Dissertations (2009 -)   Dissertations, Theses, and Professional Projects

Textbooks, Teachers, and Middle School Mathematics Student Achievement

Susan R. Monaghan
Marquette University

TEXTBOOKS, TEACHERS, AND MIDDLE SCHOOL MATHEMATICS STUDENT
ACHIEVEMENT

by

Susan R. Monaghan, B.S., A.M.

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

December 2013

ABSTRACT
TEXTBOOKS, TEACHERS AND MIDDLE SCHOOL MATHEMATICS
STUDENT ACHIEVEMENT

Susan R. Monaghan, B.S., A.M.

Marquette University, 2013

The purpose of this study was to extend the research on textbook effectiveness to a situated investigation of a single large urban school district in which middle schools were given a choice in selecting from three textbooks for mathematics instruction: a reform textbook, a commercially produced textbook developed in response to mathematics standards, and a traditional textbook. Its genesis is rooted in the efforts in the mathematics education community to investigate the interaction of teachers and mathematics curriculum materials, but in light of the shift to an accountability policy climate in public education. In particular, this study sought to determine whether the type of textbook selected by a school, moderated by the human capital of the teachers teaching mathematics, and the interaction of those variables was associated with increased student mathematics achievement on the mathematics portion of the eighth grade statewide standardized test.

Hierarchical linear modeling (HLM) was used to investigate models relating to textbook selection, components of teacher human capital, and their interaction. Contrary to the initial hypothesis, the interaction of textbook selection and components of human capital were not found to be significantly associated with student achievement. However, the selection of a reform mathematics textbook (CMP) over other more traditional texts was associated with student achievement, but accounted for very little of the variance in student test scores.

To further explicate the interaction of textbook selection with school factors, logistic regression was used to investigate the association between school factors and the selection of a reform textbook. The demographics of the school (i.e. race, SES, ELL) were not associated with the school selecting a reform mathematics textbook. However, one component of teacher human capital, expertise (a component constructed from data about teacher certification, mathematics specialization, and participation in math focused professional development) was associated with the selection of a reform textbook. This study suggests there is a connection between teacher human capital, the use of reform texts and student achievement; however further investigation is needed to understand the mechanisms at work.

ACKNOWLEDGEMENTS

Susan R. Monaghan, B.S., A.M.

This would not have been possible without the help and support of many people. If it were not for Joan Whipp, my advisor, I never would have finished this undertaking. Her continued support and motivation were invaluable throughout this entire process. The other members of my committee, Jack Moyer and Francesca Lopez, provided years of continued feedback and thought provoking comments, I cannot thank them enough. I would also like to thank all those from the school district who assisted in providing me with the data I used, in particular I would like to thank Matt Michala for his patience and help in this process. There were so many others who supported me throughout this process, but Julie Hapeman, Deb Dosemagen and Fred Call were there to listen, offer diversions, and share many meals. They, along with Bill and Nancy, who by example show the world how to live a life focused on generosity, equity and social justice, made this journey more meaningful and enjoyable.

Most importantly, I would like to thank my husband and children. Ryan and Ethan, my sons, spent much of their childhood wondering what I was doing and when I would be done. Knowing this would be a lasting memory of their childhood motivated me to continue on. And, none of this would have been possible without the love and support of my husband Sholom. His willingness to put up with my crazy desire to do this and his ability to put up with it for so long are a testament to his greatness.

DEDICATION

This work is dedicated to my extended family, particularly my parents. Without the

collective wisdom of their immigrant experience, none of this would have been possible.

Who knew someone like me could accomplish something like this?

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1: Introduction

On October 4, 1957, the Soviet Union launched Sputnik, the first artificial earth satellite, into orbit. This event is often credited with spurring a renewed focus on mathematics and science education but more accurately, Sputnik brought the conversation about the mathematics curriculum into the larger sphere of national awareness. Following Sputnik, national events such as the 1959 Report of the Commission on Mathematics (CEEB, 1959) and the Cambridge Conference of 1963 produced recommendations for reform.

The main recommendations for change included the inclusion of new topics such as logic, modern algebra, probability and statistics and the development of new high school courses that combined topics, speeding up the progression of students through mathematics. These recommendations reflected a core idea that reformers agreed upon: reorganization of school curricula around concepts, structures, and reasoning processes that modern mathematics uses as a common foundation for all specific branches of math. The reforms initially focused on high school mathematics but over time included recommendations for new curricula in elementary schools.

Despite the fact that New Math, as these reforms came to be known, is still part of our national lexicon on math reform, there was not large-scale change. Instead, the new approaches were met with skepticism from teachers; not all mathematicians and scientists supported the abstract structural argument; and the public found it to be too dissimilar from their own experiences. Research at the time showed students learned best what their curriculum focused on meaning. Students using New Math programs had a better

understanding of modern mathematical concepts and problem solving and students using traditional texts were better at arithmetic and symbolic calculation (Schneider, 2000; Ridegeway, 2003; Resendez, et al., 2005).

After this rejection of New Math, schools moved towards a back to basics approach. Reformers were left believing it was very difficult to change school and teacher practice and gain community support; and mathematics reform efforts were pushed to the background.

In 1983, the National Commission on Excellence in Education published the report *A Nation at Risk*. This report is credited with launching a new wave of educational reforms including the standards movement, but, like Sputnik, the series of events that led to the standards movement started before the national awareness developed. The first set of standards and policy statements came from the National Council of Teachers of Mathematics (NCTM). The NCTM standards are not a singular document, but a series of documents that were triggered by the *Agenda for Action* (1980). They include the *Curriculum and Evaluation Standards for School Mathematics* (1989), *Professional Standards for Teaching* (1991), *Assessment Standards for School Mathematics* (1995) and *Principles and Standards for School Mathematics* (2000).

Like the prior reform known as New Math, Standards Based Math or Reform Math suggested a shift in focus from skills and procedures to problem solving, understanding and connections within the domain and to other domains. Unlike New Math, with its focus on the structure of mathematics, reform math was also informed by advances in learning theory. Aware of the problems of New Math, this time around, reformers also tried to address the issues of teacher acceptance and public awareness.

More educators were involved in writing the standards and the standards addressed far more than a shift in content focus, including issues of pedagogy and assessment. New materials were developed that not only contained the content specified by the standards, but were field-tested and in some cases were designed to provide learning opportunities for teachers as well.

Despite greater efforts to address past problems with reform, the use of standards based materials never became the norm. These new materials were once again negatively received by a portion of the public, leading to what has become known as the math wars (Schoenfeld, 2004) . Despite the controversy, forty-one states did develop math standards or frameworks that were consistent with the NCTM standards but the larger standards movement, through the adoption of No Child Left Behind (NCLB), became a test-based output-oriented movement. These tests do not align well with state or NCTM standards, particularly because of their focus on basic skills. What started as a movement for curriculum reform morphed into a testing movement focused on the basics.

Now mathematics educators are focused on the Common Core. Whether this is a new reform or an extension of the standards movement is not yet clear. The Common Core Standards share some characteristics with the NCTM standards in that they focus on content beyond basic skills, but the Common Core does not prescribe a specific pedagogical perspective. The results of this movement are yet to be revealed, but some see promise in a shift to assessments better aligned with the more challenging content within the standards, possibly signaling a shift away from the current skills assessments.

In all these reform movements, the impetus for change in mathematics education has always been to increase student learning of mathematics. The debate over skill versus

understanding has been a constant debate since at least the early nineteenth century when books were available which taught the rule method, the inductive method and the analytic method, the latter two of which focused on thinking over rules. Sputnik, *A Nation at Risk* and the Common Core mark events in the ongoing struggle in mathematics education between skills and less rule bound approaches.

In addition to these swings or cycles in policy, the struggle at the institutional level has been to find the best math curriculum, typically meaning the curriculum that produces the greatest number of proficient scores on the annual state-wide standardized test. Given that mathematics textbooks are the de facto curriculum, the best textbook is the one that produces results on these tests. An examination of these same reforms at the district level exposes the tension between recommendations from experts and schools.

The struggle of the large urban school district that is the focus of this study to embrace promising reforms by selecting textbooks promoted by experts illustrates this tension. In the early years of New Math, this district worked with mathematics educators to provide professional development and do research on the implementation of New Math. Despite the close connection between the district and at least one developer of New Math materials, this dedication to reform ended after a few years. Push back from educators and the community resulted in the selection of three different texts for each grade level, from which schools were allowed to select the text or texts that they felt would best serve their student population.

Similarly, in 2000, the district worked with a local university in implementing a single reform textbook at each grade level. Again, after only a couple of years, schools pushed back, and were given permission to choose any text they wanted to use for

mathematics instruction. This led to schools, and even classrooms within the same school, using different texts. Eventually, at the time of the next district-wide mathematics textbook adoption in 2005-2006, the district selected three textbooks for each grade level, and allowed schools to choose from these texts. To distinguish these texts, some within the community referred to them as "traditional", "hybrid", and "reform".

Over the course of history, as national reform movements promoted different approaches to mathematics, this district initially embraced change. The response from various constituencies within the district then led to a different set of decisions, muting the effect of the change. All these decisions, whether pushed by the district or the schools and teachers, were an effort to find the best text to increase student achievement. But, there are at least two questions that should be considered when seeking the best textbook: Is there one best textbook? And how do we judge best?

Currently, for many, how we judge what is best is directly linked to No Child Left Behind. The adoption of statewide annual testing as a metric to judge districts, schools and in some cases teachers has left us with a definition of best that aligns with the test results. Many of these tests favor skill and procedure based problems so the answer to which is best becomes the textbook that is associated with higher student achievement on skills and procedures tests.

