings were not assumed to unity. Because of the small sample size (n = 55 for 36 to 48 free parameters), these measurement models will be estimated using Bayes with non-informative priors. These models were then compared through their deviance information criterion (Spiegelhalter et al., 2002, DIC), bayesian information criterion (Schwarz, 1978, BIC), and posterior predictive credible intervals and p-values (Gelman et al., 1996). The DIC and BIC values are purely for model comparison, with smaller DIC and BIC values indicating a more robust model fit. The posterior predictive credible intervals (CI) and p-values indicate general model fit, with a CI that includes zero and a p-value close to 0.5 indicating strong model fit.

Estimation

This mediation model will be estimated using the four different models examined in the simulation study: standardized averages of the mediator across time points (3-2-1 mediation); factor scores for each time point averaged across time points (ignoring the trait factor; 3-2-1 mediation; factor scores from the LST as the mediator (3-2-1 mediation); fully specified model (Multilevel SEM).

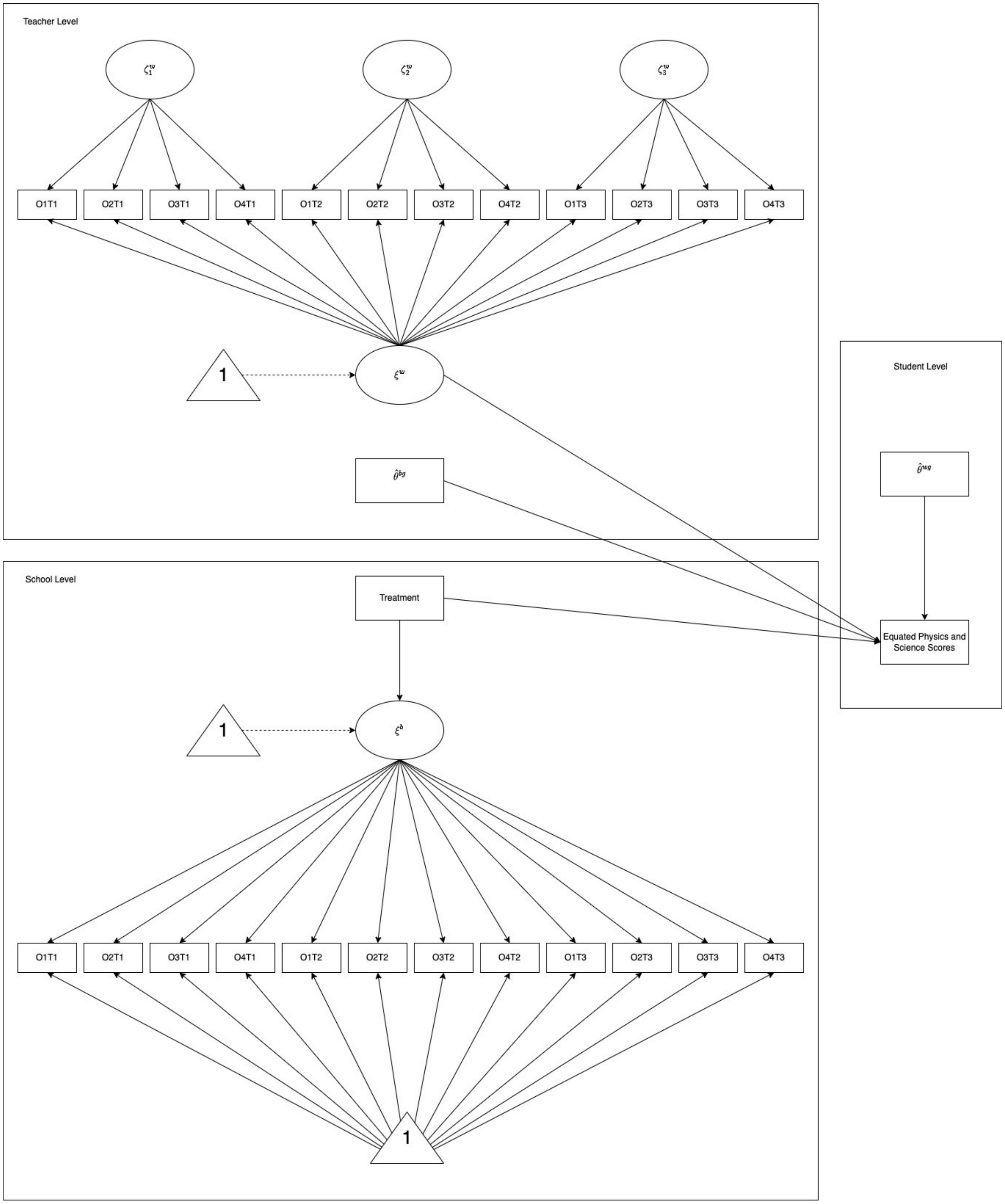

Figure 4.4 CESE Simplified Multilevel Mediation Model

Before delving into the mediation model, the treatment effect estimated in Schneider et al. (2022) will be reestimated using Bayesian multilevel modeling to replicate the findings. This will be done with the full sample that includes all teachers, even those without the observations, and then with the limited sample, which includes only the 55 teachers who had at least one observation.

After estimating the treatment effect, the following four mediation models will be estimated.

Standardized average of the mediator across time points: As was defined in Chapter 2, the averages of the mediator across the time points are defined as:

$$\bar{M}_{kl} = \frac{1}{12}\Sigma_{t=1}^{t=3}\Sigma_{i=1}^{i=4}M_{itkl}$$

Then, the following is estimated:

Level 1:

$$scienceachievement = \beta_{0kl} + \beta_{1kl}\hat{\theta}_{jkl}^{wg} + \epsilon_{jkl}$$

Level 2:

$$\beta_{0kl} = \beta_{00l} + \beta_{01l}\bar{M} + \beta_{02l}\hat{\theta}_{kl}^{bg} + \hat{r}_{0kl}$$

$$\beta_{1kl} = \beta_{10l}$$

$$\bar{M}_{kl} = \gamma_{0s} + \epsilon_{kl}$$

Level 3:

$$\beta_{00l} = \hat{\beta}_{000} + \hat{\beta}_{001}T_l + \hat{u}_{00l}$$

$$\beta_{10l} = \hat{\beta}_{100}$$

$$\beta_{01l} = \hat{\beta}_{010}$$

$$\beta_{02l} = \hat{\beta}_{020}$$

$$\gamma_{0l} = \hat{\gamma}_{00} + \hat{\gamma}_{01}T_l + \hat{r}_{0l}$$

Then $\hat{\gamma}_{01}$ is the estimated a pathway, $\hat{\beta}_{010}$ is the estimated b pathway, and $\hat{\gamma}_{01} \times \hat{\beta}_{010}$ is the indirect effect of the PBL practices on the treatment effect.

Factor scores for each time point averaged across time points: Also defined in Chapter 2, the factor for each time point will be defined as:

$$M_{it} = \alpha_{it} + \hat{\delta}_{it1}\hat{\zeta}_{tkl} + \epsilon_{it}^{M}$$

$$\bar{\zeta}_{kl} = \frac{1}{3}\Sigma_{t=1}^{t=3}\hat{\zeta}_{tkl}$$

Where $\hat{\zeta}_{tkl}$ is the estimated factor score at time point t and $\bar{\zeta}_{kl}$ is the average of those factor scores across the three time points. Then, the following mediation model is estimated:

Level 1:

$$science achievement = \beta_{0kl} + \beta_{1kl}\hat{\theta}_{jkl}^{wg} + \epsilon_{jkl}$$

Level 2:

$$\beta_{0kl} = \beta_{00l} + \beta_{01l}\bar{\zeta}_{kl} + \beta_{02l}\hat{\theta}_{kl}^{bg} + \hat{r}_{0kl}$$

$$\beta_{1kl} = \beta_{10l}$$

$$\bar{\zeta}_{kl} = \gamma_{0l} + \epsilon_{kl}$$

Level 3:

$$\beta_{00l} = \hat{\beta}_{000} + \hat{\beta}_{001}T_l + \hat{u}_{00l}$$

$$\beta_{10l} = \hat{\beta}_{100}$$

$$\beta_{01l} = \hat{\beta}_{010}$$

$$\beta_{02l} = \hat{\beta}_{020}$$

$$\gamma_{0l} = \hat{\gamma}_{00} + \hat{\gamma}_{01}T_l + \hat{r}_{0l}$$

Where once again, $\hat{\gamma}_{01}$ is the estimated a pathway, $\hat{\beta}_{010}$ is the estimated b pathway, and $\hat{\gamma}_{01} \times \hat{\beta}_{010}$ is the indirect effect of the PBL practices on the treatment effect.

