CONCEPTUALIZING CHANGE IN TEACHING AND LEARNING THROUGH STRUCTURAL EQUATION MODELING

By Mary J. Schaal

A Dissertation

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

in Educational Leadership

Northern Arizona University

May 2010

Approved:

Mary I. Dereshiwsky, Ph.D., Chair

Gary L. Emanuel, Doctor of Arts

Shirley M. Hord, Ph.D.

George J. Montopoli, Ph.D.

Richard L. Wiggall, Ed.D.

ABSTRACT

CONCEPTUALIZING CHANGE IN TEACHING AND LEARNING THROUGH STRUCTURAL EQUATION MODELING

MARY J. SCHAAL

Quality mathematics teaching that results in student learning is considered critical to heighten American competitiveness. Evaluation for verification of results of promising approaches in mathematics education is equally important for the achievement of this goal. In this study, data were reanalyzed from a study conducted by George, Hall, and Uchiyama (2000), documenting a highly-successful district-wide change in mathematics teaching and learning in a manner closely aligned with National Council of Teachers of Mathematics standards (1989). Well-specified data were collected using the Concerns-Based Adoption Model (Hall & Hord, 2006).

In this correlational, causal-comparative dissertation study, data were re-analyzed using first- and second-generation latent structural equation modeling approaches, providing insight into relationships among student outcomes and instructional quality in grades 2-8 classrooms with respect to levels of implementation behavior and fidelity of implementation of constructivist approaches to teaching mathematics. Second-generation structural equation models provided a lens through which to view dynamics of change. A model associates quality of instruction with student achievement, along with recommendations for future research.

ACKNOWLEDGEMENTS

Winn Griffin, for guiding me into critical thinking and scholarship;

Olivia Zepeda, for dreaming big for mathematical excellence in teaching and learning for all of “our” students;

Marge Barcomb and George Montopoli, for a solid foundation in statistics;

Shirley Hord, for conceptualizing professional learning communities, training me in CBAM (Hall & Hord, 2006), taking me seriously as a scholar, and for your friendship—you are my role model, the *a* variation of a friend and mentor;

Gene Hall, for foundational research on change, and permitting me to study the CBAM (2006) and your evaluation of District A;

Archie George, for your excellent research, allowing me replicate your study, and for many hours of hard work de-identifying and providing data;

The Docs to Be Study Group: Darwin Stiffler, Ed Richardson, Suzie Alka and Laura Noel, for your friendship, dedication, wisdom, and humor – it was an honor learning with you;

My Chair, Mary Dereshiwsky, for teaching me almost everything I know about research, and walking with me step by step through my scholarly journey;

My Committee: For freely giving your time and expertise to help me to succeed;

Trudy Mitchell, Donna Griffin, Ruth Whisler, Michael Todd, Vicki Bardo, Paul Neuman, Ruth Elliott, Kelli Clay, Elizabeth Hall, and Sonja Cushman Lowe, for cheerleading, encouraging, praying, and caring daily through my entire journey;

My husband, Robert “Dr Bob” Schaal, for your love, support, patience, sacrifices, and encouragement; you make my scholarship possible. Thank you!

TABLE OF CONTENTS

Chapter One: Introduction 1

Background 1

Purpose of the study 2

Statement of the problem 3

Definition of terms 3

Abbreviations used 6

Limitations 6

Delimitations 8

Assumptions 8

Significance of the study 9

Summary 10

Chapter Two: Review of Literature 12

Introduction 12

Development of the Concerns-Based Adoption Model 12

Innovation configurations 15

Specifics of IC mapping 15

Innovation configurations related to this study 16

Stages of concern 18

Levels of use 20

LoU of mathematics curricula in District A 23

Concluding statement regarding CBAM 24

Standards-Based Education 24

Educational standards 25

DoDDS adoption of constructivist approaches to teaching mathematics ..26

Measurement of learning 26

Change and Student Achievement 27

Importance of fidelity in implementation 28

George, Hall and Uchiyama's findings in District A, DoDDS 28

Summary 41

Chapter Three: Research Methodology 43

Introduction 43

Restatement of the problem 43

Research design and procedures 43

Research methodology 44

Population and sample 46

Sources of information 46

Data analysis procedures 46

Revised research methodology 47

Revised data analysis procedures 48

Chapter Four: Findings and Results 49

Introduction 49

Findings and results 50

Assumptions 50

Model relationships … 54

Goodness of fit 56

Calculation of estimates of the effects of the model 57

Revised data analysis procedures...59

Step 0. Descriptive statistics 61

Step 1. Measurement model alternatives … 63

Step 2. Transitions based on cross-sectional results 66

Step 3. Specification of the latent transition model without covariates.....67

Step 4. Inclusion of covariates 68

Final results 69

Compare present results to results from the original analyses..............................85

Summary 87

Chapter Five: Summary, Conclusions, Recommendations, and Implications...................88

Introduction..88

Summary of the study 88

Summary of findings..92

Summary of conclusions..95

Recommendations..98

Recommendations for practice 98

Recommendations for future research 99

Implications..102

Summary 104

References..106

Appendix A 113

Appendix B 115

Appendix C 118

Appendix D 119

Appendix E 120

Appendix F 121

Appendix G 123

LIST OF TABLES

Table 1. Attendance in Standards-Based Training and Associated Classroom GOALS Scores....31

Table 2. Teachers' Innovation Configuration Map Ratings and Classroom GOALS Scores, 1996-97....33

Table 3. Teachers' Innovation Configuration Map Ratings and Classroom GOALS Scores, 1996-97 and 1997-98 Combined....36

Table 4. Student Achievement as a Function of IC Map Ratings....37

Table 5. Summary Descriptive Statistics in Data Set....49

Table 6. Fall and Spring GOALS Scores....62

Table 7. Latent Class Analysis, Latent Class Factor Analysis, and Factor Mixture Analysis Model Results for Fall and Spring GOALS Scores (N=2143)....64

Table 8. Distribution of Students into Latent Performance Classes....65

Table 9. Number of Students in Each Learner Class in Fall and Spring GOALS Exams, based on Cross-Sectional LCA without Covariates....66

Table 10. Fit Indices for 4, 3 and 2 Latent Classes without Covariates....67

Table 11. Comparative Model Fit Indices for 4, 3 and 2 Latent Classes with Covariates....68

Table 12. Classification of Individuals based on their Most Likely Latent Class Membership....70

Table 13. Average Latent Class Probabilities for Most Likely Membership by Latent Class....70

Table 14. Latent Class Analysis Model Results, Part 1....73

Table 15. Latent Class Analysis Model Results, Part 2: Estimated Effects of Categorical Latent Variables, Compared to Reference Class 2 (*Novice*)....74

Table 16. Logistic Regression Odds Ratios, using Class 4 (Master) as the Reference Class....77

Table 17. Classification of Individuals based on Most Likely Latent Class Membership 78

Table 18. Average Latent Class Probabilities for Most Likely Latent Class Membership by Latent Class....79

Table 19. Latent Class Analysis Model Results, Part 1 Comparing Impacts of Medium and Low Quality Instruction on Student Learning....80

Table 20. Latent Class Analysis Model Results, Part 2: Estimated Effects of Categorical Latent Variables, Compared to Reference Class 1 (*Novice*)....81

Table 21. Logistic Regression Odds Ratios, using Class 4 (*Apprentice*) as the Reference Class....84

LIST OF FIGURES

Figure 1. Stages of Concern: Typical Concerns about the Innovation 19

Figure 2. Levels of Use of the Innovation 20

Figure 3. Proposed Model for Change in Teaching and Learning Mathematics 45

Figure 4. Scatter Plot of Fall and Spring GOALS Scores...52

Figure 5. Scatter Plot of Standardized Residuals and Standardized Predicted Values......54

Figure 6. Structural Equation Model used for this Analysis...55

Figure 7. Estimates of Effects 58

Figure 8. Scatter Plot of Latent Classes by Fall (*goalf*) and Spring (*goals*) GOALS Scores..71

Figure 9. Latent Classes by Average Fall and Spring GOALS Scores..............................72

Figure 10. Logic Model for Presentation of Results to Research Question 1 and Additional Sub-question 91

Figure 10. Sample Components from the IC Map for Teaching and Learning Mathematics 113

DEDICATION

To the students of Yuma and La Paz Counties, Arizona, who inspire me to dream big for high quality teaching and learning.

CHAPTER ONE: INTRODUCTION

Background

From Sputnik I to *A Nation at Risk* (National Center for Excellence in Education, 1983), Americans have accepted a relationship between education and American competitiveness, whether in terms of winning the Cold War and space race, or in thriving in the global economy. This link was demonstrated in 2006, when the National Math Panel was formed:

> To help keep America competitive, support American talent and creativity, encourage innovation throughout the American economy, and help State, local, territorial, and tribal governments give the Nation's children and youth the education they need to succeed, it shall be the policy of the United States to foster greater knowledge of and improved performance in mathematics among American students (Federal Register, Vol. 71, No. 77, 20519-20521).

Though much has changed politically, high expectations for teaching and learning math as a foundation for American competitiveness remain unchanged. Rep. Mike Honda, then-senator Barack Obama, and Sen. Richard Lugar co-sponsored *The Enhancing Science, Technology, Engineering, and Mathematics Education Act of 2008* (S. 3047, H.R. 6104) bill in May 2008 to enhance coordination for fulfillment of the America Competes Act.

The urgency for improving knowledge and performance in mathematics through teaching and learning is virtually unquestioned. A clarion call has gone out for schools to implement "research-based" approaches. Yet in their haste to comply, schools often overlook the matter of research verification of results during implementation (George, Hall, & Uchiyama, 2000). Such verification ensures that the results reported in research are being acquired in the local setting as a result of implementation.

Purpose of the Study

An important set of papers was published in 1999 and 2000 regarding a systematic change in teaching and learning mathematics (Hall, 1999, 2000; Thornton & West, 1999; Alquist & Hendrickson, 1999; Johnson, 2000; and George, Hall & Uchiyama, 2000). Dr. Gene Hall led a team that documented and evaluated the change process related to implementation of the constructivist math curriculum in District A of Department of Defense Dependents' Schools. The team collected well-specified data from 1996-98, and reported initial results. Findings indicated that higher levels of implementation behaviors at the classroom level and fidelity in implementation were associated with higher levels of student achievement. The research team invited replication and extension of their studies, especially related to the impact of fidelity in implementation for student learning.

Until now, these data have not been revisited. Since the study was conducted, new statistical techniques for modeling complex relationships have been developed and refined. One of these techniques, structural equation modeling, can provide nuanced insights into the interaction of measured variables that can be combined into latent variables. Established theory can be confirmed, by translating relationships into equations and comparing results from a sample covariance matrix to those of an estimated population covariance matrix. In this confirmatory study, the researcher will reanalyze this body of data through structural equation modeling, a statistical technique that was not available at the time of the original study.

Statement of the Problem

The purpose of this study is to:

1. model relationships among student outcomes, levels of implementation behavior, and fidelity of implementation of constructivist approaches to teaching mathematics, and
2. to compare results from structural equation modeling to results from the original analyses, in hopes of identifying potential similarities and differences in statistical methodologies for viewing the dynamics of change in teaching and learning mathematics.

Definition of Terms

Concerns-based adoption model: Three constructs that provide diagnostic tools to measure the change process and provide support: innovation configurations (defined below), stages of concern (predictable concerns in response to change; targeted support can enhance the change process), and levels of use of the innovation (defined below) (Hall & Hord, 2006).

Constructivist approaches to teaching mathematics: Approaches to teaching mathematics described by Alquist and Hendrickson (1999) and defined in their related innovation configuration map. In their words (1999, p. 18-19):

> … both [elementary and middle school] programs de-emphasize computation taught out of context and the rote memorization of rules and procedures that have been the mainstay of traditional mathematics instruction. Both programs emphasize using mathematics in the context of long term projects and investigations. Students are expected to work cooperatively, use a variety of techniques to solve problems, and justify their thinking, both orally and in

writing. Teachers are expected to use manipulative materials such as base 10 blocks extensively at all grade levels.

Department of Defense Dependents' Schools: K-12 schools provided by the Department of Defense (DoD) to provide quality education for minor dependents of U.S. active duty military and DoD civilian personnel stationed abroad (http://www.military.com/Resources/ ResourcesContent/0,13964,31992--,00.html referenced October 4, 2008). The Department of Defense Education Activity (DoDEA) "operates 192 schools in 14 districts located in 12 foreign countries, seven states, Guam, and Puerto Rico. All schools within DoDEA are fully accredited by U.S. accreditation agencies. Approximately 8,700 educators serve more than 84,000 DoDEA students" (http://www.dodea.edu/home/about.cfm?cId=facts referenced October 4, 2008). Data for this study were collected in District A and recoded by original researchers to preserve anonymity of participants.

Fidelity of implementation: Observed practices that range from ideal to unacceptable, as defined on an Innovation Configuration Map.

GOALS assessment: Performance-based mathematics assessment published by The Psychological Corporation for grades 2-8, aligned with NCTM standards.

Innovation: "The change itself" (Hall & Hord, 2006, p. 110). This encompasses behaviors and a mental image of the change desired. According to Hall and Hord (2006, p. 8), innovations can be products or processes (e.g., constructivist approach to teaching math), varying in complexity and time for implementation.

Innovation configuration map: A precise operational definition of the ideal implementation of the change desired and common adaptations scaled from ideal (*a*) to unacceptable (there may be a number of variations, generally from two to six or [*b*-*f*]). The purpose is to identify "different ways of doing the innovation" using a "number of components (typically eight to fifteen), and each component will have a number of variations (typically two to six). The number of components will vary depending on the complexity of the innovation and the amount of detail needed," (Hall & Hord, 2006, p. 116). Fidelity lines can be added to demark ideal implementation, acceptable variations, and unacceptable variations of the change (from the perspective of the IC Map developer and/or consensus group). An IC Map can be used to measure quality and fidelity of implementation of a given innovation (Hall & Hord, 2006, p. 129).

Levels of implementation behavior: Levels of implementation behavior correspond to levels of use of the innovation, ranging from 0-VI (Hall & Hord, 2006).

Levels of use of the innovation: A diagnostic tool of the Concerns-Based Adoption Model that describes "individuals' behaviors as they adopt and implement new ideas and innovations … the basis for describing where people are in the change process and for diagnosing their progress in implementing a change project" (Hall & Hord, 2006, p. 160). Behavior is scaled from 0-VI. The lowest three levels (0, I, and II) describe "nonusers" while levels III-VI describe users at different stages of implementation. Higher ratings relate to use designed for client benefits. Specific levels include: 0, nonuse; I, orientation; II, preparation; III, mechanical use; IVA routine; IVB refinement; V integration; and VI renewal (Loucks, Newlove, & Hall, referenced in Hall & Hord, 2006, p. 160).

Quality instruction: A variable constructed by the combination of teachers' implementation behavior and fidelity ratings. Ratings were consolidated into three categories: High (1), medium (2), and low (3). Levels of use (LoU) ratings of III were designated “low,” IVa was designated “medium,” and IVb and V were designated “high.” Innovation configuration/fidelity ratings (IC) were assigned to high, medium, and low categories by original researchers (described on p. 35). *Quality of instruction* relates specifically to participating teachers’ observed skill in implementing the constructivist mathematics approach adopted by District A (described on pages 3, 17, 18, 26, and in Appendix A) at specific times in 1996-98. It does not relate to the quality of teachers as instructors in a general sense.

Structural equation model (SEM): A system of regression equations specified to represent the underlying structure of a covariance matrix (Byrne, 1994).

Student outcomes: Mathematics learning achieved as measured by performance on GOALS performance-based assessments.

Abbreviations Used

CBAM: Concerns-based adoption model (Hall & Hord, 2006)

DoDDS: Department of Defense Dependents’ Schools

GMM: Growth mixture model

IC: Innovation configuration (Hall & Hord, 2006)

LCA: Latent class analysis

LoU: Levels of use of the innovation (Hall, et al., 1975)

NCTM: National Council of Teachers of Mathematics

SEM: Structural equation model

Limitations

Threats to internal validity in this design relate to:

1. Differences among groups that could be caused by unanticipated (confounding or intervening) variables. Two examples of confounding variables on a class' performance compared to others might be teacher illness and frequent absence during the year (substitute with a different level of comfort/experience teaching math in this manner) or something that would cause the class to perform less well on the pre- or post-test, such as an illness that impacted many in the class that altered their performance on the test. An example of an intervening variable is student participation in tutoring or a special program that enhanced math learning in addition to regular instruction compared to other classes that did not receive this. These could impact student learning or performance on the pre- or post-exam, in addition to factors that were measured. If such variables existed, it may be difficult to specify a model that adequately reflects reality.

2. Experimental mortality (students or teachers without complete data sets due to transferring, dropping- or stopping-out) could leave important factors unexamined, since only participants with complete data sets will be studied.

3. Apparent correlations could be erroneous. Causal relationships can only be identified with acceptable degrees of certainty in research with experimental control.

4. Data were collected in 1996-1998 (nearly ten years before the present study). Recovery of data and use of historical data presented unanticipated challenges in reaching conclusions when a paucity of data was discovered.

Multiple sources of evidence will be used to establish construct validity (Yin, 1994). A multi-method approach (Brewer & Hunter, 1989) will mitigate threats to internal validity. Observational data recording fidelity of implementation strengthen validity of self-reported data regarding extent of implementation. By matching observational and self-reported data related to implementation with student achievement data, concerns related to spurious correlations will be minimized. However, any apparent causal relationships would be recommended for future study.

Delimitations

Sources of information for this study were collected from teachers and grades 2-8 students in Department of Defense Dependents' Schools in District A, 1996-1998. Therefore, findings and results may not necessarily generalize to other subpopulations, locations, and/or time periods.

Assumptions

The following assumptions were derived from Ullman (2007, pp. 682-683).

Sample Size and Missing Data

Sample sizes of at least 60 (Bentler & Yuan, 1999) are required to stabilize covariances and parameter estimates. Minimum sample sizes required for sufficient power to test of goodness of fit were developed by MacCallum, Browne, and Sugawara (1996).

Multivariate Normality and Outliers

Multivariate normality is assumed among dependent and independent variables with respect to skewness, kurtosis, and outliers.

Linearity

Linear relations can be assessed in pairs of measured variables through the generation of bivariate scatter plots. If the scatter plot reveals a curvilinear relationship, a quadratic term or other transformation should be considered (Montopoli, 2007).

Absence of Multicollinearity and Singularity

SEM inverts matrices. If variables are very highly correlated or "perfect linear combinations of one another" (Ullman, 2007, p. 683) the procedure will "not converge to a solution" (Montopoli, 2007). The offending variable must be deleted or a composite variable must be created before analysis can occur (Montopoli, 2007).

Residuals

According to Ullman (p. 684): "… residuals should be small and centered around zero. The frequency of distribution of the residual covariances should be symmetrical. … When large residuals are found it is often helpful to examine the Lagrange Multiplier test … and consider adding paths to the model."

Significance of the Study

We live in a time where change in education is a matter of policy (Federal Register, 2006, Vol. 71, No. 77, 20519-20521) as well as practice. By modeling two of the three constructs of the Concerns-Based Adoption Model (CBAM, Hall & Hord, 2006) along with student achievement data for a specific instructional innovation in a multivariate environment, additional theory of change, professional learning, and the measurement of change with respect to impact on student achievement may emerge. According to Drs. Hall, Hord, and George, they know of no studies in which elements of the CBAM were modeled in a multivariate environment using a structural equation model.

Teaching and learning, particularly of mathematics, is a national priority. Additional insights to guide policy and practice relative to teaching and learning could contribute to heightened attainment of this national priority. By modeling complex relationships associated with systematic change in teaching and learning, additional insight into the dynamics of fidelity of implementation and levels of use of this specific innovation on student achievement may be achieved. This may provide a basis for additional theory that could be tested through future research.

Summary

Teaching and learning mathematics has become a matter of such urgency that it is a national policy. In order to comply, schools need to implement research-based practices, remaining mindful to verify results of implementation through evaluation. An exemplary evaluation of teaching and learning mathematics was conducted in Department of Defense Dependents' Schools in District A for the purpose of research verification. Initial findings highlighted the impact of fidelity in implementation of a specific approach to teaching mathematics and its relationship to increased student learning and achievement. In this study, the researcher will reanalyze these data in hopes of gaining additional insights into dynamics related to teaching, learning, and student achievement.

In chapter two, the researcher will provide an overview of relevant literature. Following a brief introduction, a review of the basis of change theory relevant to this study, and especially the Concerns-Based Adoption Model (Hall & Hord, 2006), will be provided. Next, a brief summary of literature related to student outcomes, standards, and assessments will provide the background to understand the approach to teaching mathematics that was evaluated in the original study.

The balance of the study will be presented in chapters three through five. Research methodology will be specified in chapter three, by restating the problem, summarizing research design and procedures, then providing in-depth descriptions of research methodology, population and sample, sources of information, and data analysis procedures. Findings and results will be presented in chapter four; while in chapter five, the researcher will provide summary, conclusions, recommendations for practice and future research, and implications.

CHAPTER TWO: REVIEW OF LITERATURE

Systemic change in teaching and learning is a topic of great interest today. Hall and Hord (2006) conceptualized change as a process rather than an event, pointing out that implementation does not occur until individuals change. In relationship to the basis for the present study, Hord (2000) concluded that "change is learning." According to Hord (Hall & Hord, 2006, p. 31; Hord & Sommers, 2008, pp. 21-22), research indicates that in order for systematic change to occur, learning would ideally take place: in a context conducive to change, with a shared vision of the desired change clearly articulated and widely communicated, where resources are planned and provided, investment is made in professional learning, progress is checked, and assistance is continuously provided.

Hall and Hord (2006) identified three aspects of the change process that can be measured to monitor implementers' progress and needs in order to provide ongoing support and to evaluate the extent of implementation of an innovation. Measurements for these aspects were specified through the Concerns-Based Adoption Model (CBAM) (Hall & Hord, 2006): Stages of Concern (SoC), Levels of Use (LoU), and Innovation Configurations (IC).