Along with most states adopting the Common Core, most states have also joined one of the consortia developing new standardized tests. These computer adaptive assessments have been described as more rigorous with more open-ended questions. Given the likelihood that these new tests will focus less on skills and procedures, it appears that evaluating effectiveness in mathematics education may change.

Recent research on textbook effectiveness using outcome measures of state or national standardized test scores do not show consistent differences in achievement between groups of students using different types of texts. The results frequently show greater achievement by students using reform texts when the outcome measures are open ended (constructed response) assessments, whether researcher created or standardized. As was the case in the era of New Math as well, differences in achievement frequently relate to the type and content of the assessment used.

Linking a type of textbook to student achievement in an effort to find the best textbook is a limited view. Although textbooks represent the de facto curriculum in mathematics, instruction is a process that involves the teacher interacting with a textbook (Brown, 2009). Given this interaction between teacher and text, investigating which text is best should include a consideration of who is using the textbook and also accepting the possibility that there may not be one best text.

If we accept that who uses the text matters, teacher qualifications need to be considered. Extensive educational research on teacher effectiveness exists, yet is not conclusive. Teachers are shown to have an impact on student achievement, but the specific characteristics or qualities that are associated with student achievement are less clear. Certification, experience, educational attainment and professional development have all been shown to be associated with increased student achievement, but the results are not consistent, nor the association strong (Hanushek, 1986; Aloe, 2013)

In addition, students experience a variety of teachers in the time they are enrolled in a school, typically a different math teacher each year. Their learning of mathematics over a number of years is not the result of instruction from a single person. Therefore,

examining the impact of a school and characteristics of that school on student achievement may increase our understanding of the impact of an education in a particular context.

The notion that the qualifications of a group of workers, in this case teachers, can impact performance has been studied in other settings. Economists use the construct of human capital, first developed by Becker (1964), to operationalize the impact of humans and groups of humans on productivity. Stein and Kim (2009) have suggested that the interaction of human capital and textbook type should be considered in determining which textbook would be best in a particular setting. However, given the plethora of data available on teachers, it would be necessary to consider how the construct of human capital can be operationalized before it could be used in guiding discussions of textbook selection. Further, the interaction of this construct with the type of textbook needs to be considered to ascertain if there is an association with student achievement.

As one looks back across time, it is doubtful there will ever be agreement on the best textbook or even the best type of textbook. The development of the Common Core will not decrease the importance of this question. The standards and assessments will set guidelines for what society believes is good mathematics instruction, but these guidelines do not explain how to increase student achievement. Schools continually struggle to balance input from mathematicians, mathematics education experts, teachers and the public in order to put together a curriculum, which in the case of math is greatly dependent upon a textbook. More guidance is needed to move textbook selection to a data driven process that considers the teachers who will be using the textbooks selected.

Overview of the study

This study intended to be an initial step in addressing the knowledge gap in textbook selection policy about the relationship among teacher human capital, textbook type and student achievement. Intrigued by the amount of support in the mathematics education community for standards based curricula, the focus on student achievement as measured by standardized tests, and the realities of schooling, I investigated which type of textbook, school characteristic of teacher human capital, and/or the interaction of these produce the highest student achievement. This was a retrospective study focused on the mathematics achievement of middle school students in a large urban district and how that achievement is related to school-wide teacher human capital and textbook type. Specifically, this study addressed these questions:

- How can the construct of human capital in a school be represented using a combination of available data?

- How does textbook type affect student achievement when moderated by human capital?

- Are there differences in student achievement among textbook type use groups on open-ended (constructed response) items on the annual mathematics exam?

Definition of terms

Although these terms will be given greater clarity through this study, brief definitions of some terms, particularly those related to textbook type are given here. Throughout the research on textbook effectiveness, texts have typically been classified as

traditional or reform, depending on whether they were developed in response to the NCTM standards and with NSF funding or commercially produced. However, publishers market their texts as meeting different needs in the market, some more focused on skills, others more focused on the standards.

Textbook types

Reform text. A reform text is a text that was developed by mathematics educators and/or mathematicians, with funding from the National Science Foundation (NSF) in conjunction with a publishing house, in response to the 1989 standards document produced by NCTM. The impetus for the development of these texts came from the research community, and the publishing houses acted predominantly as distribution agents. These texts were piloted and field-tested as part of the development process. As the review of the literature will show, reform texts are more often focused on understanding and higher order thinking than traditional texts and also may contain more limited skill and procedure practice.

Traditional text. Within the mathematics education research community, commercially produced texts are sometimes referred to as traditional texts, meaning texts produced in the traditional manner, by commercial publishing houses and taking a traditional (skill and procedure based) approach. Some traditional texts present mathematics as a series of skills and procedures to be learned. The focus is on mastery of these skills and procedures rather than understanding. These are the texts that will be referred to as traditional texts in this study.

Commercially produced standards based text (hybrid). Within the commercial textbook market, as well as skill and procedure based texts, there are texts which include

more applications problems and more connections. The publishers market these as having been developed in response to the standards, but they were not developed in conjunction with funding from the NSF nor were they field tested during the development process. Within this study, these texts will be referred to as commercially produced standards based texts.

Question Style

Multiple-choice questions. Test questions in which the respondent is asked to choose the best answer given a number of options. They are common in standardized assessment and are sometimes referred to as forced response items or selected response items.

Constructed response questions. Test questions in which the respondent is asked to answer a question including providing an explanation. They are less common on standardized tests and are sometimes referred to as short answer or open-ended items.

Significance of the Study

Through the National Science Foundation and the National Council of Teachers of Mathematics, mathematics textbooks that incorporate standards and are based on research on student learning of mathematics are available. At the same time, textbooks that reflect other perceptions of high quality mathematics curricula are also available. Current research connecting textbooks to student achievement does not show any type of text consistently produces higher student achievement. These textbooks are used by teachers for instruction, impacting achievement.

The teachers using the text mediate the effect of a textbook. The process of teaching necessarily leads to students experiencing the same text in different ways. One possible conclusion is that the best text in any particular setting may depend on the human capital (the experience, certification, education and professional development) of the teachers using the textbook. More specifically, Stein and Kim (2009) suggest it is the human capital at the school level that matters in finding which text may work best.

This connection linking textbooks and teacher human capital has not been investigated. The literature reflects attempts to link textbooks and achievement or teachers and achievement, but efforts to find an association between texts and achievement that takes into account the mediating effect of teachers on this relationship is missing. This study was an effort to fill that gap.

This study examined the effect of particular texts moderated by the teacher human capital in a school to produce student achievement in mathematics. It was an initial effort to use existing data and district and school textbook selections to determine if there are differences in student achievement associated with these factors.

Chapter 2: Literature Review

The ability of a textbook to produce improved student achievement is contingent upon who is using the textbook. To develop some understanding of the link between the type of textbook a school selects and student achievement and how that may be mediated by teacher qualifications, I review literature on standardized test quality, mathematics textbook effectiveness and teacher qualifications with particular attention to middle schools. In addition, I present information on the development of the concept of human capital as a construct for framing the qualifications of the teachers within a school.

Measuring Student Achievement

The publication of *A Nation At Risk* (National Commission on Excellence in Education, 1983) began the current accountability movement and resulted in the creation of state level commissions, increased graduation requirements, increased course loads and standardized testing (Kornhaber & Orfield, 2001). By 2000, only Iowa did not have a state mandated test and testing continued to be a policy concern, culminating in the passage of No Child Left Behind Act (NCLB) in 2002 (Jones, Jones, & Hargrove, 2003).

The main accountability component of NCLB, and thus the way policymakers most often determine student learning, is the administration of annual assessments, most of which are multiple-choice tests. From a policy and accountability perspective, multiple-choice tests have many positive characteristics; they are cost-effective, easy to administer, can be scored by machine, and allow for comparisons between students, classrooms, schools and districts. However, according to a Government Accountability Office report, "despite the cost- and time-saving benefits to states, the use of multiple

choice items on assessments has limited the content included in the assessments" (Ashby, 2010, p. 20). This conclusion is backed by studies that examine these tests in light of state standards.

Three different methods have been used to study alignment between standards and tests: each attempt to quantify how well the tests assess the content in the standards. Independent of the method, the results show that both the amount of content covered by a test and the level of difficulty of the items are below that indicated by the standards.

Studies of alignment typically focus on coherence and find a limited level of coherence between state standards and tests. Using a set of four indicators of alignment (categorical concurrence, depth of knowledge, range of knowledge and balance of representation), Webb (1999) examined the alignment of assessments in four states. One mathematics assessment from the elementary grades (grade 3 or 4) and one from middle school grades (grade 6 or 8) were examined to determine whether the assessment covered the content in the standards and the depth and breadth of that coverage. Webb found that tested material was at a lower level of cognitive demand than specified in the state standards and that the assessment covered 50% or less of the material covered in the content standards. This level of coherence was described as poor to moderate, and a follow up study of three states reached the same conclusions (Webb, 2002).

A similar alignment method, the Achieve alignment method, uses dimensions first to confirm the test blue print to assess performance centrality, source of challenge and level of cognitive demand, then secondarily to evaluate the level of challenge, balance and range of questions. This method was used to evaluate assessments in five states (Rothman et al., 2002). States with objectives written in global terms received low ratings

because of the difficulty of determining item-objective matches. Overall, individual items were found to match to content and performance standards, but assessments did not assess the full range of standards, with the most challenging standards under-sampled or omitted.