Factor scores from the LST as the mediator: Again, as was defined in Chapter 2, the factor scores from the LST are estimated by:

$$M_{itkl} = \hat{\alpha}_{it00} + \hat{\lambda}_{it10}\hat{\xi}_{kl} + \hat{\delta}_{it20}\zeta_{tkl} + \epsilon_{itkl}$$

$\hat{\xi}_{kl}$ is the estimated factor score for the teacher PBL practices trait. Then, the mediation model is estimated:

Level 1:

$$scienceachievement = \beta_{0kl} + \beta_{1kl}\hat{\theta}^{wg}_{jkl} + \epsilon_{jkl}$$

Level 2:

$$\beta_{0kl} = \beta_{00l} + \beta_{01l}\hat{\xi_{kl}} + \beta_{02l}\hat{\theta}^{bg}_{kl} + \hat{r}_{0kl}$$

$$\beta_{1kl} = \beta_{10l}$$

$$\hat{\xi}_{kl} = \gamma_{0l} + \epsilon_{kl}$$

Level 3:

$$\beta_{00l} = \hat{\beta}_{000} + \hat{\beta}_{001}T_l + \hat{u}_{00l}$$

$$\beta_{10l} = \hat{\beta}_{100}$$

$$\beta_{01l} = \hat{\beta}_{010}$$

$$\beta_{02l} = \hat{\beta}_{020}$$

$$\gamma_{0l} = \hat{\gamma}_{00} + \hat{\gamma}_{01}T_l + \hat{r}_{0l}$$

Where once again, $\hat{\gamma}_{01}$ is the estimated a pathway, $\hat{\beta}_{010}$ is the estimated b pathway, and $\hat{\gamma}_{01} \times \hat{\beta}_{010}$ is the indirect effect of the PBL practices on the treatment effect.

Fully specified model: The simplified model on page 42 is estimated as follows:

Level 1:

$$scienceachievement = \beta_{0kl} + \beta_{1kl}\hat{\theta}^{wg}_{jkl} + \epsilon_{jkl}$$

Level 2:

$$M_{itkl} = \alpha_{it0l} + \lambda_{it1l}\xi^{w}_{kl} + \delta_{it2l}\zeta_{tkl} + \epsilon^{M}_{itkl}$$

$$\beta_{0kl} = \beta_{00l} + \beta_{01l}\xi^{w}_{kl} + \beta_{02l}\theta^{\hat{bg}}_{kl} + r\hat{_{0kl}}$$

$$\beta_{1kl} = \beta_{10l}$$

Level 3:

$$\alpha_{it0l} = \hat{\alpha}_{it00} + \hat{\lambda}_{it01}\xi_l^b + \upsilon_{it0l}^M$$

$$\lambda_{it1l} = \hat{\lambda}_{it10}$$

$$\delta_{it2l} = \hat{\delta}_{it20}$$

$$\beta_{00l} = \hat{\beta}_{000} + \hat{\beta}_{001}T_l + \hat{u}_{00l}$$

$$\beta_{10l} = \hat{\beta}_{100}$$

$$\beta_{01l} = \hat{\beta}_{010}$$

$$\beta_{02l} = \hat{\beta}_{020}$$

$$\xi_l^b = \hat{\gamma}_0 + \hat{\gamma}_1 T_l + \epsilon$$

These models were estimated using Bayesian structural equation modeling with non-informative priors on the coefficients and informative prior on the school level and teacher level variances on the equated summative assessment and school level variance on the teacher PBL practices to increase power and convergence in this smaller sample size. For the school level variance for the equated summative assessment, the prior was set as school level variance follows an inverse gamma distribution (Anderson, 2007) with $\alpha = 2.25$ and $\beta = 0.19$ which yields a mean of the distribution at 0.154 (corresponding to an ICC of 0.154 since the outcome variable is standardized with a variance of 1 Spybrook et al., 2022) and the teacher level variance follows an inverse gamma distribution with $\alpha = 2.25$ and $\beta = 0.155$ which yields a mean of the distribution at 0.124 (corresponding to an ICC of 0.124 Spybrook et al., 2022). For the school level variance on teacher PBL practices, the prior follows an inverse gamma distribution with $\alpha = 2.12$ and $\beta = 0.12$, which yields a mean of the distribution at 0.11 (corresponding to an ICC of 0.11 since the teacher PBL practices are either standardized with variance of one or the latent variable is constrained to have a variance of 1Westine et al., 2020). Similar to the simulation study, the Gibbs sampler and the default Mplus settings were used to determine convergence (see pages 38-39). Again, similarly, in estimating the factor scores from the LST model and the fully specified model, the teacher-level PBL practices

Table 4.9 Model Comparison for the Teacher Observations

			Posterior Predictive Checking	
Model	DIC	BIC	95% CI	p-value
1-factor	1003.311	1080.395	[-29.262,58.977]	0.312
3-factor	974.319	1058.475	[-36.542,43.055]	0.444
Bifactor	991.967	1089.647	[-34.509,44.831]	0.452

trait and time-specific factors variance were fixed to 1, and the correlation between them was fixed to 0 for identification of the model. The Mplus code for the four different estimation methods can be found in Appendix B.

4.2 Results

Model Comparison

Across the three different models for the teacher observations, all three models had acceptable posterior predictive credible intervals and p-values (0 inclusive and not too far from 0.5 respectively), although the 3-factor solutions had a p-value closer to 0.5 than the 1 factor, indicating stronger fit and the bifactor had a p-value closer to 0.5 than both the1-factor and 3-factor models, although not substantially closer compared to the 3-factor model. The 3-factor model had the lowest DIC and BIC values compared to the 1-factor and bifactor models; however, the DIC and BIC values of the 3-factor were not substantially lower than the bifactor model (around a 20-30 difference). These results are displayed in Table 4.9.

Because the theoretical framework for the bifactor model (allowing for a trait factor of PBL practices as the mediator) fits better to the theoretical mediation model assumed compared to the 3-factor model (assuming no trait factor of PBL practices and only state factors). Since the models performed comparably, the bifactor model is deemed to be an adequate model of the teacher PBL practices. Both the 3-factor and bifactor factor loadings are reported in Table 4.10.

Here, the 3 factor loadings are similar to the state-specific factor loadings of the bifactor model (which are slightly lower in some cases). Most of the time specific factor loadings are medium sized factor loadings with some being low and some high. Most of the factor loadings for the trait factor are small (between 0.1 and 0.4), with some being medium. This aligns with the model comparisons

Table 4.10 Factor loadings for 3 factor and bifactor models

Item	3 factor model			Bifactor model			
	Time point 1 factor	Time point 2 factor	Time point 3 factor	Time point 1 factor	Time point 2 factor	Time point 3 factor	Trait factor
Timepoint 1 Item 1	0.590			0.582			0.133
Timepoint 1 Item 2	0.749			0.714			0.216
Timepoint 1 Item 3	0.647			0.527			0.322
Timepoint 1 Item 4	0.738			0.589			0.569
Timepoint 2 Item 1		0.499			0.512		0.256
Timepoint 2 Item 2		0.737			0.558		0.545
Timepoint 2 Item 3		0.434			0.346		0.286
Timepoint 2 Item 4		0.990			0.707		0.621
Timepoint 3 Item 1			0.816			0.636	0.285
Timepoint 3 Item 2			0.681			0.638	0.408
Timepoint 3 Item 3			0.216			0.183	0.256
Timepoint 3 Item 4			0.731			0.708	0.056

where the bifactor does not significantly increase model fit compared to the three-factor model. Here, in this latent state-trait bifactor model, more variance is explained at the state level than at the trait factor.

Estimation of Mediation Effects

The estimate of the treatment effects for the full sample and the limited sample are reported in Table 4.11. These effects are comparable to each other, with an estimated impact of 0.198 in the full sample and 0.186 in the limited sample. However, the limited sample is underpowered and unable to detect a treatment effect at the 5% level. These results are comparable to the estimated treatment effect found in Schneider et al. (2022) although slightly smaller in magnitude.

All four estimation models converged in estimating the four mediation effect estimation methods. Mostly, the estimates for the a, b, and product of a and b are consistent across the full and limited samples (outside of the average sum scores for pathways a and b). Almost none of the pathways nor the product of a and b were significant at the 5% level, which is unsurprising for a sample size of 61 schools. Only pathways a for the full sample in the average sum scores were significant, possibly due to chance (or bias). Outside of the average sum scores, pathway b and the product of a and b are both practically zero (although some of the credible intervals for pathway b are quite wide). These results are reported in Table 4.12.