Development of the Concerns-Based Adoption Model

The CBAM (Hall & Hord, 2006) built upon and extended several strands of research: Frances Fuller's concerns theory (1969), patterns of implementations of change (Rogers, 1971, 1995), and the need to precisely define what an innovation is and what it is not (Heck et al., 1981; Hall & Hord, 2006).

In brief, Fuller discovered four levels of concerns among student teachers that changed with their levels of experience: unrelated, self, task, and impact concerns. *Unrelated concerns* were prevalent among education students without experience with children; concerns were not related to teacher education, but to unrelated life issues (i.e., social events). *Self concerns*—concerns related to self rather than teaching or its effect on students—were apparent in beginning student teachers. *Task concerns*— concerns related to the act of teaching— were more common among intermediate student teachers. Finally, *impact concerns*—concerns related to the impact on students and how to improve teaching for greater student learning—were observed among some experienced student teachers and teachers (Hall & Hord, 2001, pp. 58-59).

Fuller then compared the sequence of education students' concerns to education curricula, and determined that curricula mirrored university faculty's, rather than student teachers', concerns. In contrast, Fuller proposed "personalized teacher education" that would better address "personalogical development" of education students (Fuller, 1970; Fuller & Brown, 1975).

Rogers (1971, 1995) developed a typology related to "diffusion of innovations" that described individuals' readiness to accept change. These patterns have been observed in various settings, and can be understood as a predictable response to change by a population. Rogers and Scott (1997) described five different types of innovators that represent various percentages of the population: *Innovators* (2.5% of population) tolerate risk, are eager to try new approaches, and cope with uncertainty related to trying new ways. They bring new ideas into a system. *Early adopters* (13.5% of population), serve as a role model for peers. They reduce uncertainty about a new idea by trying it and providing an evaluation to their peers. The *early majority* (34% of the population) deliberately adopts the innovation after early adopters. They provide an important link between early and late adopters. After risk has been removed and benefits are clear, the *late majority* (34% of population) then adopts the innovation. The final 16% of the population were typified by Rogers as *laggards,* those who are skeptical of and resistant to change, and who continue to look to the past.

During a multi-year, collaborative research effort at The University of Texas at Austin Research and Development Center for Teacher Education, led by Hall, the following principles of change were identified (headings quoted from Hall & Hord, 2006, pp. 4-14):

1. Change is a process, not an event.
2. There are significant differences in what is entailed in development and implementation of an innovation.
3. An organization does not change until the individuals within it change.
4. Innovations come in different sizes.
5. Interventions are the actions and events that are key to the success of the change process.
6. There will be no change in outcomes until new practices are implemented.
7. Administrator leadership is essential to long-term change success.
8. Mandates can work.
9. The school is the primary unit for change.
10. Facilitating change is a team effort.
11. Appropriate interventions reduce resistance to change.

12. The context of the school influences the process of change.

Hall and Hord's (2006) research demonstrated that during the process of change, various factors influenced the likelihood of successful implementation, and intervention during critical junctures in the change process could increase the likelihood of successful long-term implementation. Furthermore, key variables could be measured to chart the progress through change implementation at individual and organizational levels. Those variables relate to quality of implementation, concerns of implementers, and implementation behaviors.

Innovation configurations. In the 1970s, two independent groups of researchers identified a similar phenomenon. Labeled "mutual adaptation" by Greenwood, Mann, and McLaughlin (1975), substantive differences were noted in the way teachers implemented programs. Hall and Loucks observed "innovation adaptation" and "innovation configurations" (1981). They found that new practices could be described and analyzed as components; and that implementation occurred in various patterns among components. The *innovation configuration* (IC) map (Hall & Hord, 2006) was created to provide a precise view of ideal implementation, as well as varying degrees of implementation that may approximate but fall short of the ideal. If developed collaboratively, those implementing the change can clearly understand what this change is, and what it is not. An IC map provides a shared vocabulary and point of reference throughout the change process that can be used to chart proximity to ideal implementation (quality), and as an anchor to define ideal implementation for the other two tools.

Specifics of IC mapping. The following is quoted from "Steps in Developing an IC Matrix" (Hord, n.d.) detailing the process of IC map development:

1. Visualize and brainstorm parts of the new practice or change in terms of what the user would be doing.
2. Identify components that constitute the major pieces of the new practices—by referring to and organizing the brainstorm list, adding to it, combining, or deleting.
3. Actionalize the components by stating them in behaviors or actions/use verbs—what are the users' behaviors, what are they doing?
4. Consider the sequence of the components and reorder them to make the best sense.
5. Generate variations for each component from ideal to unacceptable—state variations in action terms also.
6. Review, refine, edit the entire document for clarity.
7. Draw lines to indicate ideal, acceptable, and unacceptable variations.

Innovation configurations related to this study. Alquist and Hendrickson (1999) developed an innovation configuration (IC) map "as a way to diagnostically assess similarities and differences in how teachers were using the mathematics program" (p. 18) in District A. The mathematics program referenced is District A's adoption of the Department of Defense Dependents' Schools (DoDDS) enriched mathematics curriculum, *DoDDS Standards and Expectancies* (1994) based on criteria from *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics [NCTM], 1989). Excerpts from the IC map used in this study are available in Appendix A.

According to Alquist and Hendrickson (1999), the NCTM standards (1989), and adapted DoDDS standards and expectancies (1994), represented a significant change from the way most teachers had learned to provide math instruction. This new approach was based on constructivist philosophy (Vygotsky, 1962) and without a textbook (elementary level) or traditional computational exercises (middle school level).

> As such, both [elementary and middle school] programs de-emphasize computation taught out of context and the rote memorization of rules and procedures that have been the mainstay of traditional mathematics instruction. Both programs emphasize using mathematics in the context of long term projects and investigations. Students are expected to work cooperatively, use a variety of techniques to solve problems, and justify their thinking, both orally and in writing. Teachers are expected to use manipulative materials such as base 10 blocks extensively at all grade levels (Alquist & Hendrickson, 1999, pp. 18-19).

Math instruction involved thirteen components (George, Hall, and Uchiyama (2000, pp. 20-22):

1. Teacher poses mathematical tasks/investigation
2. Teacher facilitation of student activity
3. Teacher use of direct instruction
4. Teacher helps students in making connections
5. Teacher achieves closure
6. Students engaged in mathematical tasks throughout the lesson
7. Students' understanding of problem solving strategies

8. Teacher uses questions or comments to promote understanding of mathematics
9. Teacher probing for a variety of solution strategies
10. Teacher establishing and maintaining procedures governing materials and student behavior
11. Teacher structuring of opportunities for student responses
12. Student communication using mathematical language
13. Classroom visual displays

Then the IC map was used during observations to assess the extent and quality of implementation of the standards-based curricula in classrooms across the district. Sample data were presented by Alquist and Hendrickson (1999), with reference to a related paper by George, Hall, and Uchiyama (2000), in which a relationship between higher fidelity of implementation and higher student achievement was identified. These findings will be discussed later in this chapter, and data collected will be reanalyzed in the present study.

Stages of concern. As The University of Texas at Austin Research and Development Center for Teacher Education team, led by Hall, built upon Fuller's research, predictable patterns of concerns that people experience during change were identified. A tool was developed to measure and chart the concerns felt by individuals. These results could also be aggregated across the institution. Later research found that providing support appropriate to a concern affected the progress of change implementation.

Stages of Concern is the tool used to measure and chart concerns that are gathered from (1) a brief interview, (2) solicited, open-ended statements, or (3) a 35-item questionnaire. Hall & Hord pictured the progression of concerns from 0-6, as in Figure 1 (2006, p. 139; also in Hord, Rutherford, Huling, & Hall, 2006, p. 31).

Figure 1. Stages of Concern: Typical Expressions of Concern about the Innovation

	Stages of Concern	**Expressions of Concern**
I	6 Refocusing	I have some ideas about something that would work even
M		better.
P	5 Collaboration	I am concerned about relating what I am doing with what
A		other instructors are doing.
C	4 Consequence	How is my use affecting kids?
T		
T		
A	3 Management	I seem to be spending all my time getting material ready.
S		
K		
S	2 Personal	How will using it affect me?
E	1 Informational	I would like to know more about it.
L	0 Awareness	I am not concerned about it (the innovation).
F		

Note. From *Taking Charge of Change* (p. 31), by S. M. Hord, W. L. Rutherford, L. Huling, and G. E. Hall, 2006, Austin, TX: Southwest Educational Development Laboratory. Copyright 2006 by SEDL. Reprinted with permission.

With respect to the present study, stages of concern data were not available.

Levels of use. The last of the three tools measures implementation behaviors. *Levels of Use* (LoU) range from 0-Nonuse to VI-Renewal, as shown in Figure 2.

Figure 2. Levels of Use of the Innovation

Levels of Use of the Innovation

Level 0—Non-use

State in which the individual has little or no knowledge of the innovation, no involvement with it, and is doing nothing toward becoming involved.

DECISION POINT A—Takes action to learn more detailed information about the innovation.

Level I—Orientation

State in which the individual has acquired or is acquiring information about the innovation and/or has explored its value orientation and what it will require.

DECISION POINT B—Makes a decision to use the innovation by establishing a time to begin.

Level II—Preparation

State in which the user is preparing for first use of the innovation.

DECISION POINT C—Begins first use of the innovation.

Level III—Mechanical use

State in which the user focuses most effort on the short-term, day-to-day use of the innovation with little time for reflection. Changes in use are made more to meet user needs than needs of students and others. The user is primarily engaged in an attempt to master tasks required to use the innovation. These attempts often result in disjointed and superficial use.

DECISION POINT D-1—A routine pattern of use is established

Level IVA—Routine

Use of the innovation is stabilized. Few if any changes are being made in ongoing use. Little preparation or thought is being given to improve innovation use or its consequences.

DECISION POINT D-2—Changes use of the innovation based on formal or informal evaluation in order to increase client outcomes.

Level IVB—Refinement

State in which the user varies the use of the innovation to increase the impact on clients (students or others) within their immediate sphere of influence. Variations in use are based on knowledge of both short and long-term consequences for clients.

DECISION POINT E—Initiates changes in use of the innovation based on input from and in coordination with colleagues for benefit of clients.

Level V—Integration

State in which the user is combining own efforts to use the innovation with related activities of colleagues to achieve a collective impact on clients within their common sphere of influence.

DECISION POINT F—Begins exploring alternatives to or major modifications of the innovation presently in use.

Level VI—Renewal

State in which the user reevaluates the quality of use of the innovation, seeks major modifications of, or alternatives to, present innovation to achieve increased impact on clients, examines new developments in the field, and explores new goals for self and the organization.

Note. From *Taking Charge of Change* (p. 55), by S. M. Hord, W. L. Rutherford, L. Huling, and G. E. Hall, 2006, Austin, TX: Southwest Educational Development Laboratory. Copyright 2006 by SEDL. Reprinted with permission.

Building on Rogers' research, Hall and Loucks (1975) defined a predictable process that individuals follow while implementing an innovation. The LoU construct was specified based on these patterns, in order to document what a person was doing in their process of change. As identified in Figure 2, individuals can be divided between nonusers (levels 0-II) and users (levels III-VI). It must be understood that this measurement scale is generic; one should not infer that an individual will necessarily progress in implementation behavior to any specific level. Rather, the instrument is intended to describe what an individual is doing (or not doing) with respect to an innovation. These data can be helpful to a change facilitator (i.e., staff developer, coach, or mentor) interested in supporting individuals to implement a specific innovation because different kinds of support are required for an individual at level 0 than for one at level III. These types of facilitating behaviors have been defined and verified through research (Hall & Hord, 2006).

Data regarding LoU are gathered through either of two interview procedures, a branching interview or focused interview. Detailed procedures for both types of interviews are specified in *Measuring Levels of Use of the Innovation: A Manual for*

Trainers, Interviewers and Raters (Loucks, Newlove & Hall, 1975). A branching interview is a less formal approach to assess an individual's level of use, based on answers to key questions during a brief interview (depicted on p. 168, Hall & Hord, 2006). The focused interview is a more in-depth procedure, using an interview protocol. This formal approach is preferable for research and evaluation (Hord & Sommers, 2008, p. 121). In order for validity to be assured, training to research criterion and certification in LoU interview procedures and rating is required (Thornton & West, 1999; Hall & Hord, 2006, p. 167; Hord & Sommers, 2008, p. 121).

LoU of mathematics curricula in District A. Thornton and West (1999) published findings from the evaluation conducted from 1995 through 1997 in District A using data collected from LoU interviews (Hall & Hord, 2006) with teachers. Data were collected by nine of the superintendent's staff who had been trained and certified by experts to research criterion. Data were collected in all 15 schools yielding 102 interviews, grades K-8 in 1995-96 and 106, grades K-8 in 1996-97. These interviews were taped and ratings were reviewed by an independent researcher. Overall inter-rater reliability alpha 0.78 was attained in 1995-96. An inter-rater reliability alpha of .86 was attained by reducing the raters to the six with the highest levels of reliability, and the other three were dropped from the interview schedule for 1996-97.

Significant findings explained by Thornton and West (1999) were:

- All teachers implemented the mathematics curricula in 1996-97, compared to 99% in 1995-96.
- At the end of 1995-96, 59% were at LoU III (Hall & Loucks, 1977). Among these, 63% moved to LoU IVA (50%) or LoU IVB (13%), respectively.

- Most (54%) teachers were at LoU IVA in 1996-97, compared to only 25% in 1995-96. No one regressed from higher levels back to LoU III; 30% of LoU IVA teachers in 1995-96 remained at LoU IVA in 1996-97.
- About a third (32%) was at LoU III in 1996-97, compared to 59% in 1995-96. Almost a quarter (22%) of teachers at LoU III in 1995-96 remained at LoU III in 1996-97.
- Teachers at LoU IVB remained nearly constant (10% 96-97, 12% 95-96).

Concluding statement regarding the CBAM. The three constructs of the CBAM (Hall & Hord, 2006) provide powerful insights. A change facilitator can provide support with specificity, and increase the tenor and success of implementation. Hord and Sommers (2008, p. 118) recommend using the CBAM (Hall & Hord, 2006) in the way that District A did. Based on the premise that: (1) “change is learning” (Hord, 2000), (2) long-term, targeted, professional learning is required in order for change of educational practice to occur, and (3) only then can benefits to student learning occur (Hord & Sommers, 2008, pp. 19-20; Joyce & Showers, 2002); the CBAM may be viewed as a set of concepts about change, each of which has a measure to infer the impact of professional learning and transfer of training.

Standards-Based Education

A Nation at Risk: The Imperative for Educational Reform (National Center for Excellence in Education, 1983) is noted by many to mark the beginning of the standards movement in American education (Spring, 2005a, p. 2; Hall & Hord, 2001, p. 23). This translated into educational policy under Presidents G. H. W. Bush, W. J. Clinton, and G. W. Bush. President G. H. W. Bush’s Goals 2000 initiative included national standards, voluntary national achievement tests, and incentives for parental choice, which President Clinton signed into law as the Goals 2000 Education Act (Spring, 2005b). Standards and assessments reached a much higher level of required implementation when G. W. Bush signed the *No Child Left Behind Act of 2001* (P. L. 107-110, 1-08-02) into law, coupled with a goal of 100% proficiency among students based on standardized tests, by 2013-2014 and penalties for schools not performing at increasingly higher levels of proficiency to meet this goal.

Educational standards. In brief, educational standards indicate what students are supposed to know and be able to do at various stages in their education. States are given responsibility to test students' abilities to demonstrate evidence of the extent that educational standards have been attained by students in various classrooms, schools and districts.

Professional organizations, such as the National Council of Teachers of Mathematics (NCTM), sought to promote exemplary content standards that could assist states, schools, and individual practitioners in defining curricular standards in their content specialization. NCTM published *Curriculum and Evaluation Standards for School Mathematics* (1989) and later republished these as *Principles and Standards for School Mathematics* (2001). According to NCTM, their purpose was to promote "precepts that are fundamental to a high-quality mathematics education" based on evidence collected through research, and to describe what mathematics instruction should equip students to know and do. NCTM sought to instill a shared vision "to guide educators as they strive for the continual improvement of mathematics education" (NCTM, 2001).

The *Principles and Standards for School Mathematics* (2001) emphasize: high expectations and strong support for all students, coherent and well articulated curricula, effective instruction that involves challenge and support, constructing understanding through experience and previous knowledge, diagnostic and formative assessment, and use of technology-supported learning. NCTM standards (2001) are divided into content and process standards. Content standards describe five strands of mathematics content that students must learn: algebra, geometry, measurement, data analysis, and probability. Process standards articulate ways to gain and apply content knowledge. These include problem solving, reasoning and proof, communication, connections, and representations.

DoDDS adoption of constructivist approaches to teaching mathematics. According to George, Hall and Uchiyama (2000), the DoDDS adapted 1989 NCTM Standards into its own Mathematics Standards and Expectancies (1994), which provided weekly and monthly expectations by grade level. As with the NCTM standards, the DoDDS approach was based on constructivist philosophy (Vygotsky, 1962), as described earlier on page 15 and Appendix A.

Measurement of learning. District A chose a performance-based assessment aligned with the NCTM standards to measure student learning, because it closely matched the way in which students were learning mathematics in the classroom and the content covered in grades 2-8, and therefore would be most likely to yield a reliable measure of content knowledge. *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 179), defined performance assessments as:

Product- and behavior-based measurements based on settings designed to emulate real-life contexts or conditions in which specific knowledge or skills are actually applied.

Fourteen of the schools in District A agreed to implement pre- and post-assessments using the GOALS performance-based assessment, Forms A and B (The Psychological Corporation). The Psychological Corporation trained teachers in the fourteen schools to administer and score the assessments (George, Hall & Uchiyama, 2000). Form A was given as a pre-test in November 1996, and Form B as post-test in April, 1997. Fall pre-tests and spring post-tests were administered annually throughout the period of the evaluation study, 1996-1998.

GOALS assessments consisted of ten items given in two separate one-hour administrations relevant to each grade level, two through eight. The first five items were given in the first sitting, and the second five in a later one-hour administration. Items were scored on a rubric where *one* indicated minimal understanding of the concept; *two*, partial understanding; *three*, full understanding; *eight*, did not answer; and *nine*, off topic or illegible. Scores of *eight* or *nine* were counted as zero in the tabulation of results. Possible scores therefore ranged from 0-30, with 0 representing the lowest possible score and 30 representing the highest possible score.

Change and Student Achievement

The primary purpose for systematic change or reform in teaching is to increase student learning and achievement. In a meta-analysis of research on teaching and student learning, Marzano, Pickering, and Pollock (2001) identified that certain teaching behaviors had greater impact on student learning and achievement than others. While an extensive body of literature has been generated along these lines (Marzano, 2007; Marzano, 2003; Marzano, Marzano & Pickering, 2003; to name a few), practitioners find varying results from attempts at classroom implementation and striking differences between teachers' self-perceived implementation and observed fidelity of implementation (Hall & Loucks, 1981; Alquist & Hendrickson, 1999; Cohen, 1990; Kilpatrick, Hancock, Mewborn, & Stallings, 1996; Lipsey & Cordray, 2000). Therefore, it is critical not only to identify "what works" but to measure the **extent of implementation** in each classroom in order to evaluate results of implementation (Hall, 1999). This will avoid "appraising nonevents in program evaluation" (Charters & Jones, 1973).

Importance of fidelity in implementation. Hall and Hord (2006) identified early on that innovation variation can produce very different results. Therefore, it is not just implementation of a practice called by a certain name, but **implementation with fidelity** of a specific practice or set of practices, that can produce reliable results. This was exemplified in George, Hall, and Uchiyama's *Extent of Implementation of a Standards-based Approach to Teaching Mathematics and Student Outcomes* (2000), where implementation with high levels of fidelity of a set of instructional practices resulted in higher levels of student achievement and implementation with low levels of fidelity resulted in lower levels of student achievement.

George, Hall and Uchiyama's findings in District A, DoDDS. Two of the three constructs from the CBAM (Hall & Hord, 2006) were used to measure extent of implementation behavior (LoU focused interviews) and fidelity of implementation (IC ratings, *a* through *d* or *e*, where *a* represented highest levels of fidelity with the innovation configuration map). These were evaluated, together with GOALS scores (classroom means), and teachers' attendance (or non-attendance) in a class designed to train teachers in this new curricular/instructional approach, to see whether relationships existed among teachers' LoU and IC map ratings, teacher training, and student achievement. Researchers matched 2,179 grades two-eight students who took fall and spring assessments, and who remained in the same teacher's classroom from fall 1996 to spring 1997. This represented 107 classrooms in 14 schools across District A.

A series of analyses were conducted, using a combination of methods including "linear model techniques described in Neter et al. (1996)" in SAS/STAT (SAS Institute, 1992), using procedures TABULATE, GLM, REG, and ANCOVA (Analysis of Covariance) (George, Hall, & Uchiyama, 2000, p. 11). GOALS fall (pre-test) scores were compared by grade level to establish equivalence of difficulty across grade levels, in order to attribute differences in students' scores to teacher characteristics. No significant differences existed among grade level mean pre-test scores. The fall class mean was then used as covariate to compare predicted spring scores to those attained, taking into consideration other variables including LoU, IC ratings, and attendance at training.