The most common instrument for assessing alignment, the Surveys of Enacted Curriculum content coding procedure, measures the extent to which the proportions in one content matrix (e.g., describing an assessment) match the proportions in another content matrix (e.g., describing standards). Results range from 0 (no alignment) to 1.0 (perfect alignment). Porter (2002) notes that there is no standard for what is "good" alignment, but indicates that this procedure works well for comparing the alignment of a test with particular state or NCTM standards.

Using the Surveys of Enacted Curriculum content coding procedure, Porter (2002) found the average alignment within a state between the state standards and test was 0.40. In examining four states, alignment between states averaged 0.39, and average state assessment-NCTM alignment was 0.39. This suggests state standards may not be specific enough for tests to be tightly aligned or states need to bring tests into alignment with standards. Porter also concluded that there was a low to moderate alignment between standards and assessments in mathematics and suggests that alignment within a state should exceed alignment between states because each state has developed its own assessments to align with that state's standards.

More recently, using data from thirty-one states and the Surveys of Enacted Curriculum content coding procedure, researchers found an average alignment index in mathematics of 0.27 (Polikoff & Porter, 2011). Although this is better than the averages

found for English/language arts/reading and science, the range in mathematics was larger, from a maximum of 0.47 for algebra in one state to a minimum 0.01 for grade 10 mathematics in another state. And although the value obtained from the alignment index is not easily classified as good or bad, these indices were judged by the authors to be quite low. Supporting this conclusion, a further analysis of alignment revealed an average misalignment of 45% between test items when considering the content and level of cognitive demand as compared to the standards.

Despite these shortcomings, a review of alignment practices has shown that individual items on an assessment map quite well to the standards. However, the tests cover a small portion of the standards, sometimes as little as 27% (Resnick et al., 2004). In addition, when assigning global ratings to the assessment, the level of challenge in the assessment are commonly judged to be inappropriately low compared to the standards.

These results, while consistent with one another, are not consistent with a survey of state education departments. With 46 states and the Department of Defense responding, when asked about alignment, almost 90% of states responded that their assessment aligned well with their standards (Reys et al., 2005). This difference between researcher coded alignment and state reported alignment may, in part, be due to a lack of a standard framework for assessing alignment. Items on the tests may map back to a standard or standards, but the standards are not fully represented by the test.

Despite the lack of a consistent framework for analyzing alignment between standards and tests, it is clear from the studies above that alignment is not strong. Assessments do not assess all or most standards, and assessment questions are frequently at a lower level than those suggested by the standards. Although these studies examine a

small number of states and making comparisons is difficult, the consistency of the findings suggests that the breadth and depth of challenge of state standardized tests is below optimal.

Mathematics Curricula

Textbook content. The extensive use of mathematics textbooks makes knowledge of their content particularly important. In mathematics, texts are often the de facto curriculum. In the middle school grades, most teachers report using a mathematics textbook most of the time (Grouws & Smith, 2000; Weiss, Banilower, McMahon, & Smith, 2001). This reliance on texts is corroborated by student report on the 2000 NAEP, in which 72% of participating eighth graders reported that they did mathematics problems from their textbooks on a daily basis (Braswell, Lutkus, Grigg, Santapau, Tay-Lim, & Johnson, 2001).

Although there is not a uniform framework for analyzing and comparing textbook content, the research literature suggests differences between texts that are similar to the differences between standards and standardized tests. In an analysis of four texts, three traditional textbooks and *Everyday Mathematics* (a reform text), traditional texts were shown to provide more tasks directly related to number relationships and more skills practice with number sense. In comparison, *Everyday Math*, a reform textbook, focused on real world connections and hands on activities that develop a variety of models and representations (Sood & Jitendra, 2007).

Similarly, an evaluation of the level of cognitive demand of probability tasks in two middle school textbook series, one traditional and one reform, found that the number of probability tasks was greater in the traditional text, but the level of cognitive demand

was greater in the reform texts (Jones & Tarr, 2007). Tasks were analyzed using the Mathematical Tasks Framework (Stein, Smith, Henningsen, & Silver, 2000). Over half the tasks in the reform text were found to be high cognitive demand, yet less than one fifth of the tasks in the traditional text were found to be high cognitive demand tasks.

In a more comprehensive review, Project 2061, a long-term American Association for the Advancement of Science initiative to advance literacy in science, mathematics and technology, analyzed thirteen middle school mathematics series, four reform series and nine commercially produced series (American Association for the Advancement of Science: Project 2061, 2000). Textbooks were evaluated based on a set of benchmarks for number concepts and skills, geometry concepts and skills, and algebra concept and skills. The series were rated as having most, partial or minimal content coverage of the benchmarks. Although none of the series were found to address all benchmarks sufficiently, four series were found to address four or more benchmarks in depth. These were the four reform series.

In general, reform texts are reviewed more favorably than traditional textbooks. The only textbook analysis that negatively reviewed reform texts was conducted by the organization Mathematically Correct, an advocacy group organized in opposition to the NCTM standards and reform curricula (Mathematically Correct, 1999). In a review of seventh grade mathematics texts, this advocacy organization considered the depth of the mathematics, the presentation of skills and concepts and the student work required. The evaluation focused on skills and procedures. Using these criteria, traditional texts did very well, with reform texts fairing poorly. Most notably, *Connected Math*, the most commonly used middle school reform mathematics text, received a grade of F.

Although analyses of content can shed light on the goals of a program and the alignment with standards, these analyses cannot determine the effectiveness of the program when implemented in a school district, school or classroom.

Textbook effectiveness. The hopes of the mathematics education community that reform textbooks would improve achievement have not been borne out in the research. These reform texts were developed with funding from the National Science Foundation in response to the NCTM standards, and are referred to throughout this review as reform textbooks. Some studies of reform textbook effectiveness show small impacts on student achievement when measured by standardized tests, rarely in composite scores, sometimes in component sub-scores, and more frequently when the outcome measure is an open-ended or researcher created assessment (Riordan, Noyce, & Perda, 2003, Eddy et al., 2008, Cai et al., 2011). Alternately, some studies have also shown positive impacts of traditional textbooks or no difference between the two (Resendez et al., 2005; Schneider, 2000; Ridgeway, 2003). This review focuses on studies of middle school mathematics textbooks.

Even though typical multiple-choice tests may not discern differences in the mathematical understanding emphasized in the NCTM standards, initial mathematics textbook effectiveness studies frequently used multiple-choice tests as the only outcome measure, possibly contributing to inconclusive or inconsistent results. On the other hand, more recent studies (e.g. Eddy et al., 2008; Cai et al., 2011) frequently use multiple outcome measures, many of which include open-ended questions.

Studies that use multiple choice tests as outcome measures show limited results. For example, two studies of CMP found a significant positive effect on some sub groups.

Twenty-one middle schools that used CMP for two to four years were matched with 34 comparable non-CMP schools. Analysis of variance showed CMP to have a significant positive effect ($d = 0.23$) (Riordan & Noyce, 2001). A follow-up study, which matched students who had used CMP for three years with students who had used another curriculum for three years, showed a small but significant positive difference on the Massachusetts Comprehensive Assessment System (MCAS) mathematics test ($d = 0.09$) (Riordan, Noyce & Perda, 2003).

A more recent study evaluating the effect of CMP2 on mathematics achievement of sixth graders used random assignment of textbooks to schools (Martin et al., 2012). This two-year study spanned the implementation year, where teachers were provided with professional development, and the data collection year (the second year teachers were using the curriculum). The TerraNova was used to collect baseline and outcome data for students. Data were collected from 65 schools across the Mid-Atlantic region, with 82% of eligible students participating in both the baseline and outcome TerraNova. The impact of CMP2 on TerraNova post-test scores was not statistically significant, and less than one point. The authors did note a difference in the type of instructional activity taking place in schools using CMP2, as well as an increase of an average of 1.18 hours per week spent on math in the CMP2 schools.

Commercially published, non-NSF funded programs are rarely the subject of research, but instead, serve as a control in studies of reform texts. The few studies of commercial texts show similar findings, that outcomes consisting of multiple-choice items only rarely give significant results. In one of the few studies of a commercially produced text, schools using McDougal-Littell Middle School Math, a commonly used

middle school math program, were compared with students in schools using a variety of other texts. There were no statistically significant differences in performance on a selection of publicly released NAEP mathematics items (Callow-Heusser, Allred, Robertson, & Sanborn, 2005).

The only commercially produced text that has been the subject of multiple studies is Saxon Mathematics. This text presents a step-by-step approach to mathematics, stressing mastery of skills and procedures. The results of these studies are mixed. Some studies find a significant difference in performance between students using Saxon and those in the matched group. Saxon Math students showed significantly higher performance on the Metropolitan Achievement Test (MAT) (Lafferty, 1994), the California Test of Basic Skills (CTBS) (Rentschler, 1994), and the Texas Learning Index (Resendez, Fahmy, & Azin, 2005). However, there was no difference between Saxon Math students and matched schools on the Criterion Referenced Competency Test (CRCT) (Resendez & Azin, 2005), and Saxon Math students were outperformed on the SAT-8 by students in the control group (Roberts, 1994).

Limitations in the outcome measures of some studies makes finding differences in student learning difficult to uncover. Studies utilizing sub scores and/or open-ended assessments in addition to composite scores on standardized tests have shown greater differences in student outcomes. This research shows favorable results among students using reform texts.