The null mediation results may be a result of several factors. The first is that this study is not powered to detect these a and b pathways, as was found in Chapter 3. The second regards the teacher observation measure itself. With small trait loadings and medium state loadings with only

Table 4.11 Treatment effects of the full and limited sample

Parameter	Full Sample Estimate	Limited Sample Estimate
Treatment	0.198	0.186
	(0.097)	(0.159)
	[0.028,0.384]	[-0.064,0.549]
Region Fixed Effects	0.002	0.004
	(0.003)	(0.006)
	[-0.005,0.008]	[-0.009,0.014]
Teacher level pretest factor	0.414	0.716
	(0.124)	(0.188)
	[0.163,0.683]	[0.409,1.191]
Chemistry	-0.504	-0.394
	(0.071)	(0.121)
	[-0.24,-0.323]	[-0.618,-0.143]
student level pretest factor	0.279	0.299
	(0.016)	(0.024)
	[0.253,0.313]	[0.255,0.348]
student level variance	0.722	0.792
	(0.016)	(0.024)
	[0.693,0.754]	[0.751,0.847]
teacher level variance	0.092	0.094
	(0.021)	(0.030)
	[0.062,0.152]	[0.053,0.170]
school level variance	0.032	0.042
	(0.012)	(0.021)
	[0.021,0.078]	[0.019,0.101]
N		
School	61	36
Teacher	102	49
Student	4238	2113

Note: posterior standard deviations are in parentheses and 95% credible intervals are in brackets.

Table 4.12 Mediation effects of the four estimation Methods

Estimation	a	b	Indirect effect (a*b)
Full Sample			
Average sum scores	-3.195	0.262	-0.693
	(1.396)	(0.351)	(1.386)
	[-6.003,-0.545]	[-0.374,1.060]	[-4.630,1.202]
Average factor scores	0.232	-0.003	0
	(0.204)	(0.039)	(0.014)
	[-0.127,0.656]	[-0.081,0.071]	[-0.032,0.028]
Bifactor factor scores	0.055	-0.071	-0.002
	(0.095)	(0.096)	(0.013)
	[-0.139,0.241]	[-0.264,0.116]	[-0.034,0.021]
Fully specified model	0.074	-0.025	0
	(0.214)	(0.099)	(0.021)
	[-0.370,0.416]	[-0.218,0.173]	[-0.042,0.049]
Limited Sample			
Average sum scores	17.281	0.303	3.952
	(8.369)	(0.415)	(8.713)
	[-0.106,35.486]	[-0.462,1.176]	[-8.146, 26.531]
Average factor scores	0.512	-0.003	-0.001
	(0.383)	(0.065)	(0.042)
	[-0.186,1.249]	[-0.128,0.125]	[-0.087,0.090]
Bifactor factor scores	0.234	-0.060	-0.007
	(0.186)	(0.147)	(0.046)
	[-0.137,0.616]	[-0.356,0.227]	[-0.120,0.068]
Fully specified model	0.091	-0.011	0
	(0.243)	(0.146)	(0.036)
	[-0.422,0.505]	[-0.294,0.279]	[-0.084,0.075]

Note: posterior standard deviations are in parentheses and 95% credible intervals are in brackets.

60 schools, the simulations from Chapter 3 indicate that there was a high likelihood of negative bias, which may be occurring here. The second issue is that this model assumes unconfoundedness after accounting for the covariates (here, student and teacher-level pretest factor scores). There may be additional variables that confound the relationship between the teacher's PBL practice trait and the student science test score that have yet to be accounted for. It may also be the case that the PBL practices are not mediating the treatment effects, which would replicate the findings from Schneider et al. (2022), which found null effects for teacher-reported incorporation of PBL as a mediator.

CHAPTER 5

DISCUSSION

5.1 Contributions

Expanding the 3-2-1 mediation model to incorporate a latent state-trait model gives a new model for estimating mediators that are assumed to be latent variables and estimated at multiple time points. This is particularly pertinent to education studies that use teacher practices observed at multiple time points or teacher ESM data as mediators for the treatment.

The simulation study shows how probable each estimation method may be and the bias and power tradeoff when choosing the method in these 3-2-1 latent trait state mediation models. In general, none of the methods were powered for any sample sizes investigated in this simulation study. This suggests that when using longitudinal mediators in a 3-2-1 model such as this one, a researcher will need a much larger sample size (more than 200) or other ways to increase power, such as using informative priors with Bayesian methods (such as was done in chapter 4) along with sum scores in reliable unidimensional outcome measures (Widaman and Revelle, 2023, which has shown to be unbiased in many situations) to decrease model complexity.

An additional essential consideration that arises from the simulation study is the importance of having a reliable measure with both strong construct and content validity. Essentially, it is a measure that is more consistent across time and less time-specific. When researchers plan an evaluation where they want to evaluate longitudinal mediators, they may want to emphasize the selection of these particular measures to ensure that their measures meet this criterion. Additionally, it is recommended that the developer of these measures report the psychometric properties so that researchers can choose their measures appropriately. If no measure is available, researchers can follow standard protocols of field testing the measure before using them to ensure that they will be adequate to answer the research question of interest.

When using a small sample of schools (30 or 60) with a measure that has medium or high general loadings, using factor scores of the latent state-trait model is the best option. However, the b path may still be biased, especially with small b pathway effects. Researchers may want to consider using bias correction techniques such as Croon's (Kelcey et al., 2021), which were not part of the investigation in this study.

The empirical study contributes to understanding how realistic these model estimation methods are. Because of the complexity of the latent model of the outcome and pretest measures, this empirical study used a standardized equated outcome and factor scores for the pretest, which may have added bias to the four estimation models. However, all four models converged, indicating that these methods can be used in an empirical setting. However, this does come with some caveats. First, with 60 schools, as was seen in the simulation study, this empirical study was not powered to detect reasonable pathway a or b effects. This empirical study highlighted additional considerations when researchers plan on using longitudinal teacher observations as the mediator, such as rater effects and adequate variance in the mediator, and how these considerations may affect the estimation of the mediation.

Implications for Education Research

As education research, along with other social study fields, face issues of replication (Wiliam, 2022), researchers are faced with the question of how the treatment effect works, often leading to an investigation into the mediation effects of the intervention. Even in reasonably non-complex multilevel models, power for these estimation methods requires either many schools ($\geq$ 200) or very high a and b pathways ($\geq$ 0.5; Kelcey et al., 2020. However, effect sizes in education research tend to be small and thus even more so underpowered (Wiliam, 2022). As researchers move from testing the efficacy of their study into scaling their study, they, as well as funding agencies, should expect a dramatic increase in the number of schools required to investigate how the treatment works through mediation. Additionally, as using informative Bayesian priors in the estimation of these mediation effects may increase the power, researchers should aim to collect preliminary data during the development and efficacy stages of their study to estimate informative priors for their scale-up investigation of the mediators. In addition, the education research community should come together to provide appropriate priors for various parameters in differing education research studies, including but not limited to the mediator a and b pathways when available. Finally, many education intervention studies occur over a specified period (such as a semester, an academic year, or multiple academic years), lending to the need for longitudinal mediator measures. In these circumstances, researchers need to invest time and resources into considering the appropriate longitudinal model for these mediation methods and choosing and/or designing the appropriate longitudinal measure. Does a LST model theoretically fit the mediation measures the best, or does a latent change model make more sense? Are the measures loading strongly enough onto the latent factors to minimize bias in the mediation estimation? These are questions researchers should consider as they design their study with longitudinal mediation. Furthermore, resources should be invested in developing and validating strong longitudinal measures for common mediators that may be investigated across numerous education intervention studies.

5.2 Limitations

Several limitations exist to the proposed model, simulation study, and empirical study. Beginning with the proposed bifactor model for estimating the latent state-trait mediator, this model was chosen because it does not require time invariance nor does it necessarily require equal distance between the time points, which other longitudinal models assume; however, this model breaks down quickly with intensive longitudinal data. As time point $t \to \infty$, the bifactor model will have a number of factors, n, that approaches $\infty + 1$. Depending on the sample size of individuals, this may cause identification issues rather quickly. The current model, which incorporates the bifactor latent state-trait model into multilevel structural equation modeling for mediation, may be able to be expanded to latent state-trait models that are more appropriate for intensive longitudinal data (Geiser, 2020).

There are numerous limitations to the simulation study that were beyond the scope of it. This simulation considers ideal conditions in a multilevel structural equation mediation analysis. It first considers balanced designs with equal number of treatment and control clusters and equal numbers of level 2 and level 1 observations within each cluster. Additionally, this study does not vary the number of level 2 and level 1 observations to see how the different estimation methods' bias, power, or convergence may differ by different numbers of observations at various levels (the current study only varies the level 3 sample sizes). One goal of this simulation study was to understand how bias and power are affected by different levels of the general and specific loadings in the bifactor latent state-trait model. However, this led to consistent sizes of loadings across the general and specific loadings, respectively, which is only partially realistic to what may be seen in the actual studies. Studies may have measures where certain items are stronger than others, as was seen in the empirical study in Chapter 4. It may also be interesting to understand how the proportion of items as strong vs. medium vs. weak affects the bias and power of the estimation methods.