Attendance at training significantly increased these teachers' predicted spring class mean scores by about 1.67 points compared to teachers who did not attend training (t=2.73, df=106, p<.05) (George, Hall, & Uchiyama, 2000, p. 12). Interaction among training and LoU ratings were explored with respect to impact on spring GOALS scores using the GLM procedure. The purpose was to isolate the effects of the various independent variables on the dependent variable. Spring classroom GOALS averages were designated dependent variable, while Training attendance, LoU, interaction of training and LoU, and fall GOALS averages were independent variables. As pictured in

Table 1, the only statistically significant effect related to LoU, attendance in training, was non-significant (F=0.77, df=1,57, p=.38), LoU was significant (F=81.70, df=3,55, p<.01), and the interaction was non-significant (F=35.25, df=2,56, p=.07). Researchers concluded from these results that participation in training did not necessarily impact student achievement; rather implementation of instructional strategies in the classroom acquired through training (measured by LoU) was associated with improved student achievement (George, Hall, & Uchiyama, 2000, p. 12).

Table 1

Attendance in Standards-Based Training and Associated Classroom GOALS Scores

Special Training for Standards Based Instruction	Level of Use	Number of Teachers	Number of Students	Mean of Classroom Means: Fall	Mean of Classroom Means: Spring	Average Change in Classroom Means: Fall to Spring
No	Not Rated	42	891	12.5	17.3	4.8
	III	13	177	11.1	16.1	5.0
	IVA	12	302	14.4	18.7	4.3
	IVB	3	104	11.6	18.7	7.0
	V	0	0			
	All	70	1474	12.5	17.4	4.8
Yes	Not Rated	6	172	12.5	17.8	5.2
	III	8	148	13.9	17.5	3.6
	IVA	14	219	13.2	21.0	7.8
	IVB	7	130	14.2	19.4	5.2
	V	2	36	12.0	21.8	9.8
	All	37	705	13.4	19.5	6.1
TOTAL	All	107	2179	12.8	18.1	5.3

Note. From “Extent of Implementation of a Standards-Based Approach to Teaching Mathematics and Student Outcomes,” by A. A. George, G. E. Hall, and K. Uchiyama, 2000, *Journal of Classroom Interaction*, *35*, p. 12. Copyright 2000 by University of Houston. Reprinted with permission.

The relationship between training and LoU was evaluated using a chi-square test. LoU IVB and V were combined due to small counts. While results were not statistically significant (chi-square=4.2, *df*=2, *p*=.12), researchers noted a pattern of higher LoU among those who attended training (George, Hall, & Uchiyama, 2000, p. 13).

LoU ratings were then evaluated in concert with fall and spring GOALS class means, and average differences between fall and spring. Researchers noted the highest increase in LoU V classrooms, followed by IVA and IVB classrooms, respectively.

> Post-hoc comparisons indicated students taught by LoU III Mechanical teachers had significantly lower predicted spring GOALS test scores than those in LoU IVA Routine (*t*=3.06, *p*<.01) and LoU V Integration (*t*=2.67, *p*=.01). LoU IVB Refinement teachers' spring expected values were also higher than LoU III Mechanical teachers', but the difference was not quite statistically significant (*t*=1.80, *p*=.08). There were no significant differences between expected values in the spring for classrooms taught by teachers at LoU IVA, LoU IVB, and LoU V. Contrasting LoU III teachers' scores with all those at higher Levels of Use indicates a significant difference (*F*=32.80, *df*=3,54, *p*<.01). These results indicate a positive relationship between teachers' use of the math program and the students' test scores. Higher Levels of Use are associated with greater student learning (George, Hall, & Uchiyama, 2000, p. 14).

Fidelity in implementation was explored in two stages, in 1996-97 and then 1997-98. Scores for 1996-97 were grouped into *high*, *medium*, and *low* fidelity groups based on

IC map ratings, and classroom averages were compared for fall, spring, and average change. Results from analyses using the GLM procedure are displayed in Table 2. Spring average classroom score was dependent variable, while teachers' fidelity group from IC maps and fall average classroom scores were independent variables. Relationships fell "short of statistical significance (F=3.43, df=3,8, p=.07)," except that "the ANCOVA model indicated a statistically significant relationship between these teachers' IC Map ratings and the expected values of the spring GOALS scores (F=4.43, df=2,8, p=.05)," (George, Hall, & Uchiyama, 2000, p. 14). However, researchers concluded that samples were too small to infer effects of fidelity on student achievement, also noting the higher LoU in medium and high fidelity groups (IVA or above, compared to III in the low fidelity group).

Table 2

Teachers' Innovation Configuration Map Ratings and Classroom GOALS Scores, 1996-97

Innovation Configuration Cluster	Number of Classrooms	Number of Students	Mean of Classroom Means: Fall	Mean of Classroom Means: Spring	Average Change in Classroom Means: Fall to Spring
Low	5	85	13.3	17.4	4.1
Medium	3	33	11.2	22.2	11.0
High	4	64	14.3	19.6	5.3
All	12	182	13.1	19.3	6.2

Note. From "Extent of Implementation of a Standards-Based Approach to Teaching Mathematics and Student Outcomes," by A. A. George, G. E. Hall, and K. Uchiyama, 2000, *Journal of Classroom Interaction*, *35*, p. 14. Copyright 2000 by University of Houston. Reprinted with permission.

Gender and ethnicity were the student characteristics explored to see whether they would have any relationship to student achievement. In brief, no significant differences could be attributed to gender, but African American students' scores were significantly lower than Caucasian and Other students' scores by about 1.7 points (t=3.23, df=301, p<.001) (George, Hall, & Uchiyama, 2000, p. 15). Similar results were found in 1997-98 (George, et al., p. 23).

Further data examining a possible relationship between fidelity and student achievement were gathered in the 1997-98 year. LoU interviews were not conducted during this phase of the evaluation. Results associated higher levels of fidelity (IC Map ratings) with higher levels of student achievement, while lower levels of fidelity were associated with lower levels of student achievement. IC ratings were available from 30 teachers in 12 schools, teaching 1,026 grades 2-8 students, while GOALS tests were available for 2,301 students who were in the same classroom in fall and spring. There were significant differences among classroom average scores, and in changes from fall to spring. As in 1996-97, differences existed between grade levels that were not statistically significant, so changes in student achievement were attributed to teacher characteristics.

Cluster analysis was used to specify and evaluate levels of fidelity on the 13 components of the IC map (Appendix A; Alquist & Hendrickson, 1999), and then to examine possible relationships among these levels and student achievement. Each component has four or five possible levels of fidelity, with *a* representing high fidelity, and *d* and *e* represent low fidelity. Original researchers converted these ratings to numbers; *a* was coded as 4, *b* as 3, *c* as 2, *d* as 1, and *e* as 0; and then performed “a type of oblique component analysis related to multiple group factor analysis” (SAS/STAT, 1992 referenced in George, Hall, & Uchiyama, 2000, p. 17). The procedure resulted in three fidelity groups, *high*, *intermediate*, and *low*, comprised of seventeen high fidelity, nine intermediate, and four low fidelity teaching variations. This was not a random sample, but the team felt these teachers were similar to fidelity patterns throughout the district, though statistical generalization of results is not possible (George, Hall, & Uchiyama, 2000, p. 17). Researchers noted that while patterns were similar to 1996-97, “small sample size and large variances between classrooms resulted in non-significant statistical tests” (George, Hall, & Uchiyama, p. 18).

IC map ratings and GOALS results from 1996-97 and 1997-98 were then combined, to increase power of statistical tests. Results, pictured in Table 4, indicate higher levels of student achievement with higher levels of teacher fidelity in implementation, and lower levels of student achievement gains for teachers with lower levels of fidelity in implementation. George, Hall, and Uchiyama noted “the implication is that the teachers’ fidelity of implementation has a significant effect in classrooms with low fall achievement scores” (2000, p. 18).

Table 3

Teachers' Innovation Configuration Map Ratings and Classroom GOALS Scores, 1996-97 and 1997-98 Combined

Innovation Configuration Cluster	Number of Classrooms	Number of Students	Mean of Classroom Means: Fall	Mean of Classroom Means: Spring	Average Change in Classroom Means: Fall to Spring
High	22	580	14.7	20.1	5.4
Medium	12	411	13.5	18.5	5.0
Low	8	217	13.6	17.6	4.0
All	42	1208	14.2	19.2	5.0

Note. From "Extent of Implementation of a Standards-Based Approach to Teaching Mathematics and Student Outcomes," by A. A. George, G. E. Hall, and K. Uchiyama, 2000, *Journal of Classroom Interaction*, *35*, p. 18. Copyright 2000 by University of Houston. Reprinted with permission.

Researchers then analyzed ratings on the thirteen individual components and students' scores, "to determine whether student achievement was related to some of the components more than others." Each individual component was analyzed using ANCOVA. In addition to fidelity ratings of *a*-*e*, raters used two additional values: "n" for "not doing" and blank, "-", in Table 4. George, Hall, and Uchiyama noted, "There are only two regression weights for each model, the first for the '*a*' teachers, the second for the '*b*' teachers, and all other teachers forming the comparison group" (2000, p. 19). Components 1, 2, 3, 5, and 6 indicated significant differences in student achievement associated with fidelity ratings.

Table 4

Student Achievement as a Function of IC Map Ratings

Component Ratings	Number of Classrooms	Number of Students	Mean of Fall Classroom Means	Mean of Spring Classroom Means	Change from Fall to Spring	Group for Regression (ANCOVA) Analysis	Regression Coefficient	Significance
All	**42**	**1208**	**142**	**192**	**50**			
1) Teacher Poses Mathematical Tasks/Investigation								
A	11	363	13.6	20.9	7.2	High	3.09	0.05
B	21	473	14.6	18.9	4.2	Medium	0.55	0.68
C	4	121	14.9	19.0	4.1	Low		
D	4	13.5	12.8	16.6	3.9	Low		
E	1	11	15.5	19.3	3.7	Low		
-	1	105	12.5	18.3	5.9	Omitted		
2) Teacher Facilitation of Student Activity								
A	22	577	14.7	20.5	5.8	High	2.68	0.03
B	6	174	13.4	18.7	5.3	Medium	1.46	0.37
C	8	280	13.3	16.6	3.2	Low		
D	6	177	14.2	18.4	4.2	Low		
3) Teacher Use of Direct Instruction								
A	10	267	12.8	20.5	7.7	High	4.17	0.00
B	19	493	15.0	20.0	5.0	Medium	1.46	0.37
C	11	393	13.7	17.3	3.6	Low		
N	2	55	15.4	15.3	-0.1	Low		
4) Teacher Helps Students in Making Connections								
A	12	316	14.6	20.6	6.0	High	1.92	0.14
B	1	15	15.5	19.5	4.0	Medium	-0.12	0.93
C	11	379	13.1	17.8	4.7	Medium		
D	8	241	15.0	19.0	4.0	Low		
E	5	83	12.8	17.5	4.6	Low		
N	3	146	15.0	19.0	4.1	Low		

-	2	28	15.8	23.2	7.5	Omitted		
5) Teacher Achieves Closure								
A	11	195	13.3	20.5	7.3	High	3.08	0.02
B	4	56	13.5	19.3	5.8	Medium	1.80	0.18
C	5	236	15.2	20.4	5.2	Medium		
D	4	116	13.0	17.6	4.6	Low		
E	10	270	15.5	19.9	4.3	Low		
N	7	235	13.7	16.0	2.2	Low		
-	1	100	15.8	19.2	3.4	Omitted		
6) Students Engaged in Mathematical Tasks throughout the Lesson								
A	19	466	14.5	20.3	5.8	High	2.85	0.03
B	10	311	14.4	19.2	4.8	Medium	1.86	0.21
C	8	263	13.8	17.1	3.4	Low		
D	2	40	11.7	15.8	4.1	Low		
N	1	16	16.0	18.4	2.4	Low		
-	2	112	13.7	20.7	7.0	Omitted		
7) Students' Understanding of Problem Solving Strategies								
A	6	126	16.3	21.3	5.0	High	1.72	0.30
B	10	261	13.5	19.2	5.8	Medium	1.86	0.21
C	13	481	14.4	17.4	3.0	Low		
D	6	102	15.2	19.1	3.9	Low		
E	2	55	10.9	16.9	5.9	Low		
N	3	68	11.5	22.5	11.0	Low		
-	2	115	14.0	21.3	7.3	Omitted		
8) Teacher Uses Questions or Comments to Promote Understanding of Mathematics								
A	9	363	15.6	21.1	5.5	High	2.20	0.14
B	12	198	13.3	18.7	5.4	Medium	0.84	0.52
C	11	385	13.5	18.5	5.0	Low		
D	6	177	14.8	18.0	3.3	Low		
E	1	37	14.7	15.0	0.3	Low		
N	2	30	14.8	19.4	4.6	Low		

9) Teacher Probing for a Variety of Solution Strategies								
A	19	509	14.3	20.2	5.9	High	1.53	0.20
B	8	218	14.2	17.8	3.6	Medium	-0.85	0.57
C	8	262	14.2	18.9	4.6	Low		
D	4	146	14.6	19.6	4.9	Low		
E	1	43	8.2	10.3	2.1	Low		
-	1	105	12.5	18.3	5.9	Omitted		
10) Teacher Establishing and Maintaining Procedures Governing Materials and Student Behavior								
A	25	807	14.7	19.7	5.0	High	2.00	0.35
B	11	227	13.9	19.5	5.5	Medium	2.11	0.31
C	4	144	13.0	17.0	3.9	Low		
-	2	30	11.5	15.5	4.0	Omitted		
11) Teacher Structuring of Opportunities for Student Responses								
A	14	326	14.0	20.3	6.3	High	2.40	0.11
B	19	659	14.9	19.2	4.4	Medium	0.92	0.52
C	8	211	12.8	16.6	3.8	Low		
D	1	12	13.7	23.4	9.8	Low		
12) Student Communication Using Mathematical Language								
A	2	51	16.2	19.2	3.0	High	2.16	0.15
B	13	424	14.5	20.1	5.6	High		
C	7	253	13.7	18.2	4.6	Medium	0.84	0.62
D	6	179	15.1	18.0	2.9	Low		
E	2	37	14.0	18.3	4.3	Low		
-	12	264	13.3	19.5	6.1	Omitted		
13) Classroom Visual Displays								
A	10	240	13.1	18.5	5.4	High	0.33	0.80
B	7	173	14.6	21.5	6.9	Medium	2.63	0.07
C	7	211	14.5	19.2	4.7	Low		
D	11	389	14.8	18.2	3.4	Low		
E	7	195	14.0	19.4	5.5	Low		

Note. From “Extent of Implementation of a Standards-Based Approach to Teaching Mathematics and Student Outcomes,” by A. A. George, G. E. Hall, and K. Uchiyama, 2000, *Journal of Classroom Interaction*, *35*, pp. 20-22. Copyright 2000 by University of Houston. Reprinted with permission.

George, Hall, and Uchiyama shared the following implications from implementation data (2000, p. 24):

1) Implementation of a major change in classroom practices truly is a multi-year process for most teachers. They are likely to be at a Mechanical Use level in their first year, and for many they remain at this level for more than one year.
2) As teachers moved to higher Levels of Use, higher levels of student learning were observed.
3) Student achievement was higher in classrooms where practices were more closely aligned with the NCTM Standards, i.e., in classrooms where there were more *a* and *b* variations on the IC Map ratings. The finding represents initial verification that, in at least one setting, the NCTM Standards led to higher levels of student learning.
4) Students in classrooms with low fall test scores seem to benefit the most from “high fidelity” implementation.

The authors closed by inviting others to review the study findings and to contribute to increased understanding of “how best to assist teachers in implementing new and better paradigms for classroom practices.”

Summary

In this chapter, Hall and Hord's research on change was chronicled, along with development of the three individual constructs of the Concerns-Based Adoption Model: Levels of Use, Stages of Concern, and Innovation Configurations (2006). This body of literature represents well-respected, foundational knowledge of change processes. In reference to the study being revisited, Hord further asserted that "change is learning" (2000). Therefore, Hord and Sommers (2008, pp. 19-20) recommend use of the CBAM (Hall & Hord, 2006) to evaluate professional learning processes with respect to transfer of learning (Joyce & Showers, 2002).

Educational leaders are critical in building a context conducive to change (Hall & Hord, 2006, p. 31; Hord & Sommers, 2008, pp. 21-22). As mentioned earlier, Hall & Hord (2006) stated: "Administrator leadership is essential to long-term change success. … Mandates can work. … The school is the primary unit for change. … Facilitating change is a team effort. … Appropriate interventions reduce resistance to change. … and, the context of the school influences the process of change." Given that all of these fall within the domain of educational leaders, changes in learning that occur from improved teaching quality could be perceived as a central goal of educational leadership.

Standards-based education was traced from 1983 to the present. Additional information on educational standards, and particularly the National Council of Teachers of Mathematics *Principles and Standards for School Mathematics* (2001), was provided. The Department of Defense Dependents' Schools adapted the NCTM standards into its Mathematics Standards and Expectancies (1994). These were implemented in District A of DoDDS.

When Hall and colleagues evaluated the extent and impact of District A's implementation of constructivist approaches to teaching mathematics 1996-98, the evaluation involved practical application of Levels of Use and Innovation Configuration Maps that were used to evaluate the fidelity and extent of implementation behaviors related to this approach to mathematics instruction, and relationship to student achievement documented through performance-based assessments. Findings from this evaluation published by George, Hall, and Uchiyama were reviewed in detail. Major implications related to:

- patterns in Levels of Use,
- relationship between higher Levels of Use and higher levels of student learning,
- higher achievement in classrooms most closely aligned with NCTM Standards (higher fidelity ratings on IC Maps), and
- apparent increased benefits from "high fidelity" implementation patterns by students with lower fall test scores.

The authors invited others to review study findings and further contribute to the body of knowledge contained in these well-specified data.

CHAPTER THREE: RESEARCH METHODOLOGY

In this chapter, the researcher will restate the problem and provide an overview of research design and procedures relevant to this study. Research methodology will be explained and depicted, to assist the reader in understanding how it corresponds to answering the research question. Population and sample will provide information about whom the data represent, and sources of information provide additional information about archival data being reanalyzed in this study. Finally, quantitative procedures will be recounted in data analysis procedures.

Restatement of the Problem

The purpose of this study is to:

1. model relationships among student outcomes, levels of implementation behavior, and fidelity of implementation of constructivist approaches to teaching mathematics, and
2. to compare results from structural equation modeling to results from the original analyses, in hopes of identifying potential similarities and differences in statistical methodologies for viewing the dynamics of change in teaching and learning mathematics.

Research Design and Procedures

In this correlational, causal-comparative study, the researcher will reanalyze data that were collected using a multimethod (Brewer & Hunter, 1989) approach. This extension of Hall's (1999, 2000), Hord's (2000), and George, Hall, and Uchiyama's (2000), research will combine analysis of interview data scored with respect to Levels of Use of the Innovation (Hall, et al., 1975), observations of classroom teaching scored by Innovation Configuration Maps (Hall & Hord, 2006), and students' pre- and post-scores on standardized examinations.

Latent variable structural equation models (Jöreskog & Sörbom, 1999) will be constructed based on original findings and used to calculate estimates of the effects of the model (Ullman, J. B., 2007; Pike & Kuh, 2005). Multigroup modeling will then be used to identify statistically significant differences, interactions, and relationships.

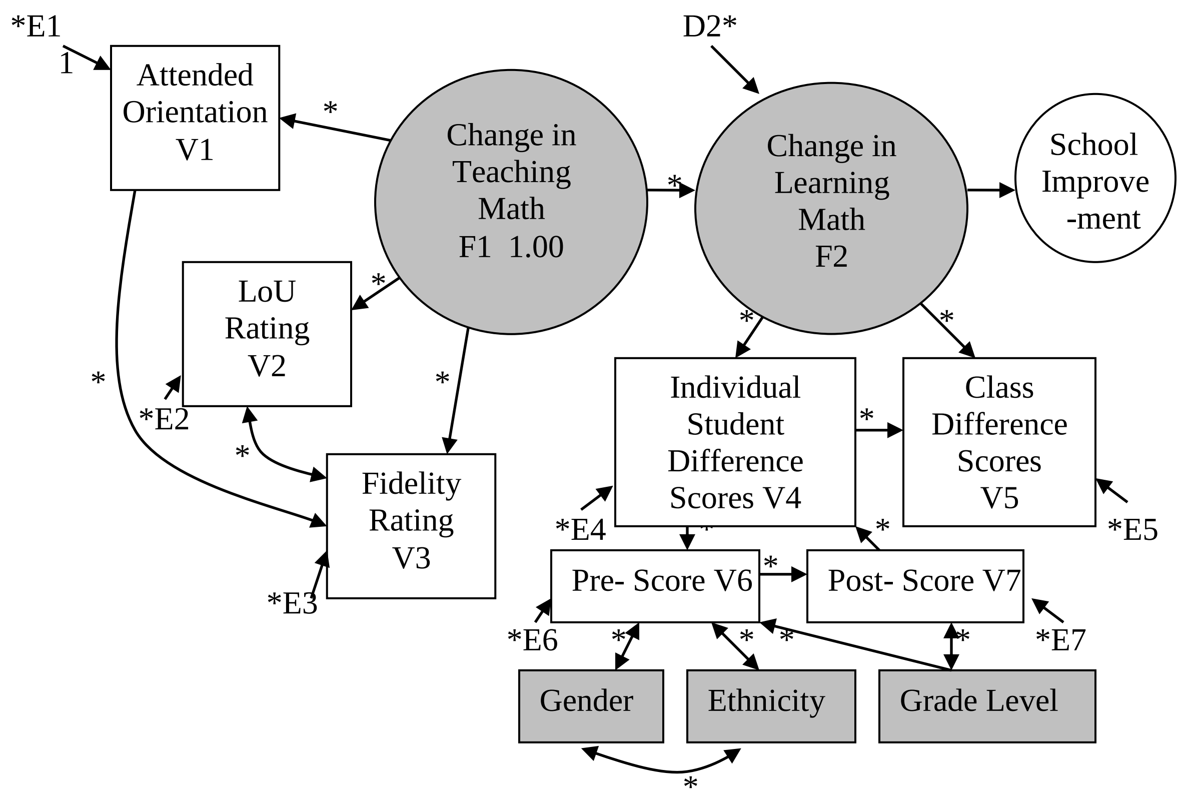
Research Methodology

The hypothesized model is illustrated in Figure 3, based on George, Hall and Uchiyama (2000), and Hord (2000). Latent variables are represented by circles and measured variables by squares. Independent variables are shaded; dependent variables are unshaded. This model contains two latent variables: Change in Teaching Math and Change in Learning Math. A third latent variable is pictured but not explored in this study, School Improvement.