A comparison of 23 CMP grade 6-8 schools in Texas with 25 schools, matched by their predicted values on the 1996 Texas Assessment of Academic Skills, found that students in the CMP schools scores statistically significantly lower ($d = -0.14$)

(Schneider, 2000). However, the What Works Clearinghouse (WWC) reanalyzed the data in this study because the unit of assignment and unit of analysis were mismatched. After statistically correcting for the mismatch, the WWC found no statistical significance between the groups (WWC, 2010).

In a similar comparison study of middle school students' mathematics achievement, student in nine schools using CMP were compared to students in nine schools not using CMP (Ridgeway et al., 2003). Schools were matched by location, student population density and student ability. Using ANOVA, the researchers found students using CMP in grade 6 had statistically significant smaller gains than those using other textbooks, with no statistical difference for 7th and 8th grade students on the Iowa Test of Basic Skills (ITBS). On the Balanced Assessment in Math (BAM), students using CMP outperformed non-CMP students in all three grades, with effect sizes of 0.15, 0.53 and 0.8 respectively.

Eddy and associates (2008) also used the ITBS and BAM as outcome measures to compare achievement in grade six mathematics for students in CMP2 and non-CMP students. Six middle schools across three states were recruited for this study. At each school, teachers were randomly assigned to treatment or control groups, resulting in 11 CMP teachers, 9 non-CMP, and an attrition rate of 18% of students between pre and post-tests. ANOVA and HLM were used to assess the effect of CMP2 on student outcomes. The HLM analysis found no statistically significant difference between students in the treatment or control groups. It is possible the small number of teachers in the study resulted in low statistical power, which would require a large effect size in order to be significant. This study is also limited in its findings because teachers were randomly

assigned within a school. There is no way to know if teachers in the same school using different texts affected teacher's instructional practices, decreasing the differences between treatment and control groups.

In a longitudinal study of algebraic learning of middle schools students from sixteen schools within one large urban school district, comparing those students who used CMP with non-CMP classrooms, four outcome measures were used: open-ended tasks, translation tasks, computation tasks and equation solving tasks (Cai, Wang, Moyer, Wang, & Nie, 2011). Overall growth rate analysis showed no difference in growth rates for CMP and non-CMP students on computation and equation solving tasks. Growth rate on open-ended tasks and translation tasks were higher for CMP students. Using growth-curve modeling over the three middle school years, the CMP students' scores on open-ended tasks increased significantly more than non-CMP student scores ($t = 2.17$, p < 0.01) CMP students gained an average of 25.09 points annually, and non-CMP students gained an average of only 19.39 points. The authors suggest that gains in conceptual understanding did not come at the expense of basic skills gains.

Similar results were found in a study of 1400 middle school students using reform curricula. Using both the SAT-9 and the New Standards Reference Exam in Mathematics (NSRE), researchers found student achievement levels on the Open-ended and Problem Solving subtests were greater than those on the Procedures subtest. The mean achievement of the students in the study exceeded the mean achievement for these nationally normed tests only on the Open-ended and Problem Solving subtests. On the Procedures subtest, student achievement was below the national norm. In this case, the students involved were in districts that adopted reform curricula for all students, so there

were no control groups. These results show students using reform curricula (*CMP* and *MATHematics*) perform better on open ended and problem solving tasks but not as well on procedural tasks (Post, Harwell, Davis, Maeda, Cutler, Andersen, Kahan, & Norman, 2008).

Overall, textbook effectiveness studies are difficult to compare and evaluate. Most studies are funded by groups who have a vested interest in the outcome of the study, whether it is the National Science Foundation, funders of reform textbook development, or publishing houses, funders of traditional textbook development. Studies also typically compare a single text with a group of other texts, giving the impression that all texts not the focus of the study are somehow equivalent. And, some of these studies are limited by their reliance on composite scores on multiple-choice tests. As noted earlier, standardized multiple-choice statewide tests are typically judged to be of poor quality and therefore may not be ideal as outcome measures in studies of textbook effectiveness. Studies that use component sub-scores or multiple outcome measures including some type of open-ended questions show more promise in uncovering the differential impact of different textbooks. Open-ended tasks may be more representative of the standards, and studies that include them do indicate that reform texts are likely to impact student achievement in important ways (Post et al., 2008; Eddy et al., 2008; Cai et al., 2011).

Teachers

Teaching can be described as the interplay between teacher and text. The text plays a role in affording and constraining teacher's actions, and teachers use texts differently given their experience, intentions and abilities (Brown, 2009). This results in different experiences for students, dependent on both textbook and teacher.

Teacher quality, teacher qualifications, teacher effectiveness, and teacher characteristics, and in the case of economics research, teacher human capital, used almost interchangeably in the literature, have long been the subject of research, with few strong conclusions. Much of the early research on teacher quality used cross-sectional data aggregated at the level of the school or district. This research attempted to find a relationship between average student test scores and measures for teachers. The results were fairly consistent, finding the average performance of individual teachers as measured by aggregate student scores differed significantly but little of the difference was accounted for by measures of teacher quality such as certification and experience (Wayne & Youngs, 2002; Lankford et al., 2002; Mueller, 2012). Hanushek (1986) reviewed 147 of these studies, noting the common limitations of aggregate data, limited teacher data, inconsistent outcome measures, and concerns that teacher effects in strong districts were overstated because of the lack of controls for prior achievement.

To compensate for some of these limitations and because of greater availability of data and analytic techniques, researchers began using gain scores and, more recently, fixed effects and multi-level models to look for teacher effects based on various observable teacher qualifications such as certification, academic degree, major or minor, experience, and professional development.

Teacher Certification. Whether teacher certification matters is the object of ongoing debate. Still, certification is the primary criteria by which individuals show they meet the minimum requirements for teaching (Greenberg, Rhodes, Ye, & Stancavage, 2004). This practice has been longstanding. Although certification started out predominantly as a way for communities to be assured of the moral character of their

teachers, by the mid nineteenth century, the majority of teachers held some type

certification, frequently based on performance on a test (Angus, 2001). With the advent

of normal schools, certification criteria shifted to completion of an educational program.

Policies today largely rely on educational program completion and testing to determine

minimum qualifications for certification.

In general, certification requirements include requirements that address content

knowledge and knowledge of teaching and learning. These requirements vary by state,

and typically include completion of a state approved teacher education program, and in

many cases, completion of a minor or major in the area of certification. Most programs or

states require student teaching between 8 and 18 weeks, and one or more test of content

knowledge, knowledge of teaching, and/or basic skills (Darling-Hammond, 2000).

Many states offer elementary and/or secondary licenses that cover the middle

school grades (K-8 licenses and/or 6-12 licenses). This practice leaves middle school

teachers with either a generalist's license, with an emphasis on pedagogy appropriate for

young children, or a content specific license, with an emphasis on subject matter

knowledge. According to one national survey, less than 10% of middle school teachers'

initial certification was specifically for middle school, with the majority of middle school

teachers holding elementary licenses (NMSA, 1996).

Despite the requirement that teachers be fully certified, most states allow for some

type of provisional, temporary, or emergency license, most teachers who are not fully

certified teach low-income students of color in under-resourced schools (National

Commission on Teaching and America's Future, 1996; Darling-Hammond, 2013). Even

though the history suggests that certification was developed as a way to ensure student

are taught by teachers with the necessary qualifications, urban districts often employ teachers with low qualifications and weak academic credentials (Murnane & Steele, 2007). Whether certification matters in producing student achievement is a much studied question.

Many studies of teacher certification show mixed results. A study of elementary students in grades three through eight matched 9849 math and reading teachers with student achievement data from 1999-2005. Using mathematics gain scores, they found students of uncertified teachers and internationally recruited teachers underperformed by a small amount, with gains of -0.005 SD and -0.05 SD respectively, when compared to certified teachers (Kane, Rockoff, & Staiger, 2006).

Similarly, a study of elementary mathematics achievement found that students taught by new, uncertified teachers did significantly worse on an achievement test than students of new, certified teachers (Laczko-Kerr & Berliner, 2002). In contrast, Betts, Zau and Rice (2003), examining data from a large number of schools between 1997-2000 found mixed results when they looked for a link between teacher certification and student achievement. Their data from 123 elementary schools, 24 middle schools, 17 high schools, and 5 charter schools within one city showed that students of certified interns (with 0-1 year experience) had lower mathematics achievement scores than students of uncertified interns with the same experience. However, high school math students taught by certified teachers had higher achievement scores in mathematics than their peers taught by uncertified teachers. No differences were found when comparing other certification types.

There are many studies that show a positive relationship between certification and student achievement (Betts, Reuben, & Danneberg, 2000; Darling-Hammond, 2000; Goldhaber & Brewer, 2000). Many of these studies use large data sets but find small effects. For example, in an examination of the effect of certification on achievement, five years of student data from more than 4,000 fourth and fifth grade teachers showed greater student achievement gains for students with certified teachers than students of uncertified teachers ($d = 0.01$ to 0.09) (Darling-Hammond & Holtzman, 2005).

In addition, many of the certification studies suggest the impact of certification may have more to do with the content knowledge implied by the certification than the actual certification status. For example, in one study of student achievement in mathematics, middle school and high school students of certified teachers outperformed students of uncertified teachers (Betts, Zau, & Rice, 2003). These researchers, however, suggest this effect may be the result of greater content knowledge implied by these certifications than by certification status. Goldhaber and Brewer (1999, 2000) also suggest this in their studies using National Education Longitudinal Study (NELS) data.