Additionally, this simulation study had issues with any methods at any sample size being powered enough to detect the true effects. With level 3 cluster sample sizes of 30, 60, and 200, these are sizes that might be expected in large cluster randomized control trials; however, given that no method was powered with these sample sizes, additional sample sizes or larger effect sizes may need to be investigated to understand at what threshold a study may be powered to detect these longitudinal mediation effects. Additionally, these simulation studies investigated these estimation methods using non-informative Bayesian multilevel structural equation methods; power may be increased by incorporating slightly informative priors. The simulations also did not include covariates in the model, which may increase the power for the a and c' pathways; however, these covariates could decrease the efficiency of the b pathway (Shen et al., 2024). This leads to another significant limitation of this simulation study: the exclusion of covariates and/or confounder variables. This mediation model assumes unconfoundedness between all the variables (conditional on any covariates). The treatment to the mediator, pathway a, is assumed to be unconfounded through random treatment assignment. However, the current simulation also assumes that the mediator-to-outcome relationship is unconfounded, which may not be reasonable in most studies. Additional research on the magnitude of added bias when including no confounders but covariates, ignoring varying levels of essential confounders, and including confounders in the model may give additional insights into bias and power for the different estimation methods.

Finally, this simulation does not consider any bias-corrected estimation methods when using the factor scores of the latent state-trait model. These models have shown to be unbiased and have fewer issues with converging (Kelcey et al., 2021). This addition may provide a more plausible method of investigating these longitudinal mediation effects for researchers.

The empirical study has several limitations. The first is the limited sample size and the mediator teacher measure of PBL practices. With a cluster sample size of 61 and a measure with small trait loadings, the estimated a and b pathways were expected to be biased based on the simulation study from Chapter 3. In addition to the bias from the small sample size and small factor loadings, other areas may have added bias to the estimation for the empirical study. There may have been missing confounders that were not considered in the model that included only the student- and teacher-level pretest. Also, using equated standardized scores for the outcome measure may have introduced bias into the b and c' pathways. Finally, ignoring the effects of the raters may have added bias to estimating all pathways.

5.3 Future Directions

As noted in the limitations section, this work has several avenues for future research. The first would be more fine-tuned simulations to focus on power and what might be required for researchers to have enough power to investigate longitudinal mediators in a 3-2-1 design similar to this one. The addition of covariates, changes in sample size makeup at different levels, different effect sizes, and the addition of using bias-corrected factor scores (Kelcey et al., 2021) or using plausible values from Bayesian factor analysis (Beauducel and Hilger, 2022) to simplify the model should be investigated to understand ways that researchers may be able to increase power reasonably. These questions of power and estimation also lead to examining the effects of confounders on the b and c' pathways on bias and causal implications.

Future research may also include exploring situations that are not ideal; for example, when the measurement has a mixture of small, medium, and large loadings and different proportions of the items, also with sample sizes that are not equal, including varying first and second level sample sizes, but also non-equal sample sizes based upon treatment conditions as well as differing number of measurement times based upon treatment conditions (such as treatment 2nd level observations having more mediator time points than the control group). A further situation to be examined would be the presence of missing data and how missing data may impact the estimation of this model. Another problem to consider would be how this model and estimation performs when the model is misspecified. For instance, if the true model is a cross-lagged mediation analysis, how much bias (if any) is introduced when using the latent state-trait model in the 3-2-1 design instead of a cross-lagged model?

Finally, in the future, the current model proposed in this dissertation can be expanded to other multilevel models, such as the 2-1-1 model or a cross-classified model. It may also be broadened to include rater effects for each item at each time point. Finally, this model should be developed to incorporate intensive longitudinal data that might be expected when using emotionality as a mediator with data collection, such as the experience sampling method (Vongkulluksn and Xie, 2022, ESM).

BIBLIOGRAPHY

Anderson, J. E. (2007). Random effect modelling using bayesian methods. *International Journal of Services Technology and Management*, 8(4-5):316–328.

Asparouhouv and Muthen (2010). *Bayesian Analysis Using Mplus: Technical Implementation*. Muthen & Muthen.

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173.

Bauer, D. J., Preacher, K. J., and Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychological methods*, 11(2):142.

Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., and Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC psychology*, 3:1–12.

Beauducel, A. and Hilger, N. (2022). Coefficients of factor score determinacy for mean plausible values of bayesian factor analysis. *Educational and Psychological Measurement*, 82(6):1069–1086.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological methodology*, pages 37–69.

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., and Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of personality*, 80(1):219–251.

DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of educational measurement*, 43(2):145–168.

Depaoli, S. and Clifton, J. P. (2015). A bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(3):327–351.

Geiser, C. (2020). *Longitudinal structural equation modeling with Mplus: A latent state-trait perspective*. Guilford publications.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760.

Goldsmith, K. A., MacKinnon, D. P., Chalder, T., White, P. D., Sharpe, M., and Pickles, A. (2018). Tutorial: The practical application of longitudinal structural equation mediation models in clinical trials. *Psychological methods*, 23(2):191.

Goldstein, E., Topitzes, J., Brown, R. L., and Barrett, B. (2020). Mediational pathways of meditation and exercise on mental health and perceived stress: A randomized controlled trial. *Journal of health psychology*, 25(12):1816–1830.

Gonzalez, O. and MacKinnon, D. P. (2018). A bifactor approach to model multifaceted constructs in statistical mediation analysis. *Educational and Psychological Measurement*, 78(1):5–31.

Gonzalez, O. and MacKinnon, D. P. (2021). The measurement of the mediator and its influence on statistical mediation conclusions. *Psychological Methods*, 26(1):1.

Hallquist, M. N. and Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, pages 621–638.

Hayes, A. F. and Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: does method really matter? *Psychological science*, 24(10):1918–1927.

Herman, K. C., Reinke, W. M., Dong, N., and Bradshaw, C. P. (2022). Can effective classroom behavior management increase student achievement in middle school? findings from a group randomized trial. *Journal of Educational Psychology*, 114(1):144.

Jose, P. E. (2016). The merits of using longitudinal mediation. *Educational Psychologist*, 51(3-4):331–341.

Kelcey, B., Cox, K., and Dong, N. (2021). Croon's bias-corrected factor score path analysis for small-to moderate-sample multilevel structural equation models. *Organizational research methods*, 24(1):55–77.

Kelcey, B., Spybrook, J., Dong, N., and Bai, F. (2020). Experimental power for cross-level mediation in school-randomized studies of teacher development. *Journal of research on educational effectiveness*.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational researcher*, 49(4):241–253.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., and Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods*, 7(1):83.

MacKinnon, D. P., Lockwood, C. M., and Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate behavioral research*, 39(1):99–128.

Matsumura, L. C., Garnier, H. E., and Spybrook, J. (2013). Literacy coaching to improve student reading achievement: A multi-level mediation model. *Learning and Instruction*, 25:35–48.

McNeish, D. and MacKinnon, D. P. (2022). Intensive longitudinal mediation in mplus. *Psychological methods*.

Muthen and Muthen (1998-2017). *Mplus User's Guide. Eight Edition*. Muthen & Muthen.

Olivera-Aguilar, M., Rikoon, S. H., Gonzalez, O., Kisbu-Sakarya, Y., and MacKinnon, D. P. (2018). Bias, type i error rates, and statistical power of a latent mediation model in the presence of violations of invariance. *Educational and Psychological Measurement*, 78(3):460–481.

O'Neill, A., O'Sullivan, K., O'Sullivan, P., Purtill, H., and O'Keeffe, M. (2020). Examining what factors mediate treatment effect in chronic low back pain: A mediation analysis of a cognitive functional therapy clinical trial. *European Journal of Pain*, 24(9):1765–1774.

Palardy, G. J. (2015). High school socioeconomic composition and college choice: Multilevel mediation via organizational habitus, school practices, peer and staff attitudes. *School Effectiveness and School Improvement*, 26(3):329–353.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., and Chen, F. (2001). Monte carlo experiments: Design and implementation. *Structural Equation Modeling*, 8(2):287–312.

Pituch, K. A., Murphy, D. L., and Tate, R. L. (2009). Three-level models for indirect effects in school-and class-randomized experiments in education. *The Journal of Experimental Education*, 78(1):60–95.

Preacher, K. J. and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, 40(3):879–891.

Preacher, K. J. and Selig, J. P. (2012). Advantages of monte carlo confidence intervals for indirect effects. *Communication Methods and Measures*, 6(2):77–98.

Preacher, K. J., Zyphur, M. J., and Zhang, Z. (2010). A general multilevel sem framework for assessing multilevel mediation. *Psychological methods*, 15(3):209.

Raykov, T., Dimitrov, D. M., and Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*, 17(2):265–279.

Reeves, E. (2015). A synthesis of the literature on trauma-informed care. *Issues in mental health nursing*, 36(9):698–709.

Ritchie, J. and Miles, R. E. (1970). An analysis of quantity and quality of participation as mediating variables in the participative decision making process. *Personnel Psychology*, 23(3).