Arrows indicate the direction of prediction. Change in Teaching Math predicts attendance at orientation to the math program, higher Levels of Use, and higher levels of fidelity on IC measurements. The researcher further postulates that an indirect effect may exist between LoU and IC ratings.

Together, the latent variable Change in Teaching Math with its measured components, will predict Change in Learning Math, a latent variable that predicts individual student and class differences on GOALS tests. Gender, ethnicity and grade level are intervening variables that may impact individual students' difference scores. The researcher postulates that grade level may also impact pre- and post-scores, and further, that an indirect effect may exist between gender and ethnicity. Individual difference scores are then aggregated into class differences, and therefore serve as predictors of the same. Together, Change in Teaching and Change in Learning Math are theorized to produce School Improvement.

Figure 3. Proposed Model for Change in Teaching and Learning Mathematics



Through structural equation modeling, relationships are translated into equations and the model is estimated as a series of equations in matrices. Results from population parameters are then compared and covariance matrices are produced. The goal is to produce a model that does not indicate significant differences between an estimated population covariance matrix and the sample covariance matrix.

Population and Sample

The findings of this study will relate to teachers and grades 2-8 students attending Department of Defense Dependents' Schools in District A, 1996-1998. A purposive sample was used involving teachers who were willing to participate in the evaluation (from 14 of the schools). Because it was a purposive sample, results cannot be generalized beyond study participants.

Sources of Information

Records were provided from a database that was recoded with research numbers by original researchers in order to preserve anonymity of original subjects. The database contains teacher information (i.e. grade taught, school, training attended), implementation behavior data (i.e. levels of use of the innovation [Hall, et. al, 1975], implementation fidelity based on IC Maps [Hall & Hord, 2006]), student characteristics (i.e. gender, ethnicity), and academic records (i.e. student pre- and post-test scores, teacher, grade level).

Data Analysis Procedures

In this study, the researcher will reanalyze George, Hall, and Uchiyama's (2000) research through the following procedures:

1. Latent variable structural equation models in *AMOS 7.0.0* (Arbuckle, 2006) will be used in concert with *SPSS 15.0 for Windows Graduate Student Version* (SPSS, Inc., 2006) to test the conceptual model for goodness of fit to the data using maximum likelihood estimation (Hu & Bentler, 1998, 1999), Root Mean Square Error of the Approximation (RMSEA), and the Standardized Root Mean Square Residuals (SRMR).

If data do not conform to the specified conceptual model, a more suitable model can be identified through re-specification and retesting.

2. Covariance matrices will be used to compare aspects of implementation and student achievement using multi-group structural equation models.

3. Factor means and intercepts will be analyzed with respect to the structural equation model to determine whether differences are directly or indirectly related to specific aspects of implementation.

4. Results from the structural equation model will be compared to original results that were obtained using a combination of Regression, ANCOVA and Cluster Analyses.

As described in Chapter 4, while attempting to interpret results from the research procedures above, the researcher realized that results were not valid because the algorithms underlying *AMOS* (Arbuckle, 2006) are not appropriate to analyze categorical data (Ullman, p. 730; Byrne, 2001, p. 72; Blunch, 2008, p. 83; and Schumacker & Lomax, 2004, pp. 68-69). Therefore, software and statistical modeling approaches were sought that would yield valid results for a model involving a mixture of continuous and categorical variables, and teacher and student levels of data.

Revised Research Methodology

Mplus (Muthén & Muthén, 1998-2009) software was designed to conduct valid statistical analyses from a mixture of continuous and categorical variables (Muthén & Muthén, 2007, p. 3; Muthén, 2001; Schumacker & Lomax, pp. 68-69; Ullman, p. 730). Muthén referred to the following model options as “second-generation structural equation modeling” (2001, p. 291).

According to Muthén and Muthén (2007, p. 510):

> The underlying model of *Mplus* consists of three parts: the measurement model for the indicators of the continuous latent variables, the measurement model for the indicators of the categorical latent variables, and the structural model involving the continuous and categorical latent variables and the observed variables that are not indicators of the continuous or categorical latent variables.

The subsequently identified *Latent Class Analysis* model (described in detail in chapter 4) used logit parameterization to estimate logistic regressions for categorical latent variables (Muthén & Muthén, 2007, p.486). The analysis employed "maximum likelihood parameter estimates with standard errors … compared using a sandwich estimator" (p. 484). The EMA algorithm "is an accelerated [expectation maximization] procedure that uses Quasi-Newton and Fisher Scoring optimization steps when needed" (p.491).

Revised Data Analysis Procedures

Steps 0-4 articulated by Nylund (2007, p. 65) were followed to identify an appropriate model using *Mplus* (Muthén & Muthén, 1998-2009):

> Step 0: Study descriptive statistics
>
> Step 1: Study measurement model alternatives for each time point
>
> Step 2: Explore transitions based on cross-sectional results
>
> Step 3: Explore specification of the latent transition model without covariates
>
> Step 4: Include covariates in the … model

Findings and results for the proposed and revised research designs and analysis procedures are detailed in Chapter 4, as well as comparison to results from the original study in District A.

CHAPTER FOUR: FINDINGS AND RESULTS

Research procedures were completed as explained in chapter three. Rather than interpreting results, "second generation structural equation modeling" (Muthén, 2001, p. 291) was used to acquire valid results for interpretation. This process and the rationale for each decision are explained below.

First, summary descriptive statistics provide an overview of the sample (Table 5). Then findings and results are presented related to assumptions for structural equation modeling, latent variable structural equation model construction, goodness of fit, revised data analysis procedures, interpretation of results, and comparison to original findings.

Table 5

Summary Descriptive Statistics in Data Set

1996-97	1997-98	Variable or Descriptor
3323	1026	Students
1706	539	Male
1562	487	Female
936	263	Black
1475	487	White
912	276	Other
		GOALS scores
2735	1026	Fall
2767	1026	Spring
2179	1026	Complete data sets by student

109	30	Teachers
36	NA	Attended training
92	16	Elementary
17	14	Middle
58	NA	Implementation behavior rating (by teacher)
20		III
26		IVa
10		IVb
2		V
12	30	Fidelity rating (by teacher)
14	13	Schools
10	8	Elementary
4	5	Middle
10	0	Complete data sets by teacher

Findings and Results

1. Assumptions

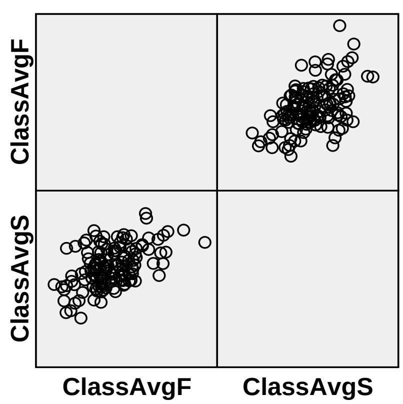
1.1 Sample size and missing data. A sample size of at least 60 is generally required to stabilize covariances in structural equation modeling (Bentler & Yuan, 1999). Original researchers combined both years of data to increase sample size and power of statistical tests. This approach was also employed for this study. While complete data sets by student exceed this criteria (2179 in 1996-97 and 1076 in 1997-98), only 10 teacher data sets contained all required data points for the analysis. Levels of Use (LoU) data were only available for 1996-97 for 58 of the 109 teachers. Innovation Configuration (IC) data were available for 12 teachers in 1996-97 and 30 teachers in 1997-98. Thirty-six teachers participated in initial training in 1996-97, and no such indicator existed in 1997-98. The 10 complete teacher data sets included LoU, IC, and attendance in initial training in 1996-97. It was not possible to conduct an SEM analysis using only original data; missing values had to be estimated in order to meet the 60 case minimum.

Several approaches were considered to estimate missing values: replacing missing values with mean values, imputation of missing values through similar response pattern imputation (Jörskog & Sörbom, 1993), and full information likelihood estimation (FIML) (Arbuckle, 1996; and Wothke, 2000). Listwise and pairwise deletion techniques were not considered because data were too sparse and, more importantly, research has demonstrated that they result in biased parameter estimates for data missing at random (Arbuckle, 1996; Muthén et al., 1987; and Wothke, 2000). Replacing missing values with mean values was ruled out because "regression and SEM studies have equivocally demonstrated that mean imputation results in biased parameter estimates under both [missing completely at random] and [missing at random] (Brown, 1994; Wothke, 2000)" (Enders & Bandalos, 2001, p. 430). Similar response pattern imputation (Jörskog & Sörbom, 1993) was considered but rejected, because Brown (1994) found that Type I error rates were inflated through this technique. Enders and Bandalos (p. 430) recommended FIML because it "yielded the lowest proportion of convergence failures and provided near-optimal Type I error rates." Schumacker and Lomax (2004, p. 43) concur, stating that "FIML is the recommended parameter estimation method when data are missing in structural equation analyses." Missing values were calculated using FIML in *AMOS (Arbuckle, 2006)* after the model was specified.

1.2 Multivariate outliers, normality and linearity. Data were screened for multivariate normality and outliers. Linear regression of quantitative variables as factors (class average for fall, class average for spring) and Teacher ID as dependent variable (for case identification) was used to evaluate Mahalanobis' Distance with 2 degrees of freedom, for a χ^2 critical value of 13.816 at p <.001 (Tabachnick & Fidell, 2007, p. 949). Values for Mahalanobis' Distance ranged from .017 to 13.606. No cases exceeded the χ^2 critical value, indicating an absence of multivariate outliers (Pallant, 2007, p. 157).

Multivariate normality and linearity between quantitative variables was assessed through examination of scatter plot matrices (Figure 4). Fall and spring GOALS scores produced generally elliptical shapes, indicating multivariate normality and linearity (Tabachnick & Fidell, 2007, p. 85).

Figure 4. Scatter Plot of Fall and Spring GOALS Scores

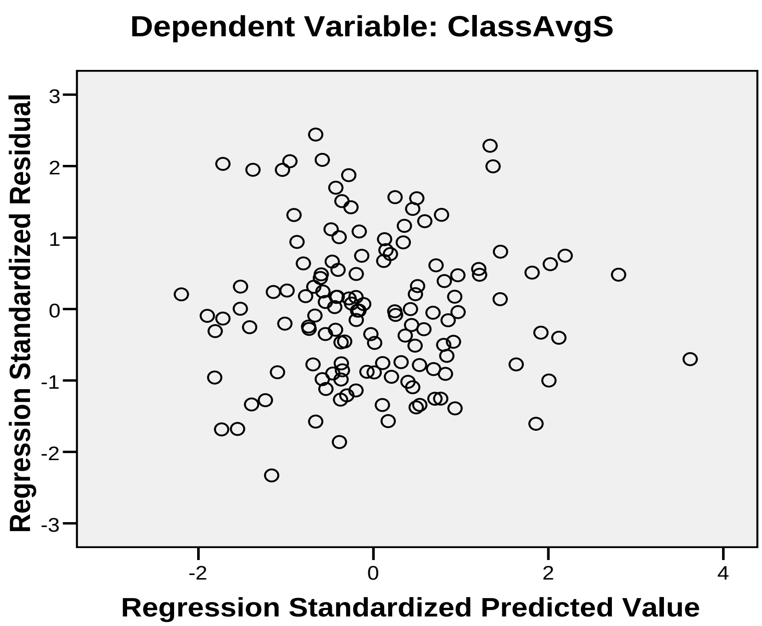


1.3 Absence of multicollinearity and singularity. Data were assessed for multicollinearity and singularity using multiple regression, with fall class average, grade, LoU ratings, and IC sums as independent variables (IV) to predict spring class averages (dependent variable, DV). Singularity was not an issue, given that Pearson Correlations between DV and IVs were greater than .3 (Pallant, 2007, p. 155). Correlation among IVs was less than .7 (Pallant, p. 155) with the exception of IC sum and LoU, which were correlated at -.712 (ClassAvgF correlated to LoU2 at .185, and IC_Sum at .016). Pallant recommends removal of one of the highly correlated variables. Given that original researchers used these measures in analyses, only 10 complete data sets existed with both ratings, and re-analyses using mean values to replace missing values resulted in a correlation of -.186, this violation of assumption was ignored.

Collinearity statistics were reviewed with respect to tolerance and variance inflation factor (VIF), and all were within acceptable levels. Tolerance coefficients were well above .1 (.921 ClassAvgF, .454 LoU2, and .470 IC_Sum) while VIF coefficients were well below 10 (1.086 ClassAvgF, 2.200 LoU2, and 2.126 IC_Sum), indicating absence of multicollinearity (Pallant, p. 156).

1.4 Residuals. Residual plots were assessed using spring class average scores as dependent variable and fall class average scores as independent variable in linear regression. Standardized residuals were compared to predicted values. The distribution of residuals in a rectangle clustered around zero, with most values between 3 and -3, indicates generally acceptable levels of multivariate homogeneity of variance-covariance (Mertler & Vannatta, 2005, p. 57).

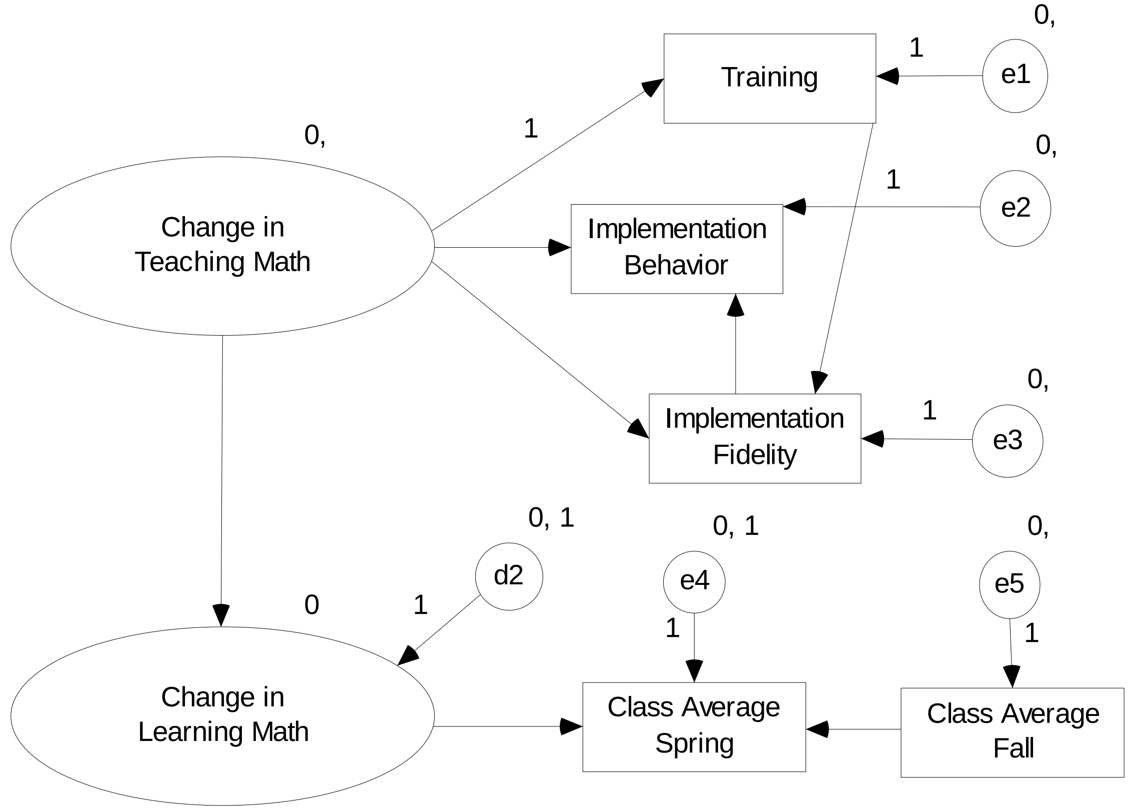
Figure 5. Scatter Plot of Standardized Residuals and Standardized Predicted Values



2. Model relationships among student outcomes, levels of implementation behavior, and fidelity of implementation of constructivist approaches to teaching mathematics.

2.1 Latent variable structural equation model construction. The originally proposed model could not be specified in *AMOS (Arbuckle, 2006)* because it combined teacher-level and student-level data. Data were recoded for analysis at the teacher level, using class means on the GOALS exams for fall and spring instead of individual student scores. Student characteristics were necessarily eliminated. An unintended benefit of decreasing the number of variables was a lower minimum number of observations for model specification (Ullman, 2007, p. 682). The revised model is illustrated in Figure 6.

Figure 6. Structural Equation Model used for this Analysis



Goodness of fit. The model was evaluated using five goodness of fit indices.

1. Convergence of the model yielded a chi-square (χ^2) value of 5.575 with 3 degrees of freedom (*df*) at probability (*p*) =.134. The *p* value is used to test the hypothesis that the "observed covariances among the measured variables arose because of the relationships between variables specified in the model" (Ullman, 2007, p. 695). H_0 was retained because p>.05, indicating that the model fits the data (Ullman, p. 695). The ratio of χ^2 to *df* is 1.86; Ullman indicates that a ratio below 2 generally indicates a "good-fitting model" (p. 715).

    Model fit indices in *AMOS (Arbuckle, 2006)* compare the default model (the one being tested) to a saturated model (a perfect model with 0 *df*) and an independence model (a model that responds to "unrelated variables" with *df* equal to data points less the number of variables estimated; Ullman, p. 716).

2. The Bentler and Bonett (1980) normed fit index (NFI) compares the default model's χ^2 to the χ^2 of the independence model. Values range from 0-1; those >.95 indicate good model fit (Ullman, p. 716). Arbuckle indicates that models yielding values <.91 can generally be improved (2007, p. 598), while Bearden, Sharma, and Teel (1982) found that NFI tends to underestimate a model that fits a small sample size well. **The default model yielded an NFI of .917, indicating a moderate fit**. Given the small sample size, Bentler's (1988) comparative fit index (CFI) provides more reliable approach to evaluating fit.

3. CFI represents the value of $1\text{-}\tau_{\text{est. model}}/\ \tau_{\text{indep. model}}$, where τ represents noncentral χ^2 distribution with noncentrality parameters τ_i. In a perfect model, τ_i=0 (Ullman, p. 717). Results range from 0-1. Hu and Bentler (1999) found that values of .95 or

greater indicated good model fit. **The CFI value of the default model is .951, indicating a good-fitting model**.

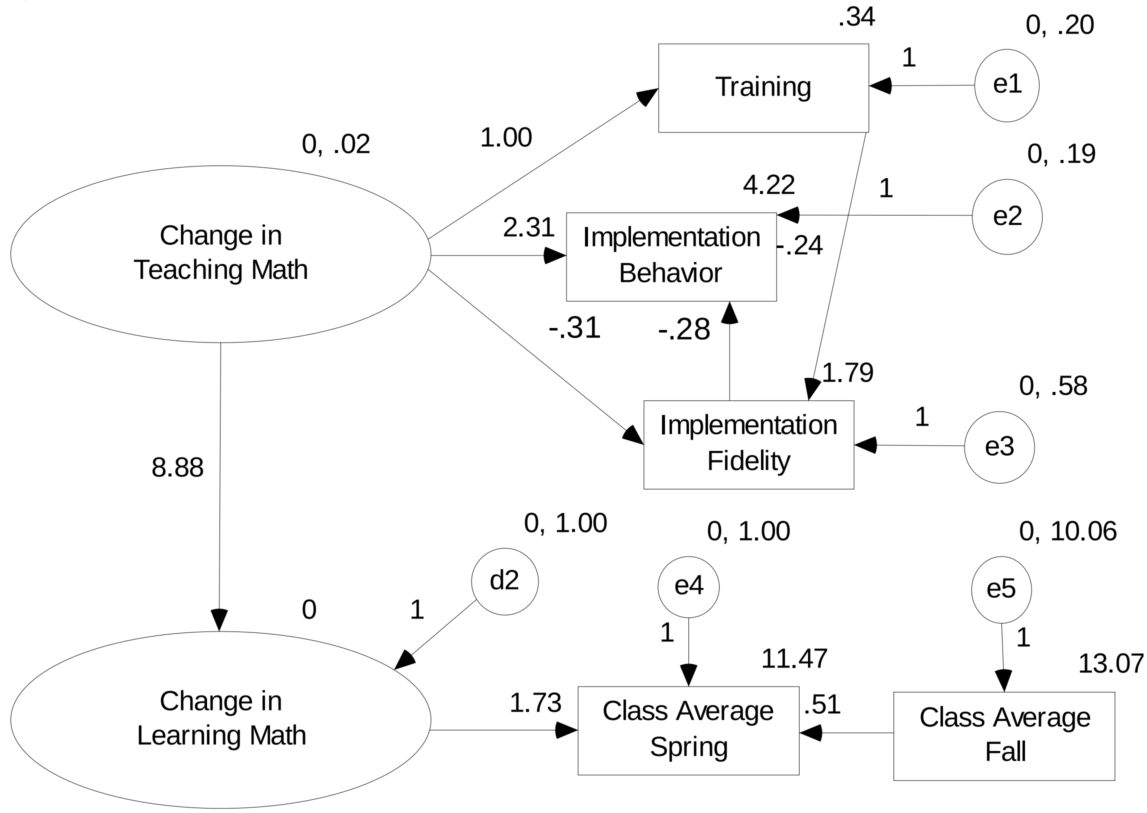
4. Root mean square error of approximation (RMSEA; Browne & Cudeck, 1993, in Ullman, p. 717) indicates absence of fit in a model compared to the saturated model. Hu and Bentler (1999) found that values less than or equal to .06 indicate good model fit relative to *df*. Browne and Cudeck (1993) found that values larger than .10 indicate a poor model fit. **The RMSEA of the default model is .079, indicating good model fit**.
5. Parsimony involves using the fewest parameters possible to achieve a certain level of fit (Schumaker & Lomax, 2004, pp. 104-105). Model parsimony is assessed using Akaike Information Criterion (AIC; Akaike, 1987). In *AMOS (Arbuckle, 2006)*, this value is calculated by adding χ^2 to double the number of parameters in a model. Lower values are more parsimonious, but no absolute scales have been established to evaluate results. **The AIC of the default model was a relatively parsimonious 39.575**; slightly lower than the saturated model, 40.000; and twice as parsimonious as the independence model, 77.369.