Two large-scale studies using data from the NELS 1988 database examined student performance and certification using student achievement data from 1990 and 1992. The NELS 1988 study collected survey and achievement data from approximately 24,000 students in 638 schools served by 2,245 teachers in 3,498 classrooms. This survey and testing was followed up in 1990, retesting 18,000 of the original 24,000 students. The first study used data from 5,149 tenth grade math students given a variety of mathematics tests based on performance on the original (eighth grade) assessment. Student performance in mathematics was higher among students taught by teachers with some

type of mathematics specific certification, including emergency certification, than by those taught by teachers with non-mathematics certification (Goldhaber & Brewer, 1999). This was followed up by a study conducted when these students were in twelfth grade, and produced the same results. Students of teachers with any type of mathematics certification outperformed students of teachers with non-mathematics certification ($d =$ 0.16) (Goldhaber & Brewer, 2000). The authors posit that this result may be indicative of the school district's screening of uncertified teachers for content competence, leading them to also conclude that subject matter knowledge may be more important than full certification for high school mathematics teachers.

Other studies suggest content specific certification at the high school level makes a difference. In an examination of 100,000 student gain scores in mathematics, students of teachers with a regular state certification in mathematics were predicted to have higher achievement in mathematics ($d = 0.11$) at the high school level. Although a similar smaller effect was found in the middle school grades, the effect was not significant (Cavalluzo, 2004).

Using data from all fifty states, including the 1993-1994 Schools and Staffing Surveys and the National Assessment of Education Progress, partial correlations showed a significant relationship between teacher quality and student achievement (controlling for student poverty and ELL status) (Darling-Hammond, 2001). The most consistent positive correlation with achievement in mathematics and reading was the proportion of well qualified teachers, defined as teachers with full certification and a major in the field they teach ($0.61 \leq r \leq 0.80$). The strongest consistently negative correlations were the

proportion of new teachers who were uncertified $(-0.40 \leq r \leq -0.63)$, and the proportion that held less than a minor in the field they teach $(-0.33 \leq r \leq -0.56)$.

In a large, school level study examining the performance of students in schools as related to the number of teachers with emergency credentials, a significant negative relationship (r = -0.055) was found between student achievement and the number of teachers with emergency certification. Using data from 6,389 schools and their annual performance index (API) for 1999-2000, regression found most of the variation among schools could be explained by demographic differences. However, the percent of teachers holding an emergency permit predicted differences in student achievement (Goe, 2002).

Opponents of current certification practices cite the inconclusive evidence that certification matters, express concern that certification is a barrier for some wishing to enter the teaching profession, and argue that the research can be interpreted to show that content knowledge is what matters (Ballou & Podursky, 2000; Paige, 2002). While the studies cited above indicate that certification itself does not necessarily increase achievement, they do suggest that secondary certification may correlate with higher student achievement, perhaps due to increased content knowledge requirements at that level.

Teacher Academic Degree, Major or Minor. A teacher's academic credentials such as their college major and highest degree held are often used to assess teacher quality. Despite policies that reward the attainment of an advanced degree, a major or minor in a content area is more likely to contribute to student achievement. Due in part to compensation policies that reward advanced degrees, data relating to and studies of

advanced degrees are far more common than studies of college major, as this data is less available.

According to one study, although many factors were associated with student achievement, the absence of a teacher with at least a minor in the subject taught accounted for almost 20% of the variation in National Assessment of Educational Progress (NAEP) scores (Darling-Hammond, 1998). The association between degree and achievement is somewhat dependent on grade level.

In one study of teachers in Florida, researchers were able to find the college minor of all teachers who had attended public universities after 1995. Findings suggest, at least for middle school, teachers with general education majors were less productive than other teachers (Harris & Sass, 2007). Similar findings suggest that higher student achievement is associated with a teacher with a math major or minor (Klecker, 2008). Using 2007 NAEP data, student performance for middle school students was higher if their teacher had a major ($d = 0.27$) or minor ($d = 0.25$) than a teacher with neither. The difference between students of teachers with a major or minor was small ($d = 0.09$).

This difference in mathematics achievement can also be found at the high school level. In a large scale study of teacher resume characteristics and student achievement found an association only for high school math (Aaronson, Barrow, & Sanders, 2007). Using data for students from all Chicago public high schools, including test results on eighth grade (ITBS) and ninth grade (Test of Achievement Proficiency) mandated tests, results show a link between college major and student achievement. A teacher holding a math or science degree was associated with an increase of 0.06 to 0.08 grade level equivalents, but other majors were not associated with increased achievement. Similar

results were found in a study of twelfth grade math achievement. Students of teachers with an undergraduate degree in mathematics had higher levels of mathematics achievement than comparable students with teachers holding majors in other fields (Goldhaber & Brewer, 2000).

This association between major or degree and student achievement is not supported by all research. There may be differences in impact at different grade levels (Hawkins, Stancavage, & Dossey, 1998). Comparing fourth and eighth grade student achievement, eighth grade students of teachers with a college major in mathematics outperformed students of teachers with majors in education or a field other than education, but fourth grade students of teachers with a major in education or mathematics education outperformed students of teachers with a major in mathematics.

Despite findings suggesting a math major or minor may matter, policy favors advanced degrees by offering increased compensation to teachers with higher educational attainment. Due in large part to policies regarding compensation, by 1996, 56.2 percent of public school teachers held advanced degrees (Skandera & Sousa, 2007). However, research on the level of education consistently fails to find a relationship between advanced degrees and student achievement (Harris & Sass, 2007; Hanushek, Kain, O'Brien, & Rivkin, 2005; Clotfelter, Ladd, & Vigdor, 2006).

In some cases, the findings are inconsistent. Using a data set that includes student achievement and teacher qualification data for all students in Florida (Harris & Sass, 2007), the impact of earning an advanced degree was positively correlated with student test scores on the Florida Comprehensive Achievement Test Norm-Referenced Test (FCAT-NRT) only in the case of middle school math ($b = 0.7246$, t(2.18), p<0.5). There

were no associations between attainment of an advanced degree and performance of elementary school teachers and a significant negative association between attainment of an advance degree and measures of productivity for high school math teachers ($b = -1.5889$, t(4.02), p < 0.01).

More commonly, no association is found between teacher attainment of an advanced degree and student achievement. Badgett (2011) investigated the association between the percent of teachers with a master's degree in a district and the percent of students attaining a label of "proficient" on the Texas Assessment of Knowledge and Skills. Data from 1,026 districts showed an association between the percentage of teachers with a master's degree and the percentage of students scoring at the "commended" level in a school district. However, the percentage of teachers with a master's degree did not impact the percentage of students labeled proficient.

In a study of over 300,000 students in grades 2-5 in Los Angeles, no relationship was found between teachers holding a master's degree and student achievement (Buddin & Zamarro, 2009). Using teacher qualifications and other factors to model student achievement gains, the investigation found no significant contribution from the degree held by the teacher. Similarly, a study using administrative data from the North Carolina Education Research Data Center for students in grades 3, 4, and 5 found no link between the level of degree held by a teacher and student achievement (Clotfelter, Ladd, & Vigdor, 2007).

A recent meta-analysis of teacher degree level and student achievement revealed little support for claims that degree matters (Aloe, 2013). Some evidence exists for the claim that there exists a positive correlation between teachers with a master's degree and

middle school mathematics achievement ($r = 0.16$). However, the research in this meta-analysis lacked consistency in methodology, level of analysis, aggregation, and available teacher and student data. The final conclusion is that there is no conclusion. The research is too inconsistent to make any generalizations.

The degree a teacher holds may impact student achievement differently at different grade levels. Although a master's degree in general does not seem to affect student achievement, a major, minor or advanced degree in mathematics may impact student achievement (Aaronson et al., 2007; Hawkins et al., 1998; Harris & Sass, 2007).

Additionally, greater educational attainment for a group of teachers may positively impact student achievement. This area is little researched, but one interesting study examined individual and group effects of educational attainment and found that higher levels of education for a group of teachers was positively associated with student gains in mathematics (Pil & Leana, 2009). Using measures that included the degree held by a teacher, the years of experience, and a teacher's score on 12 items developed by the Learning Mathematics for Teaching (LMT) project as measures of human capital, the authors modeled the effect of these and other factors on student achievement. The sample included 1,013 teachers organized into 239 grade teams at 202 schools from a large urban school district in the Northeast, and group variables were calculated by taking the average for the members of the group. Results suggest a link between group educational attainment and student achievement, with a one standard deviation increase in educational attainment for the group associated with a 5.5% gain in student achievement.

Teaching Experience. Compensation policies would suggest that experience is an important factor in teacher effectiveness, but the research is less clear. Teacher

experience is associated with student achievement in some studies (Clotfelter, Ladd, & Vigdor, 2006; Harris & Sass, 2007) , but often the effect is limited to the first few years of a teacher's career (Hanushek, Kain, O'Brien, & Rivkin, 2005; Rockoff, 2004).

Studies using state level data show inconclusive findings. Experience was found to increase student achievement in math, but only in the first two years of teaching and only on a test of mathematical computation skills, not mathematical concepts (Rockoff, 2004). Data from 1989 through 2001 for students in grades K-6 were used to determine the predictive nature of teacher experience on student achievement as measured by annual state assessments in two New Jersey school districts. Although computation scores increased as teachers gained experience over their first two years, the scores then decreased as more experience was gained. There was no change in scores on tests of mathematical concepts with greater teacher experience.