Schenke, K., Nguyen, T., Watss, T., Sarama, J., and Clements, D. (2017). Differential effects of the classroom on african american and non-african american's mathematics achievement. *Journal of Educational Psychology*, 109(6):794–811.

Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Broda, M., Spicer, J., Bruner, J., Moeller, J., Linnansaari, J., Juuti, K., and Viljaranta, J. (2016). Investigating optimal learning moments in u.s. and finnish science classes. *Journal of Research in Science Teaching*, 53(3):400–421.

Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Klager, C., Bradford, L., Chen, I.-C., Baker, Q., Touitou, I., Peek-Brown, D., et al. (2022). Improving science achievement—is it possible? evaluating the efficacy of a high school chemistry and physics project-based learning intervention. *Educational Researcher*, 51(2):109–121.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.

Schünemann, N., Spörer, N., Völlinger, V. A., and Brunstein, J. C. (2017). Peer feedback mediates the impact of self-regulation procedures on strategy use and reading comprehension in reciprocal teaching groups. *Instructional Science*, 45(4):395–415.

Selig, J. P. and Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in human development*, 6(2-3):144–164.

Shen, Z., Li, W., and Leite, W. (2024). Statistical power and optimal design for randomized controlled trials investigating mediation effects. *Psychological Methods*.

Silva, B. C., Bosancianu, C. M., and Littvay, L. (2019). *Multilevel structural equation modeling*. Sage Publications.

Smith, T. E. and Sheridan, S. M. (2019). The effects of teacher training on teachers' family-engagement practices, attitudes, and knowledge: A meta-analysis. *Journal of Educational and Psychological Consultation*, 29(2):128–157.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639.

Spybrook, J., Unlu, F., Hedberg, E., Mulolli, D., and Berglund, T. (2022). Improving the design of evaluations that include students, teachers, and schools. Paper presented at the Annual Meeting for the American Educational Research Association, San Diego, CA.

Steyer, R., Mayer, A., Geiser, C., and Cole, D. A. (2015). A theory of states and traits—revised. *Annual review of clinical psychology*, 11:71–98.

Vongkulluksn, V. W. and Xie, K. (2022). Multilevel latent state-trait models with experience sampling data: An illustrative case of examining situational engagement. *Open Education*

Studies, 4(1):252–272.

Walberg, H. J. (1969). Social environment as a mediator of classroom learning. *Journal of Educational Psychology*, 60(6p1):443.

Wang, Y., Kim, E. S., Dedrick, R. F., Ferron, J. M., and Tan, T. (2018). A multilevel bifactor approach to construct validation of mixed-format scales. *Educational and psychological measurement*, 78(2):253–271.

Westine, C. D., Unlu, F., Taylor, J., Spybrook, J., Zhang, Q., and Anderson, B. (2020). Design parameter values for impact evaluations of science and mathematics interventions involving teacher outcomes. *Journal of Research on Educational Effectiveness*, 13(4):816–839.

Widaman, K. F. and Revelle, W. (2023). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior research methods*, 55(2):788–806.

Wiliam, D. (2022). How should educational research respond to the replication "crisis" in the social sciences? reflections on the papers in the special issue. *Educational Research and Evaluation*, 27(1-2):208–214.

Williams, N. J., Preacher, K. J., Allison, P. D., Mandell, D. S., and Marcus, S. C. (2022). Required sample size to detect mediation in 3-level implementation studies. *Implementation Science*, 17(1):66.

Zhang, Q. and Phillips, B. (2018). Three-level longitudinal mediation with nested units: How does an upper-level predictor influence a lower-level outcome via an upper-level mediator over time? *Multivariate behavioral research*, 53(5):655–675.

APPENDIX A

R SIMULATION CODE

Simulation Code

```
setwd("Z:/Hannah Chair/PIRE/DATA/Analysis/Lydia/Dissertation")
getwd()

library(MplusAutomation)
source("sim-functions.R")



sss = c(30, 60, 200, 500) #school sample size
tps = c(2,5,10,20) #teahcer per school
spt = c(20, 30, 60) #students per teacher

ape = c(0.15, 0.25, 0.45) #a path effects
bpe = c(0.05, 0.15, 0.25)#b path effects

gl = c(0.3, 0.6, 0.9) #general loadings
tl = c(0.3, 0.6, 0.9) #time specific loadings

repResult <- list()
attlist <- list()
stdResult <- list()
f3Result <- list()
bfResult <- list()
setwd("Z:/Hannah Chair/PIRE/DATA/Analysis/Lydia")
```

```
t = 2
s = 30
for (ss in sss){
     for (a in ape){
       for (b in bpe){
         for (g in gl){
           for (l in tl){
             setwd("Z:/Hannah Chair/PIRE/DATA/Analysis/Lydia/
                Dissertation") #reset directory
             dir.create(paste0("ss-",ss,"t-",t,"s-",s,"a-",a,"b-",b,"g
                -",g,"l-",l))#make new directory
             setwd(paste0("ss-",ss,"t-",t,"s-",s,"a-",a,"b-",b,"g-",g,"
                l-",l))#change directory to new directory

             att <- list(ss, t, s, a, b, g, l)
             print(att)
             attlist <- append(attlist,att)
             cat(fullmodelsim(ss, t, s, a, b, g, l, processors=8),file
                = file.path(getwd(),paste0(ss,t,s,a,b,g,l,".inp")))
             runModels(file.path(getwd(),paste0(ss,t,s,a,b,g,l,".inp"))
                )
             repResult <- append(repResult,readModels(file.path(getwd()
                ,paste0(ss,t,s,a,b,g,l,".out"))))
             cat(stzmed(ss, t, s, a, b, g, l, processors=8),file = file
                .path(getwd(),paste0("std -",ss,t,s,a,b,g,l,".inp")))
             runModels(file.path(getwd(),paste0("std -",ss,t,s,a,b,g,l
                ,".inp")))
```

```
    stdResult <- append(stdResult,readModels(file.path(getwd(),paste0("std -",ss,t,s,a,b,g,l,".out"))))
    flist <- list()
    bflist <- list()
    for (i in 1:100){
      cat(threefactors(ss, t, s, a, b, g, l,i, processors=8) , file = file.path(getwd(),paste0("3f -",ss,t,s,a,b,g,l,i,".inp")))
      runModels(file.path(getwd(),paste0("3f -",ss,t,s,a,b,g,l,i,".inp")))
      cat(bifactors(ss, t, s, a, b, g, l,i, processors=8) , file = file.path(getwd(),paste0("bf -",ss,t,s,a,b,g,l,i,".inp")))
      runModels(file.path(getwd(),paste0("bf -",ss,t,s,a,b,g,l,i,".inp")))
      flist[i] <- paste0("3F", ss,t,s,a,b,g,l,"REP",i,".dat")
      bflist[i] <- paste0("BF", ss,t,s,a,b,g,l,"REP",i,".dat")
    }
    fdata <- data.frame(matrix(unlist(flist),nrow = 100, byrow=TRUE))
    bdata <- data.frame(matrix(unlist(flist),nrow = 100, byrow=TRUE))
    write.table(fdata, file = paste0("3F", ss,t,s,a,b,g,l,"REPlist.dat"), row.names=FALSE,quote=FALSE,col.names=FALSE)
    write.table(bdata, file = paste0("BF", ss,t,s,a,b,g,l,"REPlist.dat"), row.names=FALSE,quote=FALSE,col.names=
```

```
                FALSE)
              cat(f3med(ss, t, s, a, b, g, l, processors=8),file = file.
                path(getwd(),paste0("f3med -",ss,t,s,a,b,g,l,".inp")))
              runModels(file.path(getwd(),paste0("f3med -",ss,t,s,a,b,g,
                l,".inp")))
              f3Result <- append(f3Result,readModels(file.path(getwd(),
                paste0("f3med -",ss,t,s,a,b,g,l,".out"))))
              cat(bfmed(ss, t, s, a, b, g, l, processors=8),file = file.
                path(getwd(),paste0("bfmed -",ss,t,s,a,b,g,l,".inp")))
              runModels(file.path(getwd(),paste0("bfmed -",ss,t,s,a,b,g,
                l,".inp")))
              bfResult <- append(bfResult,readModels(file.path(getwd(),
                paste0("bfmed -",ss,t,s,a,b,g,l,".out"))))
            }
          }
        }
      }
}




save(repResult,attlist,stdResult,f3Result,bfResult, file="sim.RData")
```