In summary, the model indicated goodness of fit. Though NFI indicated a moderate fit, this was ignored due to the influence of sample size on the measure, and good fit in all indices not affected by small sample sizes.

2.2 Calculation of estimates of the effects of the model. When missing values are estimated, covariance matrices are not produced. Therefore it was not possible to compare aspects of implementation and student achievement as anticipated.

In order for the model to be identified, parameter values had to be fixed to 1.00 for variables d2 and e4, so that a unique numerical solution could be reached. Estimated effects of the model are displayed in figure 7.

Figure 7. Estimates of Effects



At this point, the researcher realized that results were not valid because the algorithms underlying *AMOS* (Arbuckle, 2006) are not appropriate to analyze categorical data (Ullman, p. 730; Byrne, 2001, p. 72; Blunch, 2008, p. 83; and Schumacker & Lomax, 2004, pp. 68-69). Therefore, no interpretation was provided. Instead, software and statistical modeling approaches were sought that would yield valid results for a model involving a mixture of continuous and categorical variables, and teacher and student levels of data. *Mplus* (Muthén & Muthén, 1998-2009) software was designed to conduct valid statistical analyses from a mixture of continuous and categorical variables (Muthén & Muthén, 2007, p. 3; Muthén, 2001; Schumacker & Lomax, pp. 68-69; Ullman, p. 730). Muthén referred to the following model options as "second-generation structural equation modeling" (2001, p. 291).

Revised Data Analysis Procedures

Steps 0-4 articulated by Nylund (2007, p. 65) were followed to identify an appropriate model using *Mplus* (Muthén & Muthén, 1998-2009):

Step 0: Study descriptive statistics

Step 1: Study measurement model alternatives for each time point

Step 2: Explore transitions based on cross-sectional results

Step 3: Explore specification of the latent transition model without covariates

Step 4: Include covariates in the … model

Step 5: Include distal outcomes and advanced modeling extensions

Step 5 was not used because distal outcomes were not collected, and data were collected over a time period too short to apply advanced modeling extensions.

The data set was modified to a .dat file in free format for analysis in *Mplus* (Muthén & Muthén). Free format would not work properly with missing data, so several steps were taken to decrease the number of rows with missing data and to streamline the number of parameters that would be considered in analysis.

- Rather than treat each grade level distinctly, grades were divided into elementary (1) representing grades 2-5, and middle school (2) representing grades 6-8.
- Implementation behavior and fidelity ratings were combined and consolidated into three categories: High (1), medium (2), and low (3). Levels of use (LoU) ratings of III were designated "low," IVa was designated "medium," and IVb and V were designated "high." Innovation configuration/fidelity ratings (IC) were assigned to high, medium, and low categories by original researchers. In the ten cases where a teacher had both LoU and IC ratings, if one conflicted with the other, the following rules were applied: low + high = medium, medium + high = high, low + medium = medium.
- Given the unanalyzed indirect affect postulated between gender and ethnicity, an "*egen*" variable was created to combine the two, where 11=Black male, 12=Black female, 21=White male, 22=White female, 31=Other ethnicity male, and 32=Other ethnicity female. Ethnic categories followed those assigned by original researchers.
- Analysis using a combination of 1996-97 and 1997-98 data was conducted to achieve higher power and for direct comparison to the original study. It should be noted that student-level data were not available for longitudinal study as

unique student codes were assigned by original researchers for 1996-97 and 1997-98. The same student number was used for fall and spring GOALS results within those years.

- One teacher had different quality ratings in each year. Ratings for 1996-97 were removed (*n*=5) and 1997-98 were kept because the latter score was more recent and there were many more students in the 1997-98 class.
- Rows with missing data were deleted to facilitate correct analysis with free format data. Data commonly missing were fall or spring GOALS scores, or ratings on teacher implementation behavior and fidelity.

Step 0. Descriptive statistics. Descriptive statistics were generated from the data set described above: frequencies for all variables; and mean, median, mode, standard deviation, skewness, kurtosis, and histograms for continuous variables.

Frequency tables (provided in Appendix B, due to length) indicated 2,138 complete data sets, nearly evenly distributed across 1996-97 (1135, 53.1%) and 1997-98 (1003, 46.9%). There were slightly more males than females. Nearly half (47.7%) were Black, while about one-fourth were White (26.1%) and Other (26.2%), respectively. Slightly more students attended middle school than elementary (52.2% compared to 47.8%). GOALS scores ranged from 0-30, with generally higher values in spring compared to fall. Quality of instruction, represented by the combined IC and LoU rating categories, were predominantly high (37.2%) and medium (41.9%), with a minority receiving low quality instruction (20.9%).

The following table provides descriptive statistics for the two continuous variables *goalf* and *goals*, representing fall and spring GOALS scores. The mean, median, and mode were markedly higher in spring than fall, while the standard deviation was nearly the same. This indicates overall academic growth within the student population.

Table 6

Fall and Spring GOALS Scores

Descriptive statistic	***goalf***	***goals***
N (valid)	2138	2138
Mean	13.57	18.32
Standard error of mean	.126	.141
Median	13.00	19.00
Mode	13	20
Standard deviation	5.808	6.522
Variance	33.745	42.531
Skewness	.223	-.258
Kurtosis	-.459	-.731
Standard error of kurtosis	.106	.106
Range	30	29
Minimum	0	1
Maximum	30	30

Step 1. Measurement model alternatives for each time point. Results from fall and spring GOALS scores were compared, representing two time points, irrespective of year. Table 7 demonstrates that four classes appeared to provide the best division of latent classes within the data in both fall and spring models. Latent Class Analysis (LCA) was the preferable model for determining the number of classes at both time points in terms of highest loglikelihood and lowest Bayesian Information Criterion (BIC; Schwartz, 1978) values among the models being compared.

According to Nylund (2007, pp. 31-32), LCA "uses an underlying latent variable to describe the relationship among a set of observed items," and indicates "the prevalence of each class in the population, or relative frequency of class membership."

Similar in function to indices mentioned on pages 54-55, loglikelihood and BIC are fit indices for regression models. Loglikelihood is "based on summing the probabilities associated with the predicted and actual outcomes for each case (Tabachnick & Fidell, 2007, p. 446).

$$\text{log likelihood} = \sum_{i=1}^{N}\left[Y_i \ln(Y_i) + \left(1 - \hat{Y}_i\right)\ln\left(1 - \hat{Y}_i\right)\right]$$

Therefore, the number closest to zero indicates a better fitting model. Conversely, BIC is a measure of parsimony, based on the assumption that "all other things being equal, for two models that have equal loglikelihoods, the model with the fewest parameters and larger sample size is better. … The BIC applies a penalty for the number of parameters (*g*) and the sample size, and is defined as (Nylund, 2007, p. 41)

BIC = -2 loglikelihood + g log(n)."

Table 7

Latent Class Analysis, Latent Class Factor Analysis, and Factor Mixture Analysis Model Results for Fall and Spring GOALS Scores (N=2143)

Model	**Fall Loglikelihood**	**Fall BIC**	**Spring Loglikelihood**	**Spring BIC**	**No. Parameters**
LCA, 2c	-6785.224	13601.127	-6995.904	14022.487	4
LCA, 3c	-6777.290	13600.599	-6974.115	13994.249	6
LCA, 4c	**-6772.957**	**13607.273**	**-6962.595**	**13986.550**	**8**
LCA, 5c	-6772.123	13620.945	-6959.264	13995.228	10
LCA, 6c	-6771.627	13635.294	-6951.101	13994.242	12
LCFA 1f, 2c	-8268.251	16582.523	-8460.704	16967.428	6
LCFA 1f, 3c	-8251.233	16571.495	**-8432.966**	**16934.962**	**9**
LCFA 1f, 4c	**-8244.390**	**16580.820**	-8241.490	16935.012	**12**
LCFA 1f, 5c	-8241.769	16598.588	-8417.420	16949.890	15
LCFA 1f, 6c	-8241.128	16620.315	-8416.773	16971.604	18
FMA 1f, 2c	-8268.251	16597.863	-8450.704	16982.768	8
FMA 1f, 3c	-8251.233	16586.835	-8432.966	16950.302	11
FMA 1f, 4c	**-8244.390**	**16603.830**	**-8421.490**	**16958.029**	**15**
FMA 1f, 5c	-8243.636	16633.002	-8417.421	16980.570	19
FMA 1f, 6c	Did not converge		-8410.461	16997.330	23

Note. C stands for number of latent classes, f stands for number of latent factors, LCA=latent class analysis, LCFA=latent class factor analysis, and FMA=factor mixture analysis.

Interpreting the classes. Latent class means and variances from the LCA model were used to understand the meaning of the various classes. It should be noted that classes were nominal, assigned by *Mplus* (Muthén & Muthén) based on common statistical characteristics; therefore, class 1 in fall is not necessarily the same as class 1 in spring. Labels were given by the researcher to clarify the meaning of each latent class of math learner, with respect to mastery demonstrated on fall and spring GOALS exams, respectively. *Master* indicates the highest level of demonstrated mastery, *skilled* indicates a learner that demonstrated mastery on the majority of questions, *apprentice* indicates a learner that occasionally demonstrated the required mathematical skill, and *novice* indicates a learner that rarely or never demonstrated the required mathematical skill.

Table 8

Distribution of Students into Latent Performance Classes

Latent Class - Fall	**Mean**	***SE***	**Est./*SE***	***P*-Value**	**N**	**Proportion**
1 – Skilled	16.911	0.357	47.309	0.000	699	0.32618
2 – Apprentice	11.299	0.404	27.995	0.000	895	0.41764
3 – Master	23.304	0.335	69.467	0.000	216	0.10079
4 – Novice	6.284	0.438	14.355	0.000	333	0.15539
Latent Class - Spring	**Mean**	***SE***	**Est./*SE***	***P*-Value**	**N**	**Proportion**
1 – Master	25.704	0.258	99.654	0.000	636	0.29678
2 – Apprentice	13.649	0.587	23.256	0.000	500	0.23332
3 – Skilled	19.543	0.389	50.184	0.000	703	0.32804
4 – Novice	7.803	0.341	22.883	0.000	304	0.14186

Step 2. Transitions based on cross-sectional results. Student categories were then examined in cross-sectional analysis without considering covariates (see Table 9).

Table 9

Number of Students in Each Learner Class in Fall and Spring GOALS Exams, based on Cross-Sectional LCA without Covariates

Category Fall	**Master (1) Spring**	**Skilled (3) Spring**	**Apprentice (2) Spring**	**Novice (4) Spring**	**Total**
Master (3)	173	286	166	11	**636**
Skilled (1)	35	292	323	53	**703**
Apprentice (2)	8	111	280	101	**500**
Novice (4)	0	10	126	168	**304**
Total	**216**	**699**	**895**	**333**	**2143**

The fall *master* class demonstrated a fair amount of mobility: 173 remained *master* from fall to spring, 286 fall *masters* were *skilled* in spring, 166 were *apprentices* in spring, and 11 transitioned from *master* in fall to *novice* in spring. There were 636 *master* students in fall, and only 216 in spring. The *skilled* class had about the same number of students in fall and spring (703 and 699, respectively), but still exhibited mobility among classes. Specifically, 35 were *masters* in spring, 292 remained *skilled*, 323 were *apprentices*, and 53 transitioned to *novice*. The *apprentice* class grew from 500 in fall to 895 in spring. Among the 895 spring *apprentices* were 166 fall *masters*, 323 fall *skilled*, 280 fall *apprentices*, and 126 fall *novices*. There were slightly more *novices* in spring than in fall

(333 compared to 304), including: 11 fall *masters*, 53 fall *skilled*, 101 fall *apprentices*, and 168 fall *novices*. While it may appear that a slight downward trend occurred, one should note the higher mean scores for each class in spring as compared to fall, with a greater increase between fall and spring *master* and *skilled* classes. A complete model with covariates would provide more meaningful information.

Step 3. Specification of the LCA model without covariates. A latent class analysis model was used to analyze growth classes from fall to spring. A model without covariates was considered first, in order to explore how well four latent classes described the data and to ensure model convergence (Nylund, 2007, p. 101). *Mplus* (Muthén & Muthén) input is provided in Appendix C.

The model converged for two through four latent classes without covariates. As depicted in Table 10, four classes provided the lowest BIC and highest loglikelihood (closest to zero).

Table 10

Fit Indices for 4, 3 and 2 Latent Classes without Covariates

Latent Classes	**Loglikelihood**	**BIC**	**Parameters**
4	**-13190.408**	**26480.496**	**13**
3	-13222.297	26521.271	10
2	-13389.915	26833.503	7

Step 4. Inclusion of covariates. Covariates were then added to the model (see Appendix D and E for input). As anticipated, model fit increased, indicating that gender and ethnicity, and teaching quality played an important role in the dynamics of this construct.

Table 11

Comparative Model Fit Indices for 4, 3 and 2 Latent Classes with Covariates

Model	Latent Classes	Loglikelihood	BIC	Parameters
LCA	**4**	**-13045.509**	26367.053	36
LCA	3	-13080.874	**26361.306**	**26**
LCA	2	-13279.027	26680.736	16
GMM	4	**-13590.278**	27418.252	31
GMM	**3**	-13593.378	27386.114	26
GMM	2	-13626.455	**27383.260**	**17**

Note. *LCA* stands for latent class analysis and *GMM* for 2-level growth mixture model

Fit indices for two models including covariates were compared for four, three, and two latent classes. The *LCA* model (Muthén & Muthén, 2007, p. 148) produced a higher loglikelihood and lower BIC than the 2-level *GMM* (p. 309). Analyses using *LCA* were carried out using four classes. Four classes were chosen because fit indices were very close between three and four classes, but four classes provided more even comparison group sizes. Three classes resulted in a middle level class of n=1106 compared to 527 and 505 in classes 1 and 3 respectively.

Performance comparison of two models clearly favored the *LCA*. Although two-level growth mixture modeling better represented the situation of students grouped within classes and the theory being tested, performance of the model was too unstable to ensure replicable results. A more complete data set could potentially resolve this situation by aiding in model identification. Stability concerns were not evident in the *LCA* model, so final results were provided below.

2.3 Final results. Model estimation terminated normally, indicating a loglikelihood of -13045.509, 36 free parameters, AIC of 26163.019, BIC of 26367.053, and entropy of .657. Individuals were placed into their most likely latent class membership based on the combination of fall and spring GOALS scores, as presented in Table 12, and probabilities for class membership were reviewed (Table 13). Names given to classes earlier in this chapter were also used in Table 12. *Master* indicates the highest level of demonstrated mastery, *skilled* indicates a learner that demonstrated mastery on the majority of questions, *apprentice* indicates a learner that occasionally demonstrated the required mathematical skill, and *novice* indicates a learner that rarely or never demonstrated the required mathematical skill on the GOALS fall and spring exams.

Table 12

Classification of Individuals Based on Most Likely Latent Class Membership

Latent Class	Class Counts	Proportions	Class Name
1	745	0.34846	Skilled
2	381	0.17820	Novice
3	634	0.29654	Apprentice
4	378	0.17680	Master

Average probabilities for latent classification by latent class in Table 13 indicate that latent classes strongly represent actual class membership. For example, the probability that individuals in latent class 2 – *novice* are correctly classified is 0.878, 0.000 that they are actually *skilled* (class 1) or *master* (class 4), and only 0.088 that they may actually be *apprentice* (class 3).

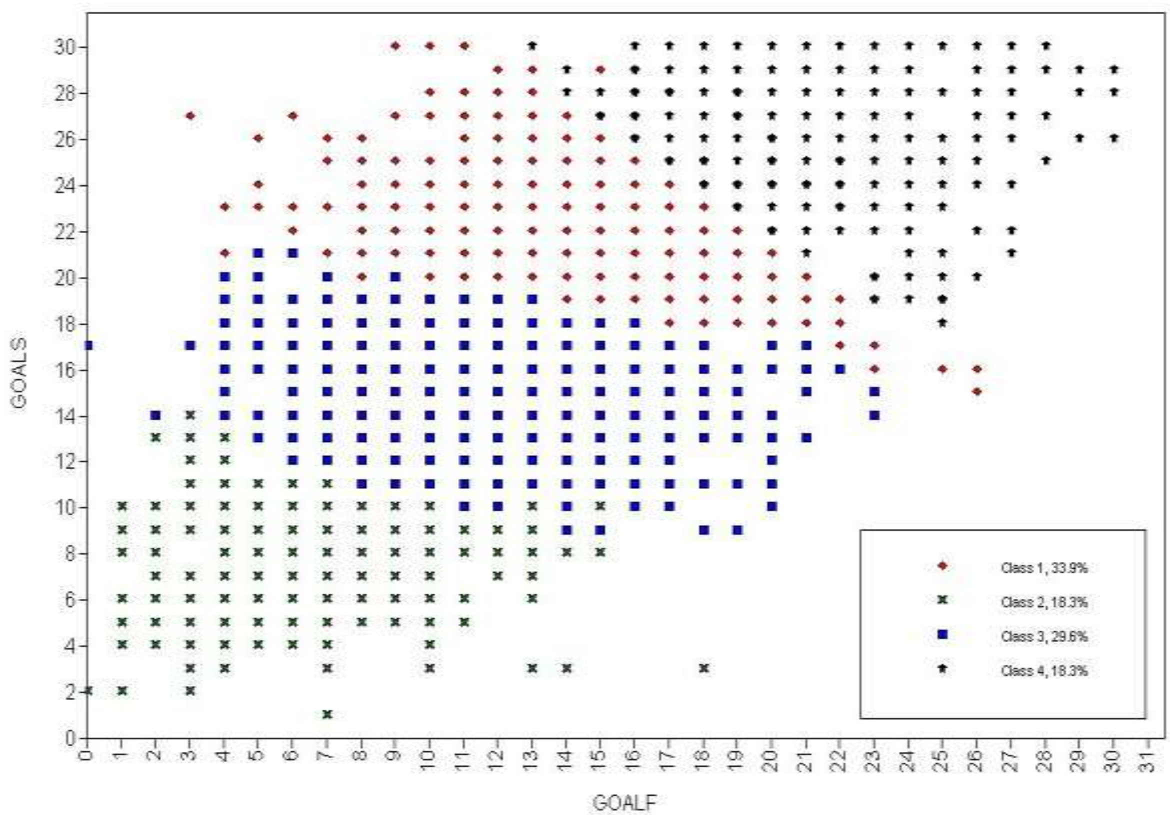
Table 13

Average Latent Class Probabilities for Most Likely Membership by Latent Class

Most likely latent class membership	Latent class 1 *Skilled*	Latent class 2 *Novice*	Latent class 3 *Apprentice*	Latent class 4 *Master*
1 – *Skilled*	0.758	0.000	0.145	0.097
2 – *Novice*	0.001	0.878	0.122	0.000
3 – *Apprentice*	0.157	0.088	0.753	0.001
4 – *Master*	0.158	0.000	0.002	0.840

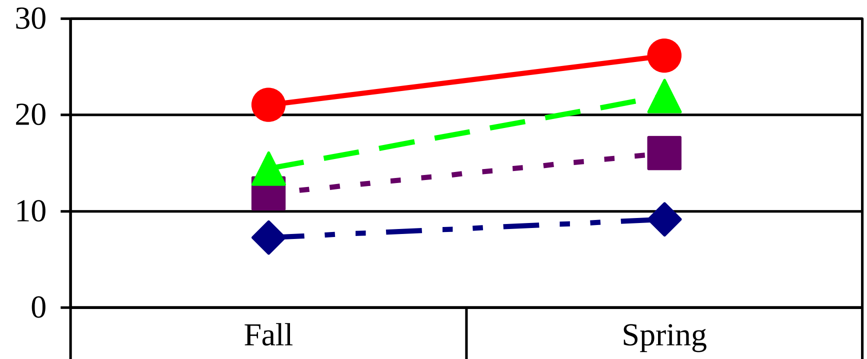
Scoring patterns in light of latent classification are depicted in Figure 8. Fall scores are represented horizontally (on the x axis) and spring scores, vertically (on the y axis). For example, *novices* (class 1) scored low in fall and spring, and therefore cluster in the lower left corner of the scatter plot while *masters* (class 4) generally scored highest in fall and spring, and therefore cluster in the upper right corner of the scatter plot.

Figure 8. Scatter Plot of Latent Classes by Fall (*goalf*) and Spring (*goals*) GOALS Scores



Average fall and spring scores in light of researcher-named latent classifications are depicted in Figure 9.

Figure 9. Latent Classes by Average Fall and Spring GOALS Scores



	Fall	Spring
Novice	7.272	9.159
Apprentice	11.832	16.028
Skilled	14.443	21.968
Master	21.048	26.167

Model results are presented in Table 14. The *estimate* is an “unstandardized regression coefficient of *B* and represents the effect the [independent variable] has on the [dependent variable]. *SE* is the standard error of *B*,” (Mertler & Vannatta, 2005, p. 320). *Est./SE* approximates a standardized score (provides a standard normal distribution [Hosmer & Lemeshow, 2000]). “The critical value for a two-tailed test at the .05 level is an absolute value greater than 1.96,” (Muthén & Muthén, 2007, p. 575). This is otherwise known as the Wald test (Hosmer & Lemeshow, p. 16). Confidence intervals were calculated, as specified in Hosmer and Lemeshow (2000, pp. 17-18), and reported with results.