In another study utilizing panel data of student test scores and teacher assignments, at the state level, the Texas Schools Project found that students of experienced mathematics teachers outperformed students of inexperienced math teachers (Rivkin, Hanushek, & Kain, 2005). Using data from the Texas School Micro data Panel which contains data on growth in mathematics achievement on the Texas Assessment of Academic Skills for grades four through eight from 1995 to 2001, researchers found students of teachers with less than three years of experience did not experience as much growth as students of teachers with more experience.

A study using data from North Carolina used data from ninth and tenth grade end of course tests for algebra and geometry for students in four cohorts of tenth graders (1999/2000 -2002/2003). Gains in achievement were associated with experience only

through the first two years ($d = 0.0503$). Estimated coefficients did continue to rise, but none of the differences were statistically significant from the coefficient for 1-2 years of experience (Clotfelter, Ladd, & Vigdor, 2007). A similar analysis of test score gains in fifth graders throughout the state found significant returns from teacher experience, with students in classrooms with experienced teachers having mathematics test scores 0.10 SD higher in mathematics than students in classrooms with less experienced teachers.

Using controls for unobserved student, teacher and school heterogeneity through the use of multiple levels of fixed effects, Harris and Sass (2008) found that experience greater than two years positively affected student achievement (between 0.04 to 0.10 SD in achievement gains which is 0.02 to 0.06 SD in achievement level). Their data from Florida included all students' math and reading scores for grades 3-10 between 1999-2000 and 2004-2005. Using the state database, students could be matched with their teachers. The longitudinal nature of the data allowed for control for time-invariant teacher characteristics via fixed effects.

Using longitudinal data from Los Angeles and value added models with adjustments for student and teacher fixed effects, Buddin and Zamarro (2009) show that an increase in student achievement in mathematics was associated with an increase in teacher experience. Using five years of achievement data with students linked to classroom teachers, the sample included over 300,000 students in grades 2-5 who were matched to over 16,000 teachers. The authors note that the effect may be more attributable to the fact that teachers perform poorly in their first one to two years. Using contemporaneous and gain value added models, they found that a five year increase in

experience was associated with only a 0.8% increase in math scores, and that as teachers acquired more experience, student achievement increased at a decreasing rate.

As available data and statistical techniques have improved, research has shifted to more complex models frequently using statewide or large school district data. This more rigorous research lends validity to the claim that experience matters, particularly early in a teacher's career, but the effect of this experience is still seen as small.

Professional development. Professional development is often seen as a way to improve teacher effectiveness. The scope and findings of this research, however, do not frequently connect professional development to student outcomes. Instead, most of the literature on professional development focuses on descriptions of high quality professional development or relies on teacher self-reports of change in practice. According to the National Math Advisory Panel, there is not sufficient evidence to make judgments about what professional development will improve student achievement (National Mathematics Advisory Panel, 2008).

A number of scholars and educational organizations have weighed in on what they believe is high-quality professional development: intensive, content-specific professional development that links directly to state or district standards (Scher & O'Reilly, 2009). A review in the AERA Research Points (2005) suggest that professional development should be content focused, aligned with work experience, aligned with the curriculum, of adequate duration and emphasize observing and understanding student understanding of the content. Elmore (2002), in reviewing professional development recommendations, found similar recommendations, but noted additional areas such as

whether a particular program should be voluntary or mandatory, system wide or narrow in focus, connected to personnel evaluations, and focused on content or pedagogy.

Studies of professional development programs focused on mathematics frequently use teacher self-reports as outcome measures. For example, a review of sites participating in the federally funded Math and Science Partnership programs found evidence that teachers at these sites were more likely than a comparison group to report changes in their classroom practice. However, the study was limited because only about half the teachers in the treatment group and one-quarter of the teachers in the comparison group stayed in the study for its duration (Smithson & Blank, 2006).

Some studies examine the impact of professional development on teacher practice, but do not look further for an association with student achievement. For example, case study analysis of long-term professional development has shown evidence of change in teacher practice. However, these case studies focus on only a small number of teachers engaged in long-term professional development that includes interaction with experts, peer discussions, and classroom practice (Borko, Davinroy, Bliem, & Cumbo, 2000; Farmer, Gerretson, & Lassak, 2003). Large-scale studies of teacher practice also suggest that professional development influences teacher practice. These studies focused on long-term interventions that reflect reform practices find a change in teacher practice, although these studies rely on self-report, not observations of actual teacher practice (Wenglinsky, 2002, Cohen & Hill, 2000, Reys, Reys, Barnes, Beem, & Papick, 1997).

Few studies measure the impact of professional development on student achievement, observed teacher practice or teacher knowledge (Weiss, 2007). Linking professional development to student achievement is difficult. Research on the

professional development of mathematics teachers has shown a small effect on student achievement ($d = 0.33$), but was derived from cross sectional survey data from NAEP data intended for other purposes (Wenglinsky, 2002). Research on professional development has also shown no effect. For example, teachers involved in a summer program focused on constructivist methods reported an increase in student understanding of key concepts, but no change was noted on standardized tests (Simon & Schifter, 1993).

In one study (Cohen & Hill, 2000), participation in professional development was shown to have an effect on student achievement. Survey data from the state of California showed teachers who participated in professional development focused on the new mathematics framework reported increased student achievement. However, student achievement results were self-reports by teachers, and only 27% of participating teachers provided testing data. In the same study, the group of teachers who indicated they had participated in year long weekly workshops on the mathematics framework showed small gains in student achievement ($d = 0.13$) over comparison classes, but this gain was limited to optional standardized test data and was not evident on an alternate assessment.

Research on professional development for mathematics teachers struggles to link outcomes to student achievement and shows mixed results. Increased student achievement may or may not be linked to professional development. In a review of studies linking professional development and student achievement in math, Scher and O'Reilly (2009) found a statistically significant effect ($d = 0.38$) across 7 studies with 14 independent effect sizes. This same review showed a greater effect on student achievement for professional development lasting longer than one year as compared to only one year. However, this review did not include any studies with short-term

professional development. Also, there was a marked difference in the effect size in studies which focused on math content and pedagogy ($d = 0.56$) as compared to those which focused only on pedagogy ($d = 0.07$, ns). Although this review notes the included studies support current ideas around what is high-quality professional development, the authors noted that further high-quality studies that include valid, useful outcomes measures are needed.

One study attempted to ascertain the impact of professional development on student achievement in struggling schools. Administrative data from Chicago public schools, including basic demographic data for students, outcome data from ITBS scores of students in third through sixth grade in the fall of 1996 (about 131,300 students) and information about professional development were all used (Aaronson, Barrow, & Sander, 2007). Some schools in Chicago were placed on probation based on low test scores. A strict cut off was established, creating a discontinuity where schools on one side of the cut off were assumed to have similar unobservable characteristics to schools on the other side of the cut off. Thus students in schools on either side of the cut off could be compared to determine the treatment effect of additional professional development on teachers in schools on probation. Results show that the additional professional development did not impact student achievement.

Only one recent study examines the impact of the amount and type of professional development engaged in by teachers on student achievement (Harris & Sass, 2008). This study examined the impact on student achievement in grades 4 – 10 in Florida public schools. After accounting for teacher effects in the model, a positive effect was found to

be associated with teacher participation in professional development for middle and high school math only, but effect sizes were quite small ($d \approx 0.004$).

Research shows teachers believe participation in professional development changes their practice and improves student understanding, but direct links between professional development and practice or achievement have not been proven (Harris & Sass, 2008; Aaronson, Barrow, & Sander, 2007; Weiss, 2007). In addition, research has that examines the cumulative impact of many years of attending varied professional development programs has not been done.

Research on the many aspects of teacher quality that are thought to impact student achievement has evolved over time, with greater data availability and improved statistical analyses. Despite the variation in findings, the main claim in most of this research is that teachers have a measurable impact on student achievement, even though the impact of any isolated characteristic appears to be not statistically significant or quite small. As Aloe (2013) cautions, the effect of teacher degree level cannot and should not be considered in isolation. This applies equally to other teacher qualifications. Teachers are a complex combination of all their characteristics (measurable or not), and they work within the ecosystem of schools within their communities. In an effort to consider multiple characteristics of teachers, this review now looks at the construct of human capital as a conceptual framework for thinking about the complex of teacher characteristics that may impact student achievement in mathematics.

Human Capital as a Conceptual Framework

The instruction, and therefore learning, resulting from the use of a particular textbook is contingent upon who is using the textbook. Theories on textbooks and teaching suggest teachers interact with textbooks to design instruction. This instruction, therefore, results from the interaction between the teacher and the text and may be influenced by both the textbook and the teacher's ability (Brown, 2009).

Economists first conceptualized human capital as a way to account for residual economic output after considering the inputs of labor and monetary capital. Production functions used the latent variable of human capital to explain output not attributable to other inputs from capital such as money and supplies. Human capital was considered an intangible, yet was often equated with education. Becker (1964), one of the first to formalize the study of human capital, described it as an individual's abilities, knowledge and skills developed through formal and informal education and experience. Although not a direct measure of ability, human capital is a measure of a person's expertise, experience and preparedness to perform a role. Measures of human capital have historically been used in education as measures of knowledge and skill.

Initially conceived as an individual attribute, human capital has also been attributed to groups (Coleman 1988). The collective human capital of a group is thought to affect and possibly improve a group's performance, producing non-additive benefits (Faraj & Sproull, 2000; Smith, Collins, & Clark, 2005). This collective human capital can be a resource, with benefits to not only the group, but also individuals within the group (Argote, 1999).