Simulation functions

```
### Generates the MPLUS syntax for all the simulations

```

```
fullmodelsim <- function(ss, t, s, a, b, g, l, processors=8){
  nobs <- ss*t*s
  te <- 0.2 - a*b
  syntax <- paste0(
    "TITLE: \n",
    "Fully Specified Model Simulation - School N =", ss, " Teacher N
       =", t,
    " Student N =", s, " a =",a," b =", b, " trait loadings =", g, "
       state loadings =",l, "\n",
    "MONTECARLO: \n",
    "NAMES ARE T PT1 PT2 PT3 PT4 TO11 TO21
    TO31 TO12 TO22 TO32 TO13 TO23 TO33;
    CATEGORICAL = PT1 PT2 PT3 PT4; \n",
    "BETWEEN = TO11 TO21
    TO31 TO12 TO22 TO32 TO13 TO23 TO33 (level3)T; \n",
    "GENERATE = PT1-PT4(1); \n",
    "CUTPOINTS = T(0);\n",
    "NOBSERVATIONS =",nobs,";\n",
    "NCSIZES = 1[1]; \n",
    "CSIZES =",ss,"[",t,"(",s,")]; \n",
    "NREP = 100;\n",
    "SEED = 2024; \n",
    "REPSAVE = ALL; \n",
    "SAVE =", ss,t,s,a,b,g,l,"REP*.dat; \n",
    "ANALYSIS: \n",
    "TYPE IS THREELEVEL \n;",
    "processors =",processors, ";\n",
```

```
"ALGORITHM=GIBBS(RW); \n",
"MODEL POPULATION: \n",
"%WITHIN% \n",
"OUTCOME BY PT1@1 PT2*0.5 PT3*0.8 PT4*0.3;\n",
"OUTCOME*1;\n",
"%BETWEEN level2% \n",
"OUTCOMET BY PT1@1 PT2*0.5 PT3*0.8 PT4*0.3;\n",
"OUTCOMET*.85; \n",
"MED BY TO11*",g," TO21-TO33*",g, ";\n",
"MEDT1 BY TO11*",l," TO21-TO31*",l,"; \n",
"MEDT2 BY TO12*",l," TO22-TO32*",l,"; \n",
"MEDT3 BY TO13*",l," TO23-TO33*",l,"; \n",
"MED-MEDT3@1; \n",
"MED-MEDT3 with MED-MEDT3@0;\n",

"%BETWEEN level3% \n",
"MEDB BY TO11*",g," TO21-TO33*",g, ";\n",
"MEDB@.2;\n",
"OUTCOMEB BY PT1@1.0 PT2*0.5 PT3*0.8 PT4*0.3;\n",
"MEDB ON T*",a,"; \n",
"OUTCOMEB ON MEDB*", b," T*",te,";\n",
"OUTCOMEB*.8; \n",

"MODEL: \n",
"%WITHIN% \n",
"OUTCOME BY PT1@1 PT2*0.5 PT3*0.8 PT4*0.3; \n",
"OUTCOME*1;\n",
```

```
    "%BETWEEN level2% \n",
    "OUTCOMET BY PT1@1 PT2*0.5 PT3*0.8 PT4*0.3;\n",
    "OUTCOMET*.85; \n",
    "MED BY TO11*",g," TO21-TO33*",g, ";\n",
    "MEDT1 BY TO11*",l," TO21-TO31*",l,"; \n",
    "MEDT2 BY TO12*",l," TO22-TO32*",l,"; \n",
    "MEDT3 BY TO13*",l," TO23-TO33*",l,";\n",
    "MED-MEDT3@1;\n",
    "MED-MEDT3 with MED-MEDT3@0; \n",

    "%BETWEEN level3% \n",
    "MEDB BY TO11*",g," TO21-TO33*",g, "; \n",
    "MEDB@.2; \n",
    "OUTCOMEB BY PT1@1.0 PT2*0.5 PT3*0.8 PT4*0.3;\n",
    "MEDB ON T*",a,";\n",
    "OUTCOMEB ON MEDB*", b," T*",te,";\n",
    "OUTCOMEB*.8; \n"


  )
  return(syntax)
}

stzmed <- function(ss, t, s, a, b, g, l, processors=8){
    te <- 0.2 - a*b
    syntax <- paste0(
      "TITLE: \n",
```

```
    "SIMULATION OF MT-MEDIATION SCHOOL N = ", ss, ", TEACHER N = ",t
        , ", STUDENT N = ", s,", A=", a, ", B= ",b,", GENERAL LOADING
        = ",g,", TIME LOADING = ",l," - STD MEDIATOR \n",
    "DATA: \n",
    "FILE IS ", ss,t,s,a,b,g,l,"REPlist.dat; \n",
    "TYPE = MONTECARLO; \n",
    "VARIABLE: \n",
    "NAMES ARE PT1 PT2 PT3 PT4 TO11 TO21
  TO31 TO12 TO22 TO32 TO13 TO23 TO33 T Tid Sid TO1 TO2 TO3 TOM;\n",
    "CATEGORICAL = PT1 PT2 PT3 PT4; \n",
    "CLUSTER = Sid Tid; \n",
    "BETWEEN = TO11 TO12
    TO13 TO21 TO22 TO23 TO31 TO32 TO33 TO1 TO2 TO3 TOM (Sid) T;\n",
    "DEFINE: \n",
    "TO1 = (TO11 + TO21 + TO31)/3;\n",
    "TO2 = (TO12 + TO22 + TO32)/3;\n",
    "TO3 = (TO13 + TO23 + TO33)/3;\n",
    "TOM = (TO1 +TO2 + TO3)/3;\n",

    "ANALYSIS: \n",
    "TYPE IS THREELEVEL;\n",
    "ESTIMATOR=Bayes;\n",
    "processors = ",processors,";\n",
    "ALGORITHM=GIBBS(RW); \n",
    "MODEL:\n",
    "%WITHIN% \n",
    "OUTCOME BY PT1@1.0 PT2*0.5 PT3*0.8 PT4*0.3; \n",
```

```
      "OUTCOME*1; \n",
      "%Between Tid% \n",
      "OUTCOMET BY PT1@1 PT2*0.5 PT3*0.8 PT4*0.3; \n",
      "OUTCOMET*.85;\n",
      "TOM@1; \n",
      "%BETWEEN Sid% \n",
      "TOM@.2;\n",
      "OUTCOMEB BY PT1@1.0 PT2*0.5 PT3*0.8 PT4*0.3; \n",
      "OUTCOMEB*.2;\n",
      "TOM ON T*",a,"; \n",
      "OUTCOMEB ON TOM*",b," T*",te,";\n",
      "OUTPUT: TECH9; \n"
  )
    return(syntax)
}

threefactors <- function(ss, t, s, a, b, g, l, i, processors=8){
  syntax<- paste0(
    "TITLE: \n",
    "SIMULATION OF MT-MEDIATION SCHOOL N = ", ss, ", TEACHER N = ", t,
        ",
    STUDENT N = ", s, ", A = ",a,", B = ",b,", GENERAL LOADING = ",g
       ,", TIME LOADING = ",l," - save 3 factor scores \n",
    "DATA: \n",
    "FILE IS ", ss,t,s,a,b,g,l,"REP",i,".dat;\n",
    "VARIABLE: \n",
    "NAMES ARE PT1 PT2 PT3 PT4 TO11 TO21
```

```
    TO31 TO12 TO22 TO32 TO13 TO23 TO33 T Tid Sid;\n",
    "MODEL: \n",
    "MEDT1 BY TO11 TO21 TO31;\n",
    "MEDT2 BY TO12 TO22 TO32; \n",
    "MEDT3 BY TO13 TO23 TO33;\n",
    "MEDT1-MEDT3@1; \n",
    "SAVEDATA: \n",
      "save=fscores;\n",
    "FILE IS 3F", ss,t,s,a,b,g,l,"REP",i,".dat;\n"
  )
  return(syntax)
}

f3med <- function(ss, t, s, a, b, g, l, processors=8){
  te <- 0.2 - a*b
  syntax <- paste0(
    "TITLE: \n",
    "SIMULATION OF MT-MEDIATION SCHOOL N = ", ss, ", TEACHER N = ",t,
       ", STUDENT N = ", s,", A=", a, ", B= ",b,", GENERAL LOADING= ",
       g,", TIME LOADING = ",l," - 3 factor MEDIATOR \n",
    "DATA: \n",
    "FILE IS 3F", ss,t,s,a,b,g,l,"REPlist.dat; \n",
    "TYPE = MONTECARLO; \n",
    "VARIABLE: \n",
    "NAMES ARE PT1 PT2 PT3 PT4 TO11 TO21
    TO31 TO12 TO22 TO32 TO13 TO23 TO33 T Tid Sid TO1 TO1_SE TO2 TO2_SE
         TO3 TO3_SE TOM;\n",
```

```
"USEVAR = PT1 PT2 PT3 PT4 T Tid Sid TO1 TO2 TO3 TOM; \n",
"CATEGORICAL = PT1 PT2 PT3 PT4; \n",
"CLUSTER = Sid Tid; \n",
"BETWEEN = TO1 TO2 TO3 TOM (Sid) T;\n",
"DEFINE: \n",
"TOM = (TO1 +TO2 + TO3)/3;\n",

"ANALYSIS: \n",
"TYPE IS THREELEVEL;\n",
"ESTIMATOR=Bayes;\n",
"processors = ",processors,";\n",
"ALGORITHM=GIBBS(RW); \n",
"MODEL:\n",
"%WITHIN% \n",
"OUTCOME BY PT1@1.0 PT2*0.5 PT3*0.8 PT4*0.3; \n",
"OUTCOME*1; \n",
"%Between Tid% \n",
"OUTCOMET BY PT1@1 PT2*0.5 PT3*0.8 PT4*0.3; \n",
"OUTCOMET*.85;\n",
"TOM@1; \n",
"%BETWEEN Sid% \n",
"TOM@.2;\n",
"OUTCOMEB BY PT1@1.0 PT2*0.5 PT3*0.8 PT4*0.3; \n",
"OUTCOMEB*.2;\n",
"TOM ON T*",a,"; \n",
"OUTCOMEB ON TOM*",b," T*",te,";\n",
"OUTPUT: TECH9; \n"
```

```
  )
  return(syntax)
}

bifactors <- function(ss, t, s, a, b, g, l, i, processors=8){
  syntax<- paste0(
    "TITLE: \n",
    "SIMULATION OF MT-MEDIATION SCHOOL N = ", ss, ", TEACHER N = ", t,
        ",
    STUDENT N = ", s, ", A = ",a,", B = ",b,", GENERAL LOADING = ",g
       ,", TIME LOADING = ",l," - save bifactor scores \n",
    "DATA: \n",
    "FILE IS ", ss,t,s,a,b,g,l,"REP",i,".dat;\n",
    "VARIABLE: \n",
    "NAMES ARE PT1 PT2 PT3 PT4 TO11 TO21
    TO31 TO12 TO22 TO32 TO13 TO23 TO33 T Tid Sid;\n",
    "MODEL: \n",
    "GM BY TO11 TO12
    TO13 TO21 TO22 TO23 TO31 TO32 TO33; \n",
    "MEDT1 BY TO11 TO21 TO31;\n",
    "MEDT2 BY TO12 TO22 TO32; \n",
    "MEDT3 BY TO13 TO23 TO33;\n",
    " GM-MEDT3@1;\n",
    "GM-MEDT3 with GM-MEDT3@0;\n",
    "SAVEDATA: \n",
    "save=fscores;\n",
    "FILE IS BF", ss,t,s,a,b,g,l,"REP",i,".dat;\n"
```

```
  )
  return(syntax)
}

bfmed <- function(ss, t, s, a, b, g, l, processors=8){
  te <- 0.2 - a*b
  syntax <- paste0(
    "TITLE: \n",
    "SIMULATION OF MT-MEDIATION SCHOOL N = ", ss, ", TEACHER N = ",t,
       ", STUDENT N = ", s,", A=", a, ", B= ",b,", GENERAL LOADING= ",
       g,", TIME LOADING = ",l," - 3 factor MEDIATOR \n",
    "DATA: \n",
    "FILE IS 3F", ss,t,s,a,b,g,l,"REPlist.dat; \n",
    "TYPE = MONTECARLO; \n",
    "VARIABLE: \n",
    "NAMES ARE PT1 PT2 PT3 PT4 TO11 TO21
    TO31 TO12 TO22 TO32 TO13 TO23 TO33 T Tid Sid \n",
    "TOG TOG_SE TO1 TO1_SE TO2 TO2_SE TO3 TO3_SE;\n",
    "USEVAR = PT1 PT2 PT3 PT4 T Tid Sid TOG; \n",
    "CATEGORICAL = PT1 PT2 PT3 PT4; \n",
    "CLUSTER = Sid Tid; \n",
    "BETWEEN = TOG (Sid) T;\n",

    "ANALYSIS: \n",
    "TYPE IS THREELEVEL;\n",
    "ESTIMATOR=Bayes;\n",
    "processors = ",processors,";\n",
```

```
    "ALGORITHM=GIBBS(RW); \n",
    "MODEL:\n",
    "%WITHIN% \n",
    "OUTCOME BY PT1@1.0 PT2*0.5 PT3*0.8 PT4*0.3; \n",
    "OUTCOME*1; \n",
    "%Between Tid% \n",
    "OUTCOMET BY PT1@1 PT2*0.5 PT3*0.8 PT4*0.3; \n",
    "OUTCOMET*.85;\n",
    "TOG@1; \n",
    "%BETWEEN Sid% \n",
    "TOG@.2;\n",
    "OUTCOMEB BY PT1@1.0 PT2*0.5 PT3*0.8 PT4*0.3; \n",
    "OUTCOMEB*.2;\n",
    "TOG ON T*",a,"; \n",
    "OUTCOMEB ON TOG*",b," T*",te,";\n",
    "OUTPUT: TECH9; \n"
  )
  return(syntax)
}
```