Mean scores increased from fall to spring in every latent class. *Low quality instruction* was associated with lower spring GOALS scores, though fall scores had stronger effects on spring scores than instruction, as anticipated. With the covariate effect of fall scores removed, gains in resulting spring GOALS scores were significantly different than 0. Confidence intervals from fall and spring were then examined for each latent class to determine whether growth was statistically significant and very unlikely to be caused by chance. *Skilled* and *master* classes indicated significant growth demonstrated (conservatively) by lack of overlap between confidence intervals. A positive treatment effect of instruction was therefore inferred.

Table 14

Latent Class Analysis Model Results, Part 1

Results	Estimate	SE	Est./SE	2-tailed *p*-value	95% CI Lower, Upper
Latent class 1 – *Skilled*					
Fall GOALS	14.443	0.898	*16.075*	0.00	12.682,16.204
Spring GOALS	21.968	1.419	*15.481*	0.00	19.187, 24.749
Latent class 2 – *Novice*					
Fall GOALS	7.272	0.425	*17.116*	0.00	6.439, 8.105
Spring GOALS	9.159	0.542	*16.913*	0.00	8.098, 10.221
Latent class 3 – *Apprentice*					
Fall GOALS	11.832	1.006	*11.761*	0.00	9.860, 13.804

Spring GOALS		16.028	1.664	*9.630*	0.00	12.766, 19.290
Latent class 4 – *Master*						
Fall GOALS		21.048	0.988	*21.306*	0.00	19.112, 22.984
Spring GOALS		26.167	0.355	*73.660*	0.00	25.471, 26.864
Values common to all latent classes						
Spring GOALS on	High quality	-0.169	0.355	-0.477	0.63	-0.865, 0.526
Spring GOALS on	Low quality	-1.241	0.607	*-2.047*	0.04	-2.430, -0.052
Variances	Fall GOALS	15.103	1.047	*14.422*	0.00	13.050,17.155
Residual variances	Spring GOALS	9.290	1.077	*8.623*	0.00	7.179, 11.402

Note. Values exceeding the critical value of 1.96 at *p*<.05 are italicized.

Using class 2 (*novice*) as a reference, estimated effects of categorical latent variables on GOALS scores were explored. There were statistically significantly positive effects of high quality instruction in classes 1, 3, and 4. White males were significantly negatively associated with classes 1 (*skilled*), 3 (*apprentice*) and 4 (*master*), while White females were negatively associated with class 3. No other results were significant.

Table 15

Latent Class Analysis Model Results, Part 2: Estimated Effects of Categorical Latent Variables, Compared to Reference Class 2 (Novice)

Categorical latent variables		Estimate	SE	Est./SE	2-tailed *p*-value
C#1 (*skilled*) on	High quality	0.930	0.233	*3.995*	0.00
C#1 on	Low quality	-0.103	0.438	-0.234	0.82
C#1 on	Black male	0.184	0.356	0.518	0.61
C#1 on	Black female	0.469	0.332	1.413	0.16
C#1 on	White male	-0.820	0.329	*-2.489*	0.01
C#1 on	White female	-0.531	0.330	-1.609	0.11
C#1 on	Other male	0.074	0.353	0.211	0.83
C#3 (*apprentice*) on	High quality	0.897	0.286	*3.141*	0.00
C#3 on	Low quality	0.157	0.232	0.676	0.50
C#3 on	Black male	-0.363	0.580	-0.626	0.53
C#3 on	Black female	-0.177	0.446	-0.398	0.69
C#3 on	White male	-1.156	0.457	*-2.532*	0.01
C#3 on	White female	-0.708	0.353	*-2.010*	0.04
C#3 on	Other male	0.021	0.460	0.046	0.96
C#4 (*master*) on	High quality	1.286	0.379	*3.392*	0.00
C#4 on	Low quality	-0.496	0.338	-1.469	0.14
C#4 on	Black male	0.198	0.357	0.553	0.58
C#4 on	Black female	0.510	0.331	1.542	0.12
C#4 on	White male	-3.317	1.079	*-3.075*	0.00

C#4 on	White female	-2.510	1.043	*-2.407*	0.02
C#4 on	Other male	-0.291	0.396	-0.735	0.46
Intercepts	C#1	0.473	0.566	0.835	0.40
Intercepts	C#3	0.637	0.318	2.003	0.05
Intercepts	C#4	0.058	0.502	0.116	0.91

Note. Values exceeding the critical value of 1.96 at $p<.05$ are italicized.

Next, logistic regression odds ratio results for each independent variable were considered. Mertler and Vannatta (2005, p. 320) state that "the odds ratio represents the increase (or decrease if *Exp [B]*, is less than 1) in odds of being classified in a category when the predictor variable increases by one." For example, Table 16 indicates that White males were 8.6 times more likely to be in class 3 (*apprentice*) than in the reference class (*master*) for each unit of increase in GOALS scores, while Black females were half as likely to be in class 3 than in the reference class for each unit of increase in GOALS scores. Odds of greatest magnitude related to White males, White females, and other males. Variability was so high; any conclusions would be tenuous at best.

With respect to *quality of instruction*, classes 1-3 were all less likely than the reference class, class 4 (*master*), to receive *high quality instruction*, and more likely to receive *low quality instruction*. For example, class 2 (*novice*) was one-quarter as likely to receive high quality instruction and 1.6 times more likely to receive *low quality instruction* (in terms of teacher implementation behavior and quality related to constructivist mathematics). Implications will be discussed in chapter five.

Table 16

Logistic Regression Odds Ratios, using Class 4 (Master) as the Reference Class

Latent variables	Predictor variable	Odds ratio exp.(*B*)	95% CI
C#1 (*skilled*) on	High quality instruction	0.701	0.400, 1.229
C#1 on	Low quality instruction	1.483	0.727, 3.023
C#1 on	Black male	0.987	0.542, 1.797
C#1 on	Black female	0.960	0.524, 1.756
C#1 on	White male	12.153	1.227, 120.339
C#1 on	White female	7.236	0.755, 69.343
C#1 on	Other male	1.442	0.608, 3.419
C#2 (*novice*) on	High quality instruction	0.276	0.132, 0.581
C#2 on	Low quality instruction	1.643	0.847, 3.185
C#2 on	Black male	0.821	0.408, 1.652
C#2 on	Black female	0.600	0.314, 1.148
C#2 on	White male	27.581	3.330, 228.437
C#2 on	White female	12.304	1.593, 95.014
C#2 on	Other male	1.338	0.615, 2.910
C#3 (*apprentice*) on	High quality instruction	0.678	0.435, 1.056
C#3 on	Low quality instruction	1.921	1.022, 3.611
C#3 on	Black male	0.571	0.266, 1.226
C#3 on	Black female	0.503	0.263, 0.961
C#3 on	White male	8.681	0.663, 113.623
C#3 on	White female	6.058	0.691, 53.128
C#3 on	Other male	1.367	0.715, 2.613

An *LCA* with 4 classes was also run using *high quality instruction* as the reference group. The original study indicates that it is unlikely for a teacher to transition from *low* to *high quality* configurations or implementation behaviors in the same year. However, transition from *low* to *medium quality instruction* was much more common. This aspect of the results indicates the difference that such a transition could make.

Model estimation terminated normally, indicating the same loglikelihood, free parameters, AIC, BIC, and entropy as reported on page 64. Individuals were placed into their most likely latent class membership based on the combination of fall and spring GOALS scores, as presented in Table 17, and probabilities for class membership were reviewed (Table 18). Model results are presented in Table 19.

Table 17

Classification of Individuals Based on Most Likely Latent Class Membership

Latent Class	Class Counts	Proportions	Class Name
1	381	0.17820	Novice
2	745	0.34846	Skilled
3	378	0.17680	Master
4	634	0.29654	Apprentice

Table 18

Average Latent Class Probabilities for Most Likely Latent Class Membership

Most likely latent class membership	Latent class 1 *Novice*	Latent class 2 *Skilled*	Latent class 3 *Master*	Latent class 4 *Apprentice*
1 – Novice	0.878	0.001	0.000	0.122
2 – Skilled	0.000	0.758	0.097	0.145
3 – Master	0.000	0.158	0.840	0.002
4 – Apprentice	0.088	0.157	0.001	0.753

Mean scores increased significantly from fall to spring in every latent class. *Low quality instruction* was associated with lower spring GOALS scores, and *medium quality instruction* was associated with higher spring GOALS scores. Fall scores had stronger effects on spring GOALS scores than quality of instruction, as anticipated. With the covariate effect of fall scores removed, gains in resulting spring GOALS scores were significantly different than 0. Confidence intervals from fall and spring were then examined for each latent class to determine whether growth was statistically significant and very unlikely to be caused by chance. *Skilled* and *master* classes indicated significant growth demonstrated (conservatively) by lack of overlap between confidence intervals. A positive treatment effect of instruction was therefore inferred.

Table 19

Latent Class Analysis Model Results, Part 1, Comparing Impacts of Medium and Low Quality Instruction on Student Learning

Results	Estimate	SE	Est./SE	2-tailed p-value	95% C.I. Lower, Upper
Latent class 1 – Novice					
Fall GOALS	7.272	0.425	*17.115*	0.00	6.439, 8.105
Spring GOALS	8.990	0.756	*11.895*	0.00	7.509, 10.472
Latent class 2 – Skilled					
Fall GOALS	14.443	0.898	*16.075*	0.00	12.682, 16.204
Spring GOALS	21.799	1.652	*13.192*	0.00	18.560, 25.037
Latent class 3 – Master					
Fall GOALS	21.048	0.988	*21.307*	0.00	19.112, 22.984
Spring GOALS	25.998	0.359	*72.370*	0.00	25.294, 26.702
Latent class 4 – Apprentice					
Fall GOALS	11.832	1.006	*11.761*	0.00	9.860, 13.804
Spring GOALS	15.859	1.882	*8.424*	0.00	12.169, 19.548
Values common to all latent classes					

Spring GOALS on	Medium quality	0.169	0.355	0.477	0.63	-0.526, 0.865
Spring GOALS on	Low quality	-1.072	0.504	*-2.126*	0.03	-2.061, -0.084
Variances	Fall GOALS	15.103	1.047	*14.422*	0.00	13.050, 17.155
Residual variances	Spring GOALS	9.290	1.077	*8.624*	0.00	7.179, 11.402

Note. Values exceeding the critical value of 1.96 at $p<.05$ are italicized.

Using class 1 (*novice*) as a reference, estimated effects of categorical latent variables on GOALS scores were explored. There were statistically significantly negative effects of medium and low quality instruction in classes 2 (*skilled*), 3 (*master*), and 4 (*apprentice*). White males were significantly negatively associated with class 2 (*skilled*), and White males and females were significantly negatively associated with classes 3 (*master*) and 4 (*apprentice*). No other results were significant.

Table 20

Latent Class Analysis Model Results, Part 2: Estimated Effects of Categorical Latent Variables, Compared to Reference Class 1 (Novice)

Categorical latent variables		**Estimate**	**SE**	**Est./SE**	**2-tailed *p*-value**
C#2 (*skilled*) on	Medium quality	-0.930	0.233	*-3.995*	0.00
C#2 on	Low quality	-1.033	0.372	*-2.779*	0.01
C#2 on	Black male	0.184	0.356	0.518	0.61
C#2 on	Black female	0.469	0.332	1.413	0.16
C#2 on	White male	-0.820	0.329	*-2.489*	0.01

C#2 on	White female	-0.531	0.330	-1.609	0.11
C#2 on	Other male	0.074	0.353	0.211	0.83
C#3 (*master*) on	Medium quality	-1.286	0.379	*-3.392*	0.00
C#3 on	Low quality	-1.782	0.364	*-4.896*	0.00
C#3 on	Black male	0.198	0.357	0.553	0.58
C#3 on	Black female	0.510	0.331	1.542	0.12
C#3 on	White male	-3.317	1.079	*-3.075*	0.00
C#3 on	White female	-2.510	1.043	*-2.407*	0.02
C#3 on	Other male	-0.291	0.396	-0.735	0.46
C#4 (*apprentice*) on	Medium quality	-0.897	0.286	*-3.141*	0.00
C#4 on	Low quality	-0.741	0.314	*-2.356*	0.02
C#4 on	Black male	-0.363	0.580	-0.626	0.53
C#4 on	Black female	-0.177	0.446	-0.398	0.69
C#4 on	White male	-1.156	0.457	*-2.532*	0.01
C#4 on	White female	-0.708	0.353	*-2.010*	0.04
C#4 on	Other male	0.021	0.460	0.046	0.96
Intercepts	C#2	1.404	0.687	*2.042*	0.04
Intercepts	C#3	1.344	0.780	1.723	0.09
Intercepts	C#4	1.535	0.464	*3.311*	0.00

Note. Values exceeding the critical value of 1.96 at $p<.05$ are italicized.

Next, logistic regression odds ratio results for each independent variable were considered. Table 21 indicates that White males were 3 times more likely to be in Class 1 (*novice*) than the reference class (*apprentice*) for each unit of increase in GOALS scores, and about one-tenth as likely as others to be in class 3 (*master*). Meanwhile, Black males and females were nearly twice as likely to be in class 3 (*master*) for each unit of increase in GOALS scores. Odds of greatest magnitude related to White males, followed by White females, and Black females and males. As observed earlier, variability was so high, that any conclusions would be tenuous at best.

With respect to quality of instruction, class 1 (*novice*) was twice as likely to receive *medium* or *low quality instruction*. Meanwhile, classes 2 and 3 were less likely than the reference class, class 4 (*apprentice*) and class 1 (*novice*), to receive *medium* and *low quality instruction*. For example, class 3 (*master*) was one-third as likely to receive *low quality instruction* and 68% as likely to receive *medium quality instruction* (in terms of implementation behavior and fidelity related to constructivist mathematics). Implications will be discussed in chapter five.

Table 21

Logistic Regression Odds Ratios, using Class 4 (Apprentice) as the Reference Class

Latent variables	**Predictor variable**	**Odds ratio exp.(*B*)**	**95% CI**
C#1 (*novice*) on	Medium quality instruction	2.453	1.401, 4.294
C#1 on	Low quality instruction	2.098	1.133, 3.885
C#1 on	Black male	1.438	0.461, 4.481
C#1 on	Black female	1.194	0.498, 2.861
C#1 on	White male	3.177	1.299, 7.775
C#1 on	White female	2.031	1.018, 4.053
C#1 on	Other male	0.979	0.398, 2.411
C#2 (*skilled*) on	Medium quality instruction	0.968	0.633, 1.479
C#2 on	Low quality instruction	0.747	0.273, 2.041
C#2 on	Black male	1.729	0.779, 3.838
C#2 on	Black female	1.908	0.984, 3.702
C#2 on	White male	1.400	0.638, 3.071
C#2 on	White female	1.194	0.637, 2.239
C#2 on	Other male	1.055	0.385, 2.886
C#3 (*master*) on	Medium quality instruction	0.678	0.435, 1.056
C#3 on	Low quality instruction	0.353	0.197, 0.633
C#3 on	Black male	1.752	0.816, 3.761
C#3 on	Black female	1.989	1.041, 3.801
C#3 on	White male	0.115	0.009, 1.508
C#3 on	White female	0.165	0.019, 1.447
C#3 on	Other male	0.732	0.383, 1.399

3. Compare present results to results from the original analyses.

Findings from George, Hall, and Uchiyama (2000) were reviewed in detail in chapter two. Major implications related to:

- patterns in Levels of Use,
- relationship between higher Levels of Use (LoU) and higher levels of student learning,
- higher achievement in classrooms most closely aligned with National Council of Teachers of Mathematics (NCTM) Standards (higher fidelity ratings on Innovation Configuration [IC] Maps), and
- apparent increased benefits from “high fidelity” implementation patterns by students with lower fall test scores.

In the present study, LoU and IC ratings were combined into a single quality rating, given the lack of data measuring both constructs, and gender and ethnicity were combined into a single variable with six categories. It was anticipated that high quality teaching implementation would be associated with higher levels of student learning, especially for students with lower fall test scores.

These results were partially verified through the present study, in the association of higher quality instruction with highest performing latent variable classes. The difference in effect of high quality instruction on students with initially lower performance did not surface, and results related to gender and ethnicities were not meaningful. It would appear that ANCOVA analyses examining student level data in greater depth were more sensitive to reflect differential effects on student outcomes. However, it is not known to what extent access to original, student-level data would have increased the acuity of this study. If a complete data set were available to complete two-level growth mixture modeling, these results would likely have become evident.

Power analyses. Post-hoc power analyses of original ANCOVA analyses related to student achievement and Levels of Use (LoU) and Innovation Configuration (IC) fidelity ratings, respectively, were performed using *G*Power* (Faul, Erdfelder, Buchner, & Lang, 1992-2001).

Original researchers found significant results for the impact of LoU on mean of classroom means on Spring GOALS (F=81.70, df=3,55, p<.01) and, in post-hoc comparisons of mean of classroom means between LoU III teachers and those with higher LoU, found a significant difference (F=32.80, df=3,54, p<.01 [George, Hall & Uchiyama, 2000, pp. 13-14]). The power of the GLM procedure related to LoU was 0.5477, critical F 1.033, 55 df (based on effect size of .25, total sample size 59, with 3,55 df, and 3 covariates). The power of the post-hoc test was 0.5484, critical F 1.033, 55 df (based on effect size of .25, total sample size 59, with 3,54 df, and 3 covariates).

With respect to the impact of fidelity/IC ratings on Spring GOALS mean of classroom means, after controlling for initial differences in Fall GOALS mean of classroom means, original researchers found "a significant difference between the slopes of the regression lines relating fall GOALS classroom averages to spring GOALS classroom averages (F=3.61, df=3,36, p=.02)," (George, Hall & Uchiyama, 2000, p. 35). The power of the ANCOVA was 0.5424 , critical F 1.035, 38 df (based on an effect size of .25, total sample size 42, with 3,36 df, and 3 covariates).

Original researchers carefully delimited results to participants in the study (a non-random sample, during the 1996-98 timeframe) and reported observed relationships between LoU, IC, and apparent impacts on student achievement, respectively. They did not claim causality, stating: “We can only look for associations, cause and effect are sometimes difficult to determine” (p. 23). They further noted small sample sizes and stated findings as “apparent..”

Power analyses in *Mplus* (Muthén & Muthén, 1998-2009) were not conducted, because they would require a separate Monte Carlo simulation, an extensive procedure that would represent an entirely separate study. It is therefore not known how the power of the *Latent Class Analysis* compares to the relatively low power of tests in the original analyses.

Summary

Quantitative research was conducted through structural equation modeling and latent class analysis. Results for four-class latent class analysis were reported and interpreted, and compared to findings from original research.

In chapter five, the researcher will provide a summary of the study, conclusions based on research, recommendations for future study, and implications from an educational leadership perspective.

CHAPTER FIVE: SUMMARY, CONCLUSIONS, RECOMMENDATIONS, AND IMPLICATIONS

Following a brief summary of the study, conclusions will be shared, followed by recommendations for practice and future research, and broader implications.

Summary of the Study

Summary of Chapters 1 through 3

In this confirmatory study, the researcher re-analyzed evaluation data collected from teachers and grades 2-8 students at District A of Department of Defense Dependents' Schools from 1996-1998, on the change process related to district-wide implementation of constructivist math curriculum. Data were collected using the Concerns-Based Adoption Model (Hall & Hord, 2006) to measure implementation behaviors (Levels of Use of the Innovation, LoU) and quality of implementation in the classroom (Innovation Configurations, IC) (Hall & Hord). Published findings from the original study (Hall, 1999, 2000; Thornton & West, 1999; Alquist & Hendrickson, 1999; Johnson, 2000; and George, Hall & Uchiyama, 2000) indicated a positive association between higher levels of implementation behaviors and fidelity in implementation with higher student achievement.

This study focused on modeling relationships among student outcomes, levels of implementation behavior, and fidelity of implementation of constructivist approaches to teaching mathematics. Results and methodology were then compared to the original study.

Hall and Hord's research on change processes, and the three individual constructs of the Concerns-Based Adoption Model (CBAM): Levels of Use (LoU), Stages of

Concern, and Innovation Configurations (IC) (2006) provided the foundation for Chapter 2. In the context of the original study, Hord clarified that "change is learning" (2000), and Hord and Sommers (2008, pp. 19-20) recommended the CBAM to evaluate professional learning processes related to transfer of learning.

Standards-based education, and particularly the National Council of Teachers of Mathematics (NCTM) *Principles and Standards for School Mathematics* (2001), provided background information on the Mathematics Standards and Expectancies (1994) implemented in District A during this study. These professional learning and change processes were evaluated by Hall and colleagues with respect to implementation behaviors (LoU) and quality of implementation (IC). Results were impressive related to district-wide implementation and increase in student learning (Hall, 1999, 2000; Thornton & West, 1999; Alquist & Hendrickson, 1999; Johnson, 2000; and George, Hall & Uchiyama, 2000).

In the present study, the researcher attempted to model relationships between latent variables *Change in Teaching Math* and *Change in Learning Math* with respect to student outcomes on pre- and post-tests. Covariance matrices and factor means and intercepts would then provide additional insight into aspects of implementation and student achievement. Results from structural equation modeling would then be compared to results from the original analyses, in hopes of identifying potential similarities and differences in statistical methodologies for viewing the dynamics of change in teaching and learning mathematics.

The study was conducted as proposed, and then "second generation structural equation modeling" (Muthén, 2001, p. 291) in the form of Latent Class Analysis was used to acquire results for interpretation and comparison. The two research questions were answered with respect to relationships among student outcomes, levels of implementation behavior, and fidelity of implementation on constructivist approaches to teaching mathematics, and comparison of results from this study to original analyses. An additional sub-question was answered, related to what impact a change in teaching quality from *low* to *medium* could have on student achievement. Findings, conclusions, recommendations for practice and future research, and implications will be organized as depicted in Figure 9.