These group attributes, such as the skills and knowledge that members bring to a team, are generally viewed as being held by individuals; yet the individual knowledge and collective knowledge are not independent (Coff, 1999). This interdependence of individual and collective knowledge or human capital can amplify differences in performance. Group membership increases individual performance to a greater extent when the members of the group are high ability (Day et al., 2003). This collective human capital has been conceptualized as the sum of individual expertise (Barrick, Steward, Neubert & Mount, 1998; DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004).

The term human capital has long been part of the research literature in education, but predominantly in research done by economists or printed in economics journals. The definition is dependent on the focus of the study. Human capital has been conceptualized in research on teachers as job manageability, collective responsibility and collective efficacy (Youngs et al., 2007) or content knowledge and the applicability of that knowledge to a specific task such as teaching math (Pil & Leana, 2009; Clark, 2010). Others do not explicitly define human capital, but examine whether additional training or experience increase human capital by examining the coefficients associated with the desired factor such as professional development (Harris & Sass, 2011) or national board certification (Clotfelter, Ladd, & Vigdor, 2010). In the case of group human capital where the group size varies, the average human capital for a group has been used (Pil & Leana, 2009).

As the idea of human capital has evolved and been used in various fields of research, including education, a broader view of human capital has emerged. No longer just a way to account for residual economic output, aspects of human capital are seen as

drivers of output. Many aspects of human capital, such as education, experience and certification are part of the policy debates about teacher and school effectiveness. Policy positions, educational research, and human capital theory, when considered together, imply that teacher human capital may impact the interaction of teacher and text, and, therefore, affect student achievement.

Stein and Kim (2009), in an effort to provide a theoretical basis for thinking about which text is best, or in their case, which text is best in a given situation, propose human capital be used in making this determination. Although they do not operationalize what is meant by human capital, prior research on human capital as well as policy and research in education provide a basis for beginning to examine the relationship between human capital, textbook choice, and student achievement.

Although human capital is sometimes used to mean something as basic as formal education, Becker (1964) and Stein and Kim (2009) all imply it is much more robust and nuanced than that. Stein and Kim (2009) describe human capital as the experience, expertise and preparedness of an individual to do a job. This conception of human capital fits into both broader discussions of human capital and existing educational research. The simplest component of human capital in general, and in an educational context, is experience. Experience in teaching is usually taken to be the number of years a teacher has been teaching, although this literature review suggests that experience in teaching mathematics may be more relevant.

Expertise from a human capital perspective is linked to formal training to do a specific job (Becker, 1964). Expertise can be conceptualized as certification and degree(s), as these represent formal training to do the job of teaching mathematics.

The last component of human capital suggested by Stein and Kim (2009) is preparedness. This is the most ill defined component of human capital. Becker's theorizing about human capital suggests that preparedness is achieved through on-the-job training. Teachers receive on-the-job training largely through various forms of professional development.

Summary

The plethora of research in many of the areas covered in this review offer little in the way of certainty, particularly in ways that might assist in selecting a textbook. Much of this research relates to the effectiveness of curricula and teachers. In many cases, the outcome measures in these studies are standardized test scores from annual large scale assessments. The quality of these tests to assess the full range of the mathematics standards is questionable. Furthermore, the lack of a consistent framework to evaluate mathematics textbooks, makes it even more difficult to use data to select a textbook.

Some insight might be garnered from research on textbook effectiveness. Although this research does not produce consistent results, studies that utilize additional assessments show more consistently that reform textbooks positively impact student achievement. However, assessing textbook effectiveness can only explain so much, as texts are used by teachers to teach, and teachers have been shown to impact student achievement. Teachers matter, but so far studies on teachers show that the effect of any particular teacher qualification is small at best.

Recently, some clarity has been brought to this research by better methodologies, particularly those that include multi-level models so less information is lost to aggregation.

Much of the research on teacher effectiveness has been advanced using economic modeling techniques. These researchers, and economists in general, frequently classify teacher qualifications such as experience, certification, degree, and professional development as aspects of human capital. Viewing teacher human capital as a characteristic of a school show promise, as Stein & Kim (2009) note that this is theoretically something that should be considered when trying to find the best textbook for a particular situation.

This leads to a number of problems to consider before data can effectively be used in aiding in textbook selection. First the human capital in a school needs to be operationalized. In addition, outcome measures beyond composite scores on annual standardized multiple-choice test will likely be needed to understand the effect of a textbook, human capital or their interaction. In the following chapters I describe the study I undertook to begin to study this complex relationship.

Chapter 3: Study Design

The continuing debate over mathematics curricula and the emphasis on increasing student achievement in mathematics suggests an unmet need in understanding the mechanisms that promote mathematics achievement. Evaluations of mathematics texts are not all the same and may produce different results. Indications are that teachers make a difference in achievement, yet little is known about the association of a textbook chosen by a school with student achievement, when moderated by the human capital in a school.

The goal of this study was to look at mathematics achievement in light of the textbook selected by a school and the human capital among the teachers in that school. This chapter includes a description of the data that was used, a description of how human capital components were developed, the hierarchical linear models developed to investigate what relationships exist among textbooks, teachers and achievement, and the logistic regression used to determine the association between school characteristics and textbook selection. Using standardized test data for 3,826 eighth grade students in 66 schools in a single large urban district, I investigated the following questions:

1. Is there a relationship between the textbook selected by a school and student achievement in mathematics?
   a. Is there a relationship between the textbook selected and the 8th grade mathematics composite score of the WKCE? If a relationship exists, what is the effect size?
   b. Is there a relationship between the textbook selected and the 8th grade mathematics strand scores for the six math strands (mathematical processes,

numbers and operations, geometry, measurement, probability and statistics, and algebra) of the WKCE? If a relationship exists, what are the effect sizes?

2. Does human capital moderate the relationship between textbook selection and student achievement?

    a. Does human capital moderate the relationship between textbook selection and the 8th grade mathematics composite score of the WKCE? If so, what is the effect size?

    b. Does human capital moderate the relationship between textbook selection and the 8th grade mathematics strand scores for the six mathematics strands of the WKCE? If so, what are the effect sizes?

3. Is there a difference in student performance on constructed response items associated with the textbook selected by a school?

    a. Is there a relationship between the textbook selected and the mean total points earned on the constructed response questions on the WKCE? Is so, what is the effect size?

    b. Does human capital moderate the relationship between textbook selection and the points earned on the constructed response items? Is so, what is the effect size?

4. Are proficiency rankings impacted by the textbook selected?

    a. Is there a difference in the proficiency rankings of students (between 6th and 8th grade), when they have been exposed to a reform mathematics textbook for at least one year of middle school?

5. Are there differences among the schools that selected different textbooks?

a. Is there a relationship between school demographic factors (racial/ethnic make up, percent low SES, percent ELL, percent special education) and textbook selection?

b. Is there a relationship between school human capital component scores and textbook selection?

Description of Setting and Population

The population for this study includes schools, teachers and students in a large urban, Midwestern school district that serves approximately 80,000 students in grades k-12. According to the district's web site, the majority of the students (56%) are African-American, followed by Hispanic (24%), and White (14%). The vast majority of students are classified as low income (83%); 67% of the students graduate from high school.

Within this school district, there are currently 156 schools, 67 of which are traditional public (non-charter) schools that serve students in grades six through eight. Students and families select the schools they wish to attend by submitting a list of preferences; the district then assigns students to schools based on those preferences.

Achievement in this district, as in many large urban districts, is below the state average. In the 2010 annual report on the results of the Wisconsin Knowledge and Concepts Exam (WKCE), 61% of fourth grade students were proficient or advanced in reading and 55% in math. Eighth grade scores were slightly higher in reading (63%), but lower in math (45%), and tenth grade scores were substantially lower, with 39% proficient or advanced in reading and 30% proficient or advanced in math.

The population for this retrospective study was students who took the WKCE in math as eighth graders in the fall of 2010 and can be associated with one of these 66

schools as a seventh grader in the fall of 2009. These students were sorted by the school they attended in 7th grade, and schools were characterized by the textbook selection and their teacher populations (the human capital of the teachers).

Study Variables

Textbook selection. During the 2006-2007 school year, all schools teaching grades six through eight in this district selected new textbooks that were then implemented in 2007-2008. Schools were given the choice of selecting from *Holt Mathematics Course 1-3*, *Glencoe Mathematics: Applications & Concepts*, or the *Connected Mathematics Project* (CMP).

These texts all claim to meet state standards and cover similar topics, yet they have different approaches to teaching those topics. Holt is the most traditional of these, with lessons that are typically intended to take one day to teach. In each lesson, students are presented with the objective of the lesson and frequently a real world context for an example that teaches a procedure. Students then work through a number of examples, guided practice and independent practice. Lessons include problem solving that connects directly to the skill or procedure taught in the lesson.

Glencoe is similar in design to the Holt text, with slight differences in the type of questions. In a typical lesson, students are presented with a lesson objective and frequently a real world context is presented. Students then work through a number of examples, guided practice and independent practice. Lessons also include problem solving and application problems that may require more than just using the skill or procedure taught in the lesson.

CMP bears little resemblance in appearance to the others, organized around units rather than daily lessons. These units include motivating questions, mathematical highlights (learning goals), several investigations intended to be completed in groups, homework problems (applications-connections-extensions), a mathematical reflection, and a unit project. In class or homework exercises that focus only on skills practice are not a focus.

For the purposes of this study, textbook selection is coded as a set of two dummy variables. The variable for reform textbook selection represents a school that selected CMP, the variable traditional textbook represents a school that selected Holt. The hybrid textbook, Glencoe, is the reference category.