APPENDIX B

EMPIRICAL STUDY MPLUS CODE

CESE Treatment Effect

```
TITLE: CESE TREATMENT - STDZ OUTCOME - WITH PRIORS
DATA: FILE IS tobs.csv;
VARIABLE: NAMES ARE STUID SID CHEM RID TID PRE1-PRE9
                  PHY1-PHY12 CHEM1-CHEM25 T POSTEQ PRESUB SPRESUB
                     OB1T1-OB1T11
                  OB2T1-OB2T11 OB3T1-OB3T11 OB4T1-OB4T11
                  OB5T1-OB5T11;
         USEVAR = SID RID TID T
                   POSTEQ CHEM PRESUB SPRESUB;
         MISSING ARE .;
         CLUSTER = SID TID;
         WITHIN = PRESUB;
         BETWEEN = (TID) CHEM SPRESUB(SID)T RID;
ANALYSIS: TYPE IS THREELEVEL;
          ESTIMATOR=BAYES;
          ALGORITHM=GIBBS(RW);
MODEL:
%WITHIN%
POSTEQ ON PRESUB;
%BETWEEN TID%
POSTEQ ON CHEM SPRESUB;
POSTEQ (TAU1);
%BETWEEN SID%
```

```
   POSTEQ (TAU2);
   POSTEQ ON T RID;

MODEL PRIORS:
   TAU1~IG(2.25,0.16);
   TAU2 ~ IG(2.25,0.19);
```

Mediation 1: Standardized averages of the mediator

```
TITLE: CESE MEDIATION - Standardized MED AND OUTCOME scores - WITH
   INFORMATIVE PRIORS
DATA: FILE IS tobs.csv;
VARIABLE: NAMES ARE STUID SID CHEM RID TID PRE1-PRE9
                 PHY1-PHY12 CHEM1-CHEM25 T POSTEQ PRESUB SPRESUB
                    OB1T1-OB1T11
                 OB2T1-OB2T11 OB3T1-OB3T11 OB4T1-OB4T11
                 OB5T1-OB5T11;
        USEVAR =SID RID TID T
                  POSTEQ CHEM PRESUB SPRESUB OB1T1-OB1T4
                 OB2T1-OB2T4 OB3T1-OB3T4 TO1 TO2 TO3 TOM;
        MISSING ARE .;
        CLUSTER = SID TID;
        WITHIN = PRESUB;
        BETWEEN = OB1T1-OB3T4 TO1 TO2 TO3 TOM (TID) CHEM SPRESUB (SID
           )T RID;

DEFINE:
TO1 = (OB1T1 + OB1T2 + OB1T3 + OB1T4)/4;
```

```
TO2 = (OB2T1 + OB2T2 + OB2T3 + OB2T4)/4;
TO3 = (OB3T1 + OB3T2 + OB3T3 + OB3T4)/4;
TOM = (TO1 +TO2 + TO3)/3;

ANALYSIS: TYPE IS THREELEVEL;
            ESTIMATOR=Bayes;
            ALGORITHM=GIBBS(RW);

MODEL:
%WITHIN%
POSTEQ ON PRESUB;
%BETWEEN TID%
    POSTEQ ON TOM (B1)
     CHEM SPRESUB;
     POSTEQ (TAU1);
     TOM ON CHEM SPRESUB;
%BETWEEN SID%
    POSTEQ (TAU2);
    TOM (TAU3);
    TOM ON T (A)
     RID;
    POSTEQ ON T RID;

MODEL PRIORS:
    TAU1~IG(2.25,0.16);
    TAU2 ~ IG(2.25,0.19);
    TAU3~IG(2.12,.12);
```

```

 MODEL CONSTRAINT:
    NEW (IND1);
    IND1 = A*B1;

OUTPUT: STANDARDIZED CINTERVAL;
```