Figure 10.

Logic Model for Presentation of Results to Research Question 1 and Additional Sub-question

Research Question	Findings	Conclusions	Recommendations Practice	 Research	Implications
Relationships among student outcomes, levels of implementation behavior, and fidelity of implementation of constructivist applications to teaching mathematics	1 Increased means	1 Impact of constructivist approaches to teaching math in District A	1 Curriculum		1 Change and professional learning for math instruction
	2 Impact of low quality instruction on GOALS	2 Relationship of low quality instruction to student achievement	2 Statistical modeling	2 SEM & 2nd gen. SEM	2 Statistical modeling 2-4 CBAM to measure professional learning
	3 Significantly positive impact of high quality instruction	3a Relationship of high quality instruction to student achievement 3b Impact of closer alignment with NCTM standards	3-4, 6-7 Professional learning focus	3 NCTM standards	2-3 Teaching quality for student learning and achievement
	4 Recipients of low quality instruction	4 Relationship of quality of instruction with mastery of learning			
	5 Gender and ethnicity	5 No conclusion		5 Replication & extension	
Comparison of student outcomes from low quality and medium quality instruction	6 Impact of medium quality on GOALS	6 Impact of movement from low to medium quality instruction	6-7 Professional learning focus (3-4 above)		6-8a Teaching quality for student learning and achievement (2-3 above)
	7 Significantly negative impact of low & medium quality instruction	7 Impact of med. and low quality on mastery			6-8b CBAM to measure teaching quality as transfer of professional learning (2-4 above)
	8 Recipients of medium or low quality instruction	8 Differences in teacher assignment		8 Extension (5 above)	

Summary of Findings

Relationships among student outcomes, levels of implementation behavior, and fidelity of implementation of constructivist approaches to teaching mathematics. Levels of implementation behavior and fidelity of implementation were combined into a single *quality of instruction* (*quality*) variable. Higher *quality* was associated with higher student outcomes, while lower *quality* was associated with lower student outcomes, as measured by standards-based examinations.

As detailed in chapter four, there were five findings, plus three related to an additional sub-question.

1. *Increased means*. Mean scores increased from fall to spring in every latent class (pp. 72-73). With the covariate effect of fall scores removed, gains in resulting spring GOALS scores were significantly different than 0. Confidence intervals from fall and spring were then examined for each latent class to determine whether growth was statistically significant and very unlikely to be caused by chance. *Skilled* and *master* classes indicated significant growth demonstrated (conservatively) by lack of overlap between confidence intervals. A positive treatment effect of instruction was therefore inferred.
2. *Impact of low quality instruction on GOALS*. As mentioned in chapter 4 (pp. 75-76 and 81-83), *low quality instruction* was associated with lower spring GOALS scores in each of the four latent classes. The standardized difference in spring GOALS scores was -2.047 compared to *medium quality instruction*, and -2.126 compared to *high quality instruction*.

3. *Significantly positive impact of high quality instruction*. As presented on pages 75-76, using class 2 (*novice*) and *medium quality instruction* as a reference, there were statistically significantly positive effects of *high quality instruction* on spring GOALS scores in classes 1, 3, and 4 by the following standardized amounts: 3.995 on class 1 (*skilled*), 3.141 on class 3 (*apprentice*), and 3.392 on class 4 (*master*).
4. *Recipients of low quality instruction*. As stated on page 76, "With respect to *quality of instruction*, classes 1-3 were all less likely than the reference class, class 4 (*master*), to receive *high quality instruction*, and more likely to receive *low quality instruction*. For example, class 2 (*novice*) was one-quarter as likely to receive high quality instruction and 1.6 times more likely to receive *low quality instruction* (in terms of teacher implementation behavior and quality related to constructivist mathematics)."
5. *Gender and ethnicity*. Odds ratios were examined to determine whether gender and ethnicity could be clearly associated with various latent classes. Variability was so high; any conclusions would be tenuous at best.

Additional sub-question: Comparison of student outcomes from low quality and medium quality instruction

6. *Impact of medium quality on GOALS*. As shown in Table 19 on page 80, the impact of *medium quality instruction* was not significant to all latent classes, and did not reach the critical value of 1.96 or *p* value of .05. One can state with 95% confidence that *medium quality instruction* impacted spring

GOALS scores in all latent classes from -0.526 to 0.865, compared to a significantly negative impact of -2.126, $p < .03$ of *low quality instruction*.

7. *Significantly negative impact of low and medium quality instruction*. As stated on page 81, using class 1 (*novice*) and *high quality of instruction* as a reference, there were statistically significantly negative effects of *medium* and *low quality instruction* in classes 2 (*skilled*), 3 (*master*), and 4 (*apprentice*). The standardized difference on spring GOALS scores of receiving *medium*, rather than *high*, *quality instruction* was -3.995 (*skilled*), -3.392 (*master*), and -3.141 (*apprentice*), respectively. The standardized difference of receiving *low* rather than *high quality instruction* was -2.779 (*skilled*), -4.896 (*master*), and -2.356 (*apprentice*).
8. *Recipients of medium or low quality instruction*. As stated on page 83, "With respect to quality of instruction, class 1 (*novice*) was twice as likely to receive *medium* or *low quality instruction*. Meanwhile, classes 2 and 3 were less likely than the reference class, class 4 (*apprentice*) and class 1 (*novice*), to receive *medium* and *low quality instruction*. For example, class 3 (*master*) was one-third as likely to receive *low quality instruction* and 68% as likely to receive *medium quality instruction* (in terms of implementation behavior and quality related to constructivist mathematics)."

Comparison of results from this study to original analyses. Findings were consistent with those of George, Hall, and Uchiyama (2000) that higher quality instruction (a combination of higher Levels of Use and higher fidelity ratings on IC Maps—instruction more closely aligned with NCTM Standards) was related to higher levels of student learning. Unlike the original analysis, increased benefits by students with lower fall test scores who then received “high fidelity” teaching, were not observed. Post-hoc power analyses of original ANCOVA and GLM statistical tests indicated relatively low power, 0.5424, 0.5477, and 0.5484 respectively. This would indicate an approximate 45% chance of making a Type I error (falsely rejecting the null hypothesis).

Summary of Conclusions

Relationships among student outcomes, levels of implementation behavior, and fidelity of implementation of constructivist approaches to teaching mathematics. In District A, higher quality constructivist mathematics instruction was related to higher levels of mastery in learning mathematics.

The following conclusions were reached:

1. *Impact of constructivist approaches to teaching math in District A.* Constructivist approaches to teaching mathematics in this District were effective to increase student learning each year in each of the four latent classes. Gains were most significant in the two classes indicating greatest mastery of learning, *skilled* and *master*.
2. *Relationship of low quality instruction to student achievement*. Student achievement on the standardized spring GOALS exams among students who received *low quality instruction* was lower than those who received *medium quality instruction* and significantly lower than those who received *high quality instruction*. Therefore one can conclude that *low quality instruction* as defined in this study had a significantly negative impact on student learning of

mathematics, compared to the positive learning outcomes related to *medium* and *high quality instruction*.

3. a. *Relationship of high quality instruction to student achievement*. Conversely, *high quality instruction* positively impacted student learning, with the greatest impact on students in the *skilled* mastery of learning latent class.

   b. *Impact of closer alignment with NCTM standards*. In this study, *high quality instruction* was defined by higher fidelity of implementation behaviors to the *Innovation Configuration* in Appendix A, and higher levels of implementation behaviors defined by *Levels of Use* ratings of IVb and V. These ratings coincide directly with closer adherence to National Council of Teachers of Mathematics (NCTM) Standards (1989). Therefore, closer alignment of instruction with NCTM Standards (aka *high quality instruction*) in District A resulted in higher student learning and achievement.

4. *Relationship of high quality instruction with mastery of learning*. The four latent classes in this study indicate progressive levels of mastery of learning: *novice*, *apprentice*, *skilled*, and *master* (defined on p. 65), based on pre- and post- scores on a performance-based assessment. *High quality instruction* in District A made the largest difference in mastery of learning for all students, in the following order: *skilled*, *master*, *apprentice*, and then *novice* classes. Predominant assignment of teachers with lower quality implementation patterns to students with lower levels of mastery in learning may obscure understanding of the impact that *high quality instruction* would have on *novice* and *apprentice* learners.

5. *No conclusion*. There was too much variability in student outcomes by ethnicity and gender to make any conclusion about differential effects that may exist related to gender and ethnicity.

Additional sub-question: Comparison of student outcomes from *low quality* and *medium quality instruction*

6. *Impact of movement from low to medium quality instruction*. Positive movement along the spectrum of professional growth from *low* to *medium quality* was beneficial to student learning, though the estimated effect was small.
7. *Impact of medium and low quality instruction on mastery of learning*. *Low quality instruction* was most damaging to mastery of learning among students in the *master* class. *Medium quality instruction* was less effective than *low quality* to promote mastery of learning among *skilled* and *apprentice* latent classes.
8. *Differences in teacher assignment*. Students with the lowest initial status were far more likely to receive *low quality instruction* than students with higher initial status. *Medium quality instruction* had more positive effects on students in the lower two latent classes than *low quality instruction*, though the *novice* class was least impacted by quality of instruction. It is unclear whether student mastery of learning is reinforced or could be caused by such teacher assignment patterns. Though anything less than *high quality instruction* negatively impacted student learning gains, students in the *master* class were most negatively impacted by *medium*, rather than *low quality instruction*. As

mentioned before, predominant assignment of teachers with lower quality implementation patterns to students with lower levels of mastery in learning may obscure understanding of the impact that *high quality instruction* would have on *novice* and *apprentice* learners, as well as the full impact that *low quality instruction* would have on *skilled* and *master* latent classes of learners.

Comparison of results from this study to original analyses. ANCOVA analyses were more sensitive to differential effects on individual student outcomes than Latent Class Analysis under the circumstances in which identifiable student-level data were not available. Given the unavailability of comparable data, it was not possible to reach further conclusions. However, post-hoc power analyses of original studies indicated a high risk for Type I error, based on relatively low power of statistical tests.

Recommendations

Recommendations for Practice

School administrators are encouraged to consider the following three recommendations for immediate, practical use:

1. *Curriculum*. Constructivist approaches to teaching mathematics, specifically those detailed in the Innovation Configuration in Appendix A, are recommended for use with students approximating *skilled* and *master* levels of learning mastery in elementary and middle school mathematics.
2. *Use statistical modeling to gain insights on the dynamics of teaching and learning*. Many statistical modeling approaches exist to aid school administrators in understanding the dynamics of change. Systematic collection, analysis, and study of

professional development and student achievement data is highly recommended for decision making as an instructional leader.

3. *Professional learning focus*. The differential effects of implementation behaviors and fidelity on student achievement in mathematics are very important. A coordinated long-term approach to professional learning is recommended in order to support teachers in the attainment of higher instructional quality. In this study, *low quality instruction* was defined by mechanical implementation behaviors (LoU III), and a predominance of lower fidelity instructional actions, based on an explicitly defined Innovation Configuration.

   Professional learning must focus on the ideal configuration, rather than perfunctory behavior. As teachers experience higher levels of professional learning, they appear to become more effective at assisting their students to learn. Mechanical behavior in implementation of an instructional approach should be viewed as a step in a professional learning continuum, rather than the destination, in order to facilitate higher levels of student learning and achievement.

Recommendations for Future Research

1. *Structural equation modeling*. SEM provides an appealing context for studies of change when continuous variables are of interest. Conducting a study based on the proposed model using class means and separate SEMs to compare specific sub-populations, such as males and females, would likely yield interesting results. Both well-specified and complete data are needed for such a study to be successful.
2. *Second-generation SEM*.

a. Replication of the present study with longitudinal student-level and teacher-level data with at least four measurement points would be ideal using a multi-level Growth Mixture Modeling. This would yield clearer results that account for students within classrooms, and teacher effects on classrooms as a whole. Data sets greater than 60 for each variable are recommended. This technique, combined with more robust data collection (e.g., for each measurement each year), as well as inclusion of data regarding student characteristics (i.e., poverty, gender, ethnicity, and number of days absent), and more complete description of teachers' number of hours being coached, would more adequately mitigate the influence of confounding variables.

b. Second-generation SEM techniques (Muthén, 2001) seem ideal for modeling the three independent constructs of the CBAM (Hall & Hord, 2006) along with student achievement data for a specific instructional innovation in a multivariate environment. Additional theory of change, professional learning, and the measurement of change with respect to impact on student achievement may emerge. According to Drs. Hall, Hord, and George, they know of no studies in which the three elements of the CBAM were modeled in a multivariate environment using a structural equation model.

3. *Alignment of mathematics instruction with NCTM standards*. In District A, mathematics instruction more closely aligned with National Council of Teachers of Mathematics (NCTM) standards (1989) was associated with higher student learning. Administrators and policy makers are encouraged to further research and examine the

results of closely aligning mathematics instruction with NCTM standards for increased student learning in local and national contexts.

4. *Extension of the original study in other school districts*.
    a. Replication of the original study, with the addition of collecting Stages of Concern (Hall & Hord, 2006) data is recommended. It is not known to what extent aspects of curricular implementation and professional learning may differ between DoDDS districts such as District A and non-DoDDS districts. Cross-case analysis (Yin, 1994) may provide insights into the applicability of results from DoDDS districts to non-DoDDS contexts.
    b. Furthermore, the implementation in District A resulted in Levels of Use (LoU) III and higher among all teachers. Comparing and contrasting impacts on student achievement of LoU 0-II to those of LoU III and above could provide additional information for teachers, professional developers, school and district administrators, researchers, and policy makers on the nuances of studying and evaluating effectiveness of professional development in schools and districts.
5. *Extension in different educational contexts*. Modeling associations between change in teaching and change in learning in various educational contexts is also recommended. For example, it would be interesting to understand whether impacts of teacher quality are different in higher education with adult learners than in a K-12 context. Comparative studies among elementary, middle and high schools, and higher education may increase understanding of the impacts of differing quality of instruction upon student learning at different stages in the learning continuum.

6. *Comparative power analyses among statistical tests*. A comparison of statistical power of ANCOVA, GLM, SEM, and second-generation SEM approaches, with minimum sample and group sizes required to reach power levels of .8 or greater could greatly contribute to research related to education and educational leadership.

Implications

There are several implications from this study to consider.

1. *Change and professional learning for mathematics instruction*. Much is known about change processes and how to provide support to bring about successful implementation over time. District-wide implementation of change in mathematical instruction **can** succeed, as the original study demonstrated, but change only occurs if teachers **implement instructional practices** in their classrooms **with fidelity** to the intended ideal configuration. Views of professional learning must extend beyond attendance at a class or series of coaching events to implementation in the classroom with increasing quality over time. In District A, this was accomplished through a systematic, comprehensive, long-term, district-wide approach. Rarely does district-wide curricular adoption result in 100% of teachers demonstrating Levels of Use III or greater implementation behaviors by the end of the second year. It could be inferred that results reflected the quality of the overall implementation process as well as the quality of teachers' professional and students' individual learning.
2. *Statistical modeling can provide insights to guide instructional leadership*. ANCOVA, SEM, and second-generation SEM can provide unique windows into the dynamics of complex relationships, including those related to teaching and learning. School and district leaders and policy makers would be well served by seeking

statistical research to inform their decisions on behalf of teachers and students. It is critical, however, that assumptions are carefully reviewed and data are scrupulously interpreted to avoid irresponsible conclusions. Such leaders would benefit from studying advanced multivariate statistics, in order to make informed decisions that would result in fulfillment of national priorities to improve teaching and learning of mathematics.

3. *Use the CBAM (Hall & Hord, 2006) for measurement of professional learning, evaluation, and verification of research-based practices*. The CBAM (Hall & Hord, 2006) may be viewed as a set of concepts about change, each of which has a measure to infer the impact of professional learning and transfer of training. This study demonstrates the value of the CBAM for measurement of professional learning, and as a tool to distinguish between levels of implementation and fidelity, to better evaluate the impact of instructional innovations on student learning.
4. *High quality mathematics instruction is associated with higher student achievement*. *Higher quality of instruction* (a combination of implementation behaviors and fidelity of implementation to the Innovation Configuration Map) of constructivist approaches to teaching mathematics in District A was associated with higher levels of student learning and achievement, and *lower quality of instruction* was associated with lower student learning and achievement in mathematics in District A. The implication is that professional learning and implementation of research-verified innovations are crucial to improving student learning and achievement in mathematics.

Summary

The study of change in professional learning, in order to increase student achievement, is a worthy endeavor. Findings and conclusions were summarized related to the positive impact of constructivist approaches to teaching mathematics in District A, the damaging effects of *low quality instruction* on student achievement and the beneficial relationship of *high quality instruction*—which related to closer alignment with NCTM standards (1989)—to student achievement and mastery of learning. Positive results were inferred from movement from *low* to *medium quality instruction* related to student learning and achievement, though patterns in teacher assignment indicated differential effects of *quality of instruction* among the four latent classes of learners. ANCOVA analyses from the original study provided additional insights related to individual students that could not be detected using the current methodology.

Recommendations for practice in school administration related to inclusion of constructivist approaches to teaching mathematics for students with greater levels of mastery in mathematics; practical uses of statistical modeling for the oversight of teaching and learning; and a coordinated, long-term approach to professional learning focused on ideal implementation.

Future research was recommended using first- and second-generation structural equation modeling to replicate and extend techniques from the present study to various education contexts, and further study the efficacy of mathematics instruction aligned with NCTM standards (1989). Extension of the original study to other school districts, including those outside of the Department of Defense Dependents' Schools context, was recommended, as well as extension to different educational levels, comparing results from higher education to those garnered from K-12 contexts. Finally, research related to sample sizes needed to attain statistical power of .8 or greater is recommended.

The implications related first to district-wide change and professional learning for mathematics instruction—it can be successful. Measurement of success relates to the extent of implementation and fidelity of implementation that result in improved student achievement, rather than a series of events. In the end, this boils down to the effectiveness of an implementation process that facilitated both teachers' professional and students' individual learning. Statistical modeling is one way to monitor and evaluate outcomes, in tandem with the tools embedded in the Concerns-Based Adoption Model (Hall & Hord, 2006). Finally, and most important, *high quality mathematics instruction* is associated with higher student achievement. These important implications can assist in heightened global competitiveness as a result of increased student learning and achievement.

REFERENCES

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

Alquist, A., & Hendrickson, M. (1999). Mapping the configurations of mathematics teaching. *Journal of Classroom Interaction*. (*34*)1, 18-26.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling*, 243-277. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Arbuckle, J. L. (2006). AMOS 7.0.0. (Build 1140). Spring House, PA: Amos Development Corp.

Arbuckle, J. L. (2007). *AMOS 16.0 user's guide*. Spring House, PA: Amos Development Corp.

Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models. *Journal of Marketing Research*, *19*, 425-430.

Bentler, P. M. (1988). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

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

Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, *34*(2), 181-197.

Blunch, N. J. (2008). *Introduction to structural equation modeling using SPSS and AMOS*. Thousand Oaks, CA: Sage Publications.

Brewer, J., & Hunter, A. (1989). *Multimethod research: A synthesis of styles*. Newbury Park, CA: Sage Publications.

Brown, R. L. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling*, *1*, 287-316.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.). *Testing Structural Models*. Newbury Park, CA: Sage Publications.

Bush, G. W. (2006). Presidential executive order 13398 of April 18, 2006: National mathematics advisory panel. *Federal Register*, *Vol. 71*, No. 77, 20519-20521. Retrieved September 27, 2008 from http://www.ed.gov/about/bdscomm/list/mathpanel/report/appendix-a.pdf.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Taylor & Francis Group.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Thousand Oakes, CA: Sage.

Charters, Jr., W. W., & Jones, J. E. (Nov. 1973). On the risk of appraising non-events in program evaluation. *Educational Researcher*, *2*(11), 5-7. American Educational Research Association. Retrieved November 11, 2008 from http://www.jstor.org/stable/1175174.

Cohen, D. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational evaluation and policy analysis*, *12*, 327-345.

Department of Defense Dependent Schools (1994). Mathematics standards and expectancies, *DS Manual 2320.1*.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling, 8*(3), 430-457. Lawrence Erlbaum Associates, Inc.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (1992-2009). G*Power (Version 3.1.2). Kiel, Germany: University of Kiel.

Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal*, *6*(2), 207-236.

Fuller, F. F. (1970). *Personalized education for teachers: An introduction for teacher educators*. Austin: The University of Texas at Austin, Research and Development Center for Teacher Education.

Fuller, F. F., & Brown, O. H. (1975). *Becoming a teacher: Teacher education 1975*. Chicago: The National Society for the Study of Education.

George, A. A., Hall, G. E., & Uchiyama, K. (2000). Extent of implementation of a standards-based approach to teaching mathematics and student outcomes. *Journal of classroom interaction*. (*35*)1, 8-25.

Hall, G. E. (1999). Using constructs and techniques from research to facilitate and assess implementation of an innovative mathematics curriculum. *Journal of Classroom Interaction*. (*34*)1, 1-8.

Hall, G. E. (2000). Evaluation of systematic approach to changing classroom practices for the teaching of mathematics. *Journal of Classroom Interaction*. (*35*)1, i-ii.

Hall, G. E., & Hord, S. M. (2001). *Implementing change: Patterns, principles, and potholes*. Boston, MA: Allyn & Bacon.

Hall, G. E., & Hord, S. M. (2006). *Implementing change: Patterns, principles, and potholes* (2nd ed.). Boston, MA: Allyn & Bacon.

Hall, G. E., & Loucks, S. F. (1975). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal*, *14*(3), 263-276.