Human Capital. Human capital has become a common phrase used to describe the effect of human workers on the output of a system. Its measure is rarely specified. Stein and Kim (2009) reference Becker's (1964) definition of human capital as the experience, expertise and preparedness of individuals for the role they are expected to perform. They further state "In a given organizational unit (grade-level, school), teacher human capital can be characterized as limited (most teacher have a low degree of experience or capability), high (most teachers have a high level of experience or capability) or variable" (Stein & Kim, 2009, p. 40). Although Stein and Kim (2009) define human capital as the experience, expertise and preparation of a teacher, they do not suggest how to operationalize it.

For the purposes of this study, the construct of human capital was operationalized as a combination of teacher characteristics available from school district data. Using information about years of experience, years of experience within the school district,

certification type, mathematics specialization, and professional development (both general and math specific that were offered by the school district), principal component analysis (PCA) was used as a data reduction technique, producing two components that I labeled experience and expertise. Because I was interested in the human capital within a school, specifically within the math teachers, component scores were aggregated to the school level.

The teacher experience component was made up of a combination of the teacher's number of years of experience, number of years with this district, and the highest degree attained. Year of experience in general and with the district were represented by continuous variables, based on data from the spring of 2010. The highest degree attained by a teacher was an ordinal variable with 1 representing a bachelor's degree, 2 a math major with any type of degree and 3 a master's degree or higher.

The teacher expertise component was made up of a combination of the type of certification a teacher has (emergency or grades 1-5, 1-8, 6-9, 6-12), whether the teacher had some type of math specializations (0 for no math, 1 for any type of endorsement or subject specific math license), and the average number of hours of mathematics professional development per year over the three year period 2007-2010. The overall number of hours of professional development was dropped as a result of the PCA.

Student achievement. A variety of measures of student achievement were used as outcome variables. All student data came from the fall 2010 administration of the math portion of the WKCE administered to 8th graders. Students each received a composite score, strand scores, and a proficiency ranking. This study used all of these as outcome variables. Composite scores were simply a numeric score given to represent a students

overall performance on the WKCE. Strand scores were given in the form of a standards performance index (SPI) for each strand. The SPI is an estimate of how many items covering a topic or strand a student would correctly answer if there were 100 such items on the test. It is not simply a percentage, but a score based on performance on test items measuring the content standard and related performance on other test items (Administrator's Interpretive Guide, 2011-2012).

In addition to these scores that are routinely provided to schools and teachers, I also used the mean number of points earned on the constructed response items on this administration of the WKCE. This version of the test had 7 constructed response items, meaning the student had to write in an answer and may have been asked for an explanation. (see Appendix A for sample questions). Students were not usually given a separate score for these items as they fit into different strands and were used along with the multiple-choice questions to produce a strand score. The total points a student earned was calculated by adding the seven scores together; then z-scores were calculated.

The final outcome measure used in these analyses was the proficiency rank of students. Cut scores were established by the state to rank students on a scale of 1 to 4, 1 for minimal performance, 2 for basic, 3 for proficient and 4 for advanced. Because the tests were not vertically aligned across grades and scores were not intended to be used to measure growth, proficiency rank was used to examine whether students' rank increased between 6th and 8th grade.

Data Analysis Procedures

Because of the nested nature of the data, with outcome measures at the student level and predictor variables at the school level, Hierarchical Linear Modeling (HLM)

was used to analyze the data. This addressed the analytic challenge of considering

varying levels of data, allowing for level-two variables to explain between group variance

in the level-one intercept (Raudenbush & Bryck, 2002). HLM was also appropriate given

my conceptual framework, as it enabled me to consider how school level characteristics,

in this case the selection of a type of text and components of human capital, were

associated with student level test scores.

I used SPSS 19 and HLM 7 software to analyze the data. I used my research

questions as a guide in constructing the multilevel model, and below I describe the

process I used in answering each research question.

Is there a relationship between the textbook selected by a school and student achievement

in mathematics?

I began by creating models of each of the outcome variables with no predictor

variables, commonly referred to as the null model. These models, equivalent to a one-

way analysis of variance, allowed me to estimate the variation in student test scores at the

individual level and across schools. The null model at level-one was given by the

equation:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

where Y_{ij} is the outcome variable (student test score) for student i in school j, β_{0j} is the

intercept for school j, and r_{ij} is the individual, or level-one, residual. The level-two

model (school level) was given by the equation:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

where intercept β_{0j} becomes the outcome variable, γ_{00} is the grand mean effect or fixed intercept at level-two, and u_{0j} is the level-two residual. This results in a combined level-one and level-two equation:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

where Y_{ij} is the outcome variable (student test score) for student i in school j, γ_{00} is the grand mean effect or fixed intercept at level-two, u_{0j} is the residual at the school level and r_{ij} is the residuals at the student level. The null model allowed me to determine that the student test scores were statistically different across schools. I did this by determining the i correlations, which I calculated by dividing the variance in level-two (between school) residuals by the total variance of level-one and level-two. The null model also provided me with a baseline from which to gauge the reduction in variance in subsequent models.

In the next step of the modeling process, I created the individual or level-one models. I entered only one additional variable, an indicator of socio-economic status. The American Psychological Association recommends accounting for differences in SES in all sociological and educational research (APA, 2013). This is suggested because of the variety of effects associated with low SES. For example, a teacher's years of experience and quality of training is correlated with children's academic achievement (Gimbert, Bol, & Wallace, 2007). Test results are also affected by SES. Children with higher SES backgrounds are more likely to be proficient on tasks of addition, subtraction, ordinal sequencing, and math word problems than children with lower SES backgrounds (Coley, 2002).

Socio-economic status was entered as a dichotomous variable, based on the school district supplied variable of "SES indicator" such that a value of 1 corresponds to a student receiving free or reduced lunch. The level-one model then became

$$Y_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij}) + r_{ij}$$

In the final step of the modeling process to answer this question, I developed the school-level or level-two model. I was interested in which of the level-two variables, dummy variables for textbook type, explained some of the variance in student test scores. Using the scale core and six stand scores on the WKCE as outcomes in seven models, I modeled the level-one intercept with this set of dummy variables, which I added to the previous equation:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(TRADITIONAL_j) + \gamma_{02}(REFORM_j) + u_{0j}$$

Resulting in the mixed level-1 and level-2 model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}(TRADITIONAL_j) + \gamma_{02}(REFORM_j) + \beta_{1j}(SES_{ij}) + u_{0j} + r_{ij}$$

In this model, any significant reduction in between-school variance was due to the explanatory power of the group-level variables relating to textbook type.

Does human capital moderate the relationship between textbook selection and student

achievement?

In order to construct the complete level-two models necessary to answer this question, I first used SPSS 19 to develop school level human capital components. Starting with the seven teacher quality related variables (years of experience, years of experience within this school district, highest degree earned, average number of hours of professional development attended per year over a three year period, average number of hours of mathematics professional development attended within the district per year over that same three year period, grade levels covered by certification, and the presence of any type of math specialty such as a math major or math endorsement). Correlations among these variables led to a decision to use a data reduction technique to produce components that would represent a teacher's human capital (a measure of their perceived ability or preparation to perform the tasks of their job).

Using two components, experience and expertise, I calculated individual component scores for each teacher in the study. I then aggregated these scores to the school level. This was done for theoretical reasons. In this study I viewed human capital as a school characteristic because textbooks are selected and used school-wide and I was interested in how school factors effected the association between textbooks and student test scores.

The first component, *experience*, is comprised of the years of experience for a teacher, the years of experience with this school district, and the highest degree attained. All three of these factors relate to experience. Two are direct measures of experience; the third, the highest degree earned, correlates highly (0.54) with total years of experience,

possibly because both of these measures are used in determining salary, so are incentivized, resulting in many teachers pursuing a master's degree as they gain more experience.

The second component, *expertise,* is comprised of the amount of mathematics professional development attended, the grades covered by the certification, and whether the teacher has some type of math endorsement. The amount of professional development was a measure of the average amount of mathematics professional development within the district, attended by the teacher, per year, over a three-year period from 2008 to 2010. The grades covered by certification was an ordinal variable in which 0 represents a teacher with an emergency certification, 1 represents a teacher certified to teach elementary school, 2 represents a teacher certified to teach elementary and middle school, 3 represents a teacher certified to teach only middle school, and 4 represents a teacher certified to teach middle and high school. The final factor in this component was a dichotomous variable indicating a teacher with some type of math endorsement. This included teachers with a math degree, teachers with a subject specific certification, and teachers given a special endorsement, usually as a result of majoring or minoring in mathematics. Over time the requirements for and availability of a math endorsement have changed, so all teachers with a subject specific license and math major or minor were included in the group identified as having a math endorsement.

I then went back to my HLM models from the previous analysis and augmented them to reflect my model that textbook selection is moderated by teacher human capital in affecting student test scores. Having already modeled the level-one intercept for each outcome variable with the level-two textbook selection variables (traditional and reform),

I added each group-level principal component scores (experience and expertise) and the interaction variables (traditional x experience, traditional x expertise, reform x experience and reform x expertise). This resulted in the final model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}*TRADITIONAL_j + \gamma_{02}*REFORM_j + \gamma_{03}*EXPERIENCE_j + \gamma_{04}*EXPERTISE_j + \gamma_{05}*TRADITIONAL_j*EXPERTISE_j + \gamma_{06}*TRADITIONAL_j*EXPERIENCE_j + \gamma