Mediation 2: Averages of the factor socres

factor score estimation

```
TITLE: CESE MEDIATION - SAVE FACTOR SCORES;
DATA: FILE IS tobs.csv;
VARIABLE: NAMES ARE STUID SID CHEM RID TID PRE1-PRE9
                    PHY1-PHY12 CHEM1-CHEM25 T POSTEQ PRESUB SPRESUB
                        OB1T1-OB1T11
                    OB2T1-OB2T11 OB3T1-OB3T11 OB4T1-OB4T11
                    OB5T1-OB5T11;
          USEVAR ARE SID RID TID T
                     POSTEQ CHEM PRESUB SPRESUB OB1T1-OB1T4
                    OB2T1-OB2T4 OB3T1-OB3T4;
          MISSING ARE .;
          CLUSTER = TID;
          BETWEEN = OB1T1-OB3T4;

ANALYSIS: TYPE IS TWOLEVEL;
           ESTIMATOR =Bayes;

MODEL:
```

```
%WITHIN%

%BETWEEN%
MEDT1 BY OB1T1* OB1T2 OB1T3 OB1T4;
MEDT2 BY OB2T1* OB2T2 OB2T3 OB2T4;
MEDT3 BY OB3T1* OB3T2 OB3T3 OB3T4;
MEDT1-MEDT3@1;

SAVEDATA:
save=fscores(50 10);
FILE IS 3FSCORES.csv;
```

Estimation of Mediation

```
TITLE: CESE MEDIATION - AVERGAED FACTOR SCORES - INFPRIORS
  DATA: FILE IS 3FSCORES.csv;
  VARIABLE: NAMES ARE OB1T1-OB1T4
                      OB2T1-OB2T4 OB3T1-OB3T4
                      SID RID T POSTEQ CHEM PRESUB SPRESUB MEDT1 MEDT1m
                          MEDT1sd
                      MEDT125 MEDT198 MEDT2 MEDT2m MEDT2sd
                      MEDT225 MEDT298 MEDT3 MEDT3m MEDT3sd
                      MEDT325 MEDT398 IG1-IG35 TID;
            USEVAR = SID RID T
                       POSTEQ CHEM PRESUB SPRESUB MEDT1 MEDT2 MEDT3 TID
                           TOM;
            MISSING ARE *;
            CLUSTER = SID TID;
```

```
          WITHIN = PRESUB;
          BETWEEN = MEDT1 MEDT2 MEDT3 TOM (TID)CHEM SPRESUB (SID)T
             RID;

DEFINE:
TOM = (MEDT1 + MEDT2 + MEDT3)/3;

STANDARDIZE TOM;

ANALYSIS: TYPE IS THREELEVEL;
            ESTIMATOR=Bayes;
            ALGORITHM=GIBBS(RW);

MODEL:
   %WITHIN%
POSTEQ ON PRESUB;
%BETWEEN TID%
    POSTEQ ON TOM (B1)
     SPRESUB CHEM;
     TOM ON SPRESUB CHEM;
     POSTEQ (TAU1);
%BETWEEN SID%
    POSTEQ (TAU2);
    TOM (TAU3);
    TOM ON T (A)
     RID;
    POSTEQ ON T RID;
```

```
MODEL PRIORS:
    TAU1~IG(2.25,0.16);
    TAU2 ~ IG(2.25,0.19);
    TAU3~IG(2.12,.12);


   MODEL CONSTRAINT:
      NEW (IND1);
      IND1 = A*B1;

  OUTPUT: CINTERVAL;
```

Mediation 3: Factors from LST

Factor scores estimation

```
TITLE: CESE MEDIATION - SAVE BIFACTOR SCORES;
DATA: FILE IS tobs.csv;
VARIABLE: NAMES ARE STUID SID CHEM RID TID PRE1-PRE9
                    PHY1-PHY12 CHEM1-CHEM25 T POSTEQ PRESUB SPRESUB
                        OB1T1-OB1T11
                    OB2T1-OB2T11 OB3T1-OB3T11 OB4T1-OB4T11
                    OB5T1-OB5T11;
          USEVAR ARE SID RID TID T
                     POSTEQ CHEM PRESUB SPRESUB OB1T1-OB1T4
                    OB2T1-OB2T4 OB3T1-OB3T4;
          MISSING ARE .;
          CLUSTER = SID TID;
          WITHIN = PRESUB;
```

```
        BETWEEN = (TID) OB1T1-OB3T4 CHEM SPRESUB (SID)T RID;

ANALYSIS: TYPE IS THREELEVEL;
          ESTIMATOR=BAYES;
         ALGORITHM=GIBBS(RW);
MODEL:

%BETWEEN TID%
   MED BY OB1T1* OB1T2-OB3T4;
   MEDT1 BY OB1T1* OB1T2-OB1T4;
   MEDT2 BY OB2T1* OB2T2-OB2T4;
   MEDT3 BY OB3T1* OB3T2-OB3T4;
   MED-MEDT3@1;
   MED-MEDT3 WITH MED-MEDT3@0;


SAVEDATA:
save=fscores(50 10);
FILE IS BFSCORES.csv;
```

Mediation estimation

```
TITLE: CESE MEDIATION - BIFACTOR SCORES - INFPRIORS
DATA: FILE IS BFSCORES.csv;
VARIABLE: NAMES ARE RID T POSTEQ CHEM PRESUB SPRESUB OB1T1-OB1T4
                 OB2T1-OB2T4 OB3T1-OB3T4
                 TOM TOMm TOMsd TOM25 TOM98 MEDT1 MEDT1m MEDT1sd
                 MEDT125 MEDT198 MEDT2 MEDT2m MEDT2sd
```

```
                  MEDT225 MEDT298 MEDT3 MEDT3m MEDT3sd
                  MEDT325 MEDT398 IG1-IG10 SID TID;
         USEVAR = SID RID T
                   POSTEQ CHEM PRESUB SPRESUB TID TOM;
         MISSING ARE *;
         CLUSTER = SID TID;
         WITHIN = PRESUB;
         BETWEEN = TOM(TID)CHEM SPRESUB (SID)T RID;


ANALYSIS: TYPE IS THREELEVEL;
           ESTIMATOR=Bayes;
           ALGORITHM=GIBBS(RW);

MODEL:
   %WITHIN%
POSTEQ ON PRESUB;
%BETWEEN TID%
    POSTEQ ON TOM (B1)
     SPRESUB CHEM;
     TOM ON SPRESUB CHEM;
     POSTEQ (TAU1);
%BETWEEN SID%
    POSTEQ (TAU2);
    TOM (TAU3);
    TOM ON T (A)
     RID;
```

```
   POSTEQ ON T RID;

MODEL PRIORS:
     TAU1~IG(2.25,0.16);
     TAU2 ~ IG(2.25,0.19);
     TAU3~IG(2.12,.12);

 MODEL CONSTRAINT:
   NEW (IND1);
   IND1 = A*B1;
```

Mediation 4: Fully specified model

```
TITLE: CESE full mediation model - STDZ OUTCOME -inprior
DATA: FILE IS tobs.csv;
VARIABLE: NAMES ARE STUID SID CHEM RID TID PRE1-PRE9
                 PHY1-PHY12 CHEM1-CHEM25 T POSTEQ PRESUB SPRESUB
                    OB1T1-OB1T11
                 OB2T1-OB2T11 OB3T1-OB3T11 OB4T1-OB4T11
                 OB5T1-OB5T11;
        USEVAR = SID RID TID T
                  POSTEQ CHEM PRESUB SPRESUB OB1T1-OB1T4
                 OB2T1-OB2T4 OB3T1-OB3T4;
        MISSING ARE .;
        CLUSTER = SID TID;
        WITHIN = PRESUB;
        BETWEEN = OB1T1-OB1T4
```

```
                  OB2T1-OB2T4 OB3T1-OB3T4 (TID) CHEM SPRESUB (SID)T
                     RID;
ANALYSIS: TYPE IS THREELEVEL;
          ESTIMATOR=BAYES;
          ALGORITHM=GIBBS(RW);
          BITERATIONS = 100000;
MODEL:
%WITHIN%
POSTEQ ON PRESUB;
%BETWEEN TID%
   POSTEQ ON MED (B1)
    CHEM SPRESUB;
    MED ON CHEM SPRESUB;
   MED BY OB1T1* OB1T2-OB3T4;
   MEDT1 BY OB1T1* OB1T2-OB1T4;
   MEDT2 BY OB2T1* OB2T2-OB2T4;
   MEDT3 BY OB3T1* OB3T2-OB3T4;
   MED-MEDT3@1;
   MED-MEDT3 WITH MED-MEDT3@0;
   POSTEQ (TAU1);
%BETWEEN SID%
   POSTEQ (TAU2);
   MEDB BY OB1T1* OB1T2-OB3T4;
   MEDB (TAU3);
   MEDB ON T (A)
    RID;
   POSTEQ ON T RID;
```

```
MODEL PRIORS:
     TAU1~IG(2.25,0.16);
     TAU2 ~ IG(2.25,0.19);
     TAU3~IG(2.12,.12);

 MODEL CONSTRAINT:
    NEW (IND1);
    IND1 = A*B1;
```