Hall, G. E., Loucks, S. F., Rutherford, W. L., & Newlove, B. W. (1975). Levels of use of the innovation: A framework for analyzing innovation adoption. *Journal of Teacher Education*, *26*(1), 52-56

Heck, S., Stiegelbauer, S. M., Hall, G. E., & Loucks, S. F. (1981). *Measuring innovation configurations: Procedures and applications* (Report No. 3108). Austin: The University of Texas at Austin, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 204 147).

Hord, S. M., Rutherford, W. L., Huling, L., & Hall, G. E. (2006). *Taking charge of change*. Austin, TX: Southwest Educational Development Laboratory.

Hord, S. M. (n.d.) Steps in developing an IC matrix, *Creating an innovation configuration*. Austin, TX: Southwest Educational Development Laboratory.

Hord, S. M. (2000). Change is learning: It's as simple and complicated as that. *Journal of Classroom Interaction*. (*35*)1, 26-27.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons, Inc.

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

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

Johnson, M. H. (2000). A district-wide agenda to improve teaching and learning in mathematics. *Journal of Classroom Interaction*. (*35*)1, 1-7.

Jöreskog, K. G., & Sörbom, D. (1993). PRELIS 2 user's reference guide [Computer software]. Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1999). LISREL 8.3 [Computer software]. Chicago: Scientific Software International.

Joyce, B. R., & Showers, J. (2002). *Student achievement through staff development*, 3rd ed. Alexandria, VA: Association for Supervision and Curriculum Development.

Kilpatrick, J. (2000). Reflections on verifying change in school mathematics. *Journal of Classroom Interaction*. (*35*)1, 28-30.

Kilpatrick, J., Hancock, L., Mewborn, D. S., & Stallings, L. (1996). Teaching and learning cross-country mathematics: A story of innovation in precalculus. In S. A. Raizen & E. D. Britton (Eds.), *Bold ventures*. *Vol. 3: Case studies of innovations in U. S. innovations in mathematics education*, 133-243. Dordrecht: Kluwer Academic Publishers.

Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology*, *51*, 345-375.

Loucks, S. F., Newlove, B. W., & Hall, G. E. (1975). *Measuring levels of use of the innovation: A manual for trainers, interviewers, and raters*. Austin, TX: Southwest Educational Development Laboratory.

MacCallum, R. C., Brown, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Marzano, J. S., & Pickering, D. J. (2003). *Classroom management that works: Research-based strategies for every teacher*. Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mertler, C. A., & Vannatta, R. A. (2005). *Advanced and multivariate statistical methods: Practical application and interpretation* (3rd ed.). Glendale, CA: Pyrczak Publishing.

Miles, M., & Huberman, M. (1984). *Qualitative data analysis: A source book for new methods*. Beverly Hills, CA: Sage Publications.

Montopoli, G. (2007). *Course content: EPS 725*. Unpublished instructor notes distributed in class October 19, 2007. Northern Arizona University, Yuma, AZ.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In Collins, L. M. & Sayer, A. (Eds.), *New methods for the analysis of change*, *82*: 291-322. Washington, D.C.: APA.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431-462.

Muthén, B. O. & Muthén, L. K. (1998-2009). Mplus. (Version 5.2.1, base program and combination add-on, 32-bit). Los Angeles: Muthén & Muthén.

Muthén, L. K. & Muthén, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Muthén & Muthén.

National Center for Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, D.C.: U.S. Department of Education. Retrieved November 2, 2008 from http://www.ed.gov/pubs/NatAtRisk/risk.html.

National Council of Teachers of Mathematics (2001). *Executive summary: Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics. Retrieved November 2, 2008 from http://www.nctm.org/uploadedFiles/Math_Standards/12752_exec_pssm.pdf

Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Boston, MA: McGraw-Hill.

Nylund, K. L. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization*. Los Angeles, CA: University of California Los Angeles.

Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows* (3rd ed.). New York: McGraw-Hill Education.

Pike, G. R., & Kuh, G. D. (2005). First- and second- generation college students: A comparison of their engagement and intellectual development. *The Journal of Higher Education, 76*(3), 276-300.

Public law 103-227 (3-31-94). *The goals 2000: Educate America act*. Retrieved November 2, 2008 from http://www.ed.gov/legislation/GOALS2000/TheAct/index.html.

Public law 107-110 (1-08-02). *The no child left behind act of 2001*. Retrieved November 2, 2008 from http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf.

Rogers, E. M. (1971). *Diffusion of innovations*. New York: Free Press.

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

Rogers, E. M. & Scott, K. L. (1997). *The diffusion of innovations model and outreach from the National Network of Libraries of Medicine to Native American Communities*. Seattle, WA: National Network of Libraries of Medicine. Retrieved November 2, 2008 from http://nnlm.gov/archive/pnr/eval/rogers.html.

SAS Institute, Inc. (1992). *SAS/STAT user's guide*, release 6.03. Cary, NC: SAS Institute, Inc.

Schumacker, R. E. & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). New York: Taylor & Francis Group.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464.

Spring, J. (2005a). *Conflict of interests: The politics of American education* (5th ed.). New York: McGraw-Hill.

Spring, J. (2005b). *The American school: 1642-2004* (6th ed.). New York: McGraw-Hill.

SPSS, Inc. (2006). *SPSS 15.0 for Windows graduate student version*.

Tabachnick, B. G. & Fidell, L. S. (Eds.). (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education, Inc.

Thornton, E. & West, C. E. (1999). Extent of teacher use of mathematics curriculum innovation in one district: Years 1 and 2 levels of use (LoU). *Journal of Classroom Interaction*. (*34*)1, 9-17.

Ullman, J. B. (2007). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (5th ed.), pp. 676-780. Boston, MA: Pearson Education.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Wothke, W. (2000). Longintudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples*, 219-240. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

APPENDIX A

Figure 10. Sample Components from the IC Map for Teaching and Learning Mathematics

1) <u>Teacher Poses Mathematical Tasks/Investigations</u> {poses, frequency, open-ended questions, language}

a	b	c	d	e
Teacher poses open-ended problem, highlights mathematical aspects and asks students to determine how to figure them out. Open-ended questions are used to pose problems, not only at the beginning but also throughout the lesson. Teacher uses mathematical language to present tasks/investigations .	Teacher identifies mathematical aspects of tasks/ investigations and explains how to figure them out. Teacher directions are clear. Some mathematical language is used. Some open-ended questions are asked.	The teacher presents the activity with little or no explanation. Teacher uses little or no mathematical language. Some teacher directions are clear. Nearly all questions require one-word answers.	Teacher structures activity and directs students' activity. Questions requiring one-word answers are used to check for student understanding. Isolated use of math vocabulary. .	Teacher presents/explains concept or procedure and assigns individual student work. Questions requiring one-word answers are used to check for student understanding. Isolated use of math vocabulary.

4) <u>Teacher helps students in making connections</u> {making connections among mathematical topics and/or other subject areas}

a	b	c	d
The teacher guides the students in making connections within the discipline of mathematics and/or to other subject areas. The teacher elicits connections from the students based on the context of the lesson or investigation.	The teacher tends to state the mathematical and/or other subject area connections. Teacher elicitation of connections from the students is minimal.	The teacher states only the mathematical connections in the lesson or investigation. The teacher makes no attempt to elicit connections from the students.	The teacher makes no attempt to communicate the mathematical connections in a lesson or investigation.

6) Students Engaged in Mathematical Tasks throughout the Lesson {engagement, time}

a	b	c	d
Most students are engaged in mathematical tasks, most of the time.	Most students are engaged in mathematical tasks, part of the time.	Some students are engaged in mathematical tasks. Many are off task most of the time.	Few students are engaged any of the time.

7) Students' Understanding of Problem Solving Strategies {knowing your goal, knowing where you are now, knowing the steps to get to the goal, reflection}

a	b	c	d	e
Students view the open-ended problem as a whole and analyze its parts. They create, select, and test a range of strategies. Students reflect upon the reasonableness of the strategies and the solution.	Students grasp the open-ended problem as a whole and analyze its parts. Students pick an established/traditional strategy to try to solve the problem, which is applied without considering alternatives. Students reflect upon the reasonableness of the solution but not the strategy.	Students approach the open-ended problem as a whole but do not have a clear understanding of the parts. The primary focus is on getting an answer. The students' reflection is on whether the answer is right rather than the reasonableness of the strategy.	Students approach open-ended problems as unconnected/ unrelated parts and do not see the problem as a whole. Students may manipulate materials and numbers, but are not clear about the reason/ purpose. If observable, reflection is about procedures.	Students calculate and compute using rote and routine procedures. Students are not clear about the final goal or the relationship of the tasks to that goal. There is little or no reflection about what is being learned.

14) Teacher Use of Visual Displays and Tools

Check all that apply.

Manipulative Materials	**Use of Technology**	**Other Resources**
_____ Organized	_____ Calculators	_____ Games
_____ Accessible to students during the lesson	_____ Computer	_____ Puzzles
_____ Accessible to students during the day	_____ Overhead	_____ Math dictionaries
	_____ Other	_____ Other

Note. From "Mapping the Configurations of Mathematics Teaching," by A. Alquist and M. Hendrickson, 1999, *Journal of Classroom Interaction, 34*, pp. 23-24.

APPENDIX B

Statistics

		goalf	goals	egen	gr	tchrid	Stuid	year
N	Valid	2138	2138	2138	2138	2138	2138	2138
	Missing	0	0	0	0	0	0	0

Note. Goalf = fall GOALS, goals = spring GOALS, egen = ethnic-gender, gr = grade, stuid = student ID, year = year

Egen

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	11	556	26.0	26.0	26.0
	12	463	21.7	21.7	47.7
	21	291	13.6	13.6	61.3
	22	268	12.5	12.5	73.8
	31	283	13.2	13.2	87.0
	32	277	13.0	13.0	100.0
	Total	2138	100.0	100.0	

Note. 11=Black Male, 12=Black Female, 21=White Male, 22=White Female, 31=Other Male, 32=Other Female

Grade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1021	47.8	47.8	47.8
	2	1117	52.2	52.2	100.0
	Total	2138	100.0	100.0	

Note. 1=Elementary, grades 2-5; 2=Middle School, grades 6-8.

Year

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	9697	1135	53.1	53.1	53.1
	9798	1003	46.9	46.9	100.0
	Total	2138	100.0	100.0	

Quality

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	796	37.2	37.2	37.2
	2	896	41.9	41.9	79.1
	3	446	20.9	20.9	100.0
	Total	2138	100.0	100.0	

Note. 1=High, 2=Medium, 3=Low

APPENDIX B - Continued

Goalf

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2	.1	.1	.1
	1	13	.6	.6	.7
	2	15	.7	.7	1.4
	3	26	1.2	1.2	2.6
	4	56	2.6	2.6	5.2
	5	61	2.9	2.9	8.1
	6	68	3.2	3.2	11.3
	7	90	4.2	4.2	15.5
	8	113	5.3	5.3	20.8
	9	114	5.3	5.3	26.1
	10	137	6.4	6.4	32.5
	11	138	6.5	6.5	39.0
	12	130	6.1	6.1	45.0
	13	143	6.7	6.7	51.7
	14	117	5.5	5.5	57.2
	15	135	6.3	6.3	63.5
	16	127	5.9	5.9	69.5
	17	112	5.2	5.2	74.7
	18	109	5.1	5.1	79.8
	19	80	3.7	3.7	83.5
	20	80	3.7	3.7	87.3
	21	56	2.6	2.6	89.9
	22	48	2.2	2.2	92.1
	23	46	2.2	2.2	94.3
	24	44	2.1	2.1	96.4
	25	24	1.1	1.1	97.5
	26	27	1.3	1.3	98.7
	27	13	.6	.6	99.3
	28	7	.3	.3	99.7
	29	3	.1	.1	99.8
	30	4	.2	.2	100.0
	Total	2138	100.0	100.0	

APPENDIX B - Continued

Goals

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	.0	.0	.0
	2	3	.1	.1	.2
	3	10	.5	.5	.7
	4	16	.7	.7	1.4
	5	26	1.2	1.2	2.6
	6	35	1.6	1.6	4.3
	7	46	2.2	2.2	6.4
	8	44	2.1	2.1	8.5
	9	72	3.4	3.4	11.8
	10	50	2.3	2.3	14.2
	11	70	3.3	3.3	17.4
	12	72	3.4	3.4	20.8
	13	86	4.0	4.0	24.8
	14	87	4.1	4.1	28.9
	15	90	4.2	4.2	33.1
	16	93	4.3	4.3	37.5
	17	110	5.1	5.1	42.6
	18	117	5.5	5.5	48.1
	19	126	5.9	5.9	54.0
	20	127	5.9	5.9	59.9
	21	113	5.3	5.3	65.2
	22	108	5.1	5.1	70.3
	23	102	4.8	4.8	75.0
	24	118	5.5	5.5	80.5
	25	102	4.8	4.8	85.3
	26	77	3.6	3.6	88.9
	27	69	3.2	3.2	92.1
	28	67	3.1	3.1	95.3
	29	68	3.2	3.2	98.5
	30	33	1.5	1.5	100.0
	Total	2138	100.0	100.0	

APPENDIX C

Mplus input for LCA model without covariates, p. 55

```
TITLE:
Change in Learning LCA model

DATA:
FILE = chgtl_cat7trim.dat;

VARIABLE:
NAMES ARE goalf goals egen gr stu year tchr qual H M L c1 c2 ms u1-u16;
USEVAR = goalf goals;
CLASSES = c(4);

DEFINE:
if (egen eq 11) then bm=1;
if (egen ne 11) then bm=0;
if (egen eq 12) then bf=1;
if (egen ne 12) then bf=0;
if (egen eq 21) then wm=1;
if (egen ne 21) then wm=0;
if (egen eq 22) then wf=1;
if (egen ne 22) then wf=0;
if (egen eq 31) then om=1;
if (egen ne 31) then om=0;

ANALYSIS:
TYPE = mixture;

OUTPUT:
TECH1 TECH8 modindices(all);
```

APPENDIX D

Mplus input for 4-class LCA model with covariates, p. 56

```
TITLE:
Change in Learning LCA model

DATA:
FILE = chgtl_cat7trim.dat;

VARIABLE:
NAMES ARE goalf goals egen gr stu year tchr qual H M L c1 c2 ms u1-u16;
USEVAR = goalf goals h l bm bf wm wf om;
CLASSES = c(4);

DEFINE:
if (egen eq 11) then bm=1;
if (egen ne 11) then bm=0;
if (egen eq 12) then bf=1;
if (egen ne 12) then bf=0;
if (egen eq 21) then wm=1;
if (egen ne 21) then wm=0;
if (egen eq 22) then wf=1;
if (egen ne 22) then wf=0;
if (egen eq 31) then om=1;
if (egen ne 31) then om=0;

ANALYSIS:
TYPE = mixture;
STARTS = 50 5;

MODEL:
%OVERALL%
c on h l bm bf wm wf om;
goals on h l;

OUTPUT:
tech1;

PLOT:
plot3;
```

APPENDIX E

Mplus input for 2-level GMM model with covariates, p. 56

```
TITLE:
Change in Learning 2-level GMM model
DATA:
FILE = chgtl_cat7trim.dat;
VARIABLE:
NAMES ARE goalf goals egen gr stu year tchr qual H M L c1 c2 ms u1-u16;
USEVAR = goalf goals h m bm bf wm wf om;
CLASSES = c(3);
WITHIN = bm bf wm wf om;
BETWEEN = h m;
CLUSTER = tchr;
DEFINE:
if (egen eq 11) then bm=1;
if (egen ne 11) then bm=0;
if (egen eq 12) then bf=1;
if (egen ne 12) then bf=0;
if (egen eq 21) then wm=1;
if (egen ne 21) then wm=0;
if (egen eq 22) then wf=1;
if (egen ne 22) then wf=0;
if (egen eq 31) then om=1;
if (egen ne 31) then om=0;
ANALYSIS:
TYPE = twolevel mixture;
STARTS = 0;
MODEL:
%WITHIN%
%OVERALL%
iw | goalf@0 goals@1;
c on bm bf wm wf om;
%BETWEEN%
%OVERALL%
ib | goalf@0 goals@1;
ib on h m;
c#1 on h m;
c#2 on h m;
c#1*1
%c#1%
[ib];
%c#2%
[ib*5];
OUTPUT:
tech1;
```

APPENDIX F

ADVANCING RESEARCH
SEDL
IMPROVING EDUCATION

SEDL License Agreement

TO: Mary Schaal (Licensee)
5468 E. 34th St.
Yuma, AZ 85365

FROM: Nancy Reynolds
Information Associate
SEDL
Information Resource Center
4700 Mueller Blvd.
Austin, TX 78723

SUBJECT: License Agreement to reprint and distribute SEDL materials

DATE: January 14, 2009; revised February 2, 2010

Thank you for your interest in using the excerpts **Figure 3.1. Stages of Concern: Typical Expressions of Concern About the Innovation**, p. 31 and **Figure 4.1. Levels of Use of the Innovation**, p. 55 from the book *Taking Charge of Change*, revised ed., written by Shirley M. Hord, William L. Rutherford, Leslie Huling, and Gene E. Hall, and published by SEDL in 2006. These excerpts will be referred to as the "works" in this permission agreement.

SEDL is pleased to grant permission for use of the works cited above by the Licensee in her dissertation entitled, *Conceptualizing Change in Teaching and Learning through Structural Equation Modeling* at Northern Arizona University. The following are the terms, conditions, and limitations governing this limited permission to reproduce the works:

1. All reprinting and distribution activities shall be solely in the medium in which the works have been made available for your use, *i.e.*, a printed copy of the book *Taking Charge of Change*, and shall be solely for educational, non-profit use only. Precise compliance with the following terms and conditions shall be required for any permitted reproduction of the work described above.

2. No adaptations, deletions, or changes are allowed in the material, nor shall any derivative work based on or incorporating the work be created without the prior written consent of the SEDL.

Voice: 800-476-6861
Fax: 512-476-2286
www.sedl.org
4700 MUELLER BLVD. | AUSTIN, TX 78723

SEDL License Agreement, p.2

3. This permission is non-exclusive, non-transferable, and limited to the one-time use specified herein. This permission is granted solely for the period January 14, 2009 through February 2, 2011, inclusive. SEDL expressly reserves all rights in this material.

4. You must give appropriate credit: "Reprinted with permission of SEDL" All reproductions of the materials used by you shall also bear the following copyright notice on each page of use: "Copyright © 2006, SEDL."

5. An exact copy of any reproduction of the work you produce shall be promptly provided to SEDL. All copies of the work produced by you not distributed or used shall be destroyed or sent to SEDL, save and except a maximum of three archival copies you are permitted to keep in permanent records of the activity you conducted.

6. This License Agreement to reproduce the works is limited to the terms hereof and is personal to the person and entity to whom it has been granted; and it may not be assigned, given, or transferred to any other person or entity.

7. SEDL is not charging the Licensee a copyright fee to use the works.

I'm sending you two copies of this agreement. Please review this license agreement and sign one copy below, indicating that you understand and agree to comply with the above terms, conditions and limitations, and send the original back to me. If you wish to keep a copy with original signatures, please also sign, and return the second copy and, after I receive and sign them, I'll return one copy with both of our signatures to you.

Thank you, again, for your interest in using excerpts from SEDL's ***Taking Charge of Change***. If you have any questions, please contact me at 800-476-6861, ext. 6548 or 512-391-6548, or by e-mail at nancy.reynolds@sedl.org.

Sincerely,

Nancy Reynolds

Nancy Reynolds for SEDL

Feb. 4, 2010

Date signed

Agreed and accepted:

Signature: Mary J. Schaal

February 2, 2010

Date signed

Printed Name: Mary J. Schaal

APPENDIX G

University of Houston
Intellectual Property Use Permission Form

Thank you for your interest in the University of Houston. Please fill out the following information to submit your request to utilize the University's marks, website, or other communications materials.

NAME: Mary J. Schaal DATE: February 1, 2010

COMPANY:

ADDRESS: 5468 E. 34th St., Yuma, AZ 85365

PHONE NO: 928.344.0587

I am interested in using the University of Houston (check all applicable):

☐ Logo

☐ Athletics ☐ Institutional ☐ Learning. Leading.

☐ UH ☐ Other

☐ Seal

☐ Cougar image

☐ UH Website

☐ Link to UH ☐ Link from UH to other

☐ URL: ________________________

☐ UH Photo(s)
Name specific photo(s): ________________________

☐ UH Video(s)
Name specific video(s): ________________________

☑ UH Publication(s)
Name specific publication(s): Journal of Classroom Interaction, Vol. 34 No. 1 and Vol. 35 No. 1

I plan to use this/these resource(s) for: ☐commercial ☑ non-commercial purposes.

I plan to use this/these resource(s) in the following manner:

In the literature review of my dissertation: I am re-analyzing data related to those published in Vol. 34 No. 1: figure 2 on pages 22-23, and Vol. 35 No. 1: table 2 on p. 12; table 5 on p. 14; table 9 on p. 18; and table 10 on pp. 20-22. In my dissertation, *Conceptualizing Change in Teaching and Learning through Structural Equation Modeling*, I will compare results re-analyzing original data (provided by original researchers) using a structural equation model to results published in JCL gained through traditional analysis using ANCOVA, GLM, etc. The sample components from the IC Map are needed to help the reader understand the meaning of IC data and results being discussed.

I plan to use this/these resource(s) to support the following program(s):

To fulfill degree requirements of doctor of education in educational leadership at Northern Arizona University in Flagstaff, AZ.

I plan to use this/these resource(s) over the following period of time:

February-August 2010

I am working with Northern Arizona University and Dr. Archie George, University of Idaho (UH Department/Organization) in partnership on this project.

Mary J. Schaal

Return this form to the following address for approval:

Office of Public Affairs
204 E Cullen Building
Houston, Texas 77204-2021

Approval is granted ☐ denied for limited use of ______ for the following program and time period:

______

______

Karen B Clarke 2/17/10
Office of Public Affairs Date

______ ______
Office of the General Counsel Date

License Fee: ______ Paid: ______ ______
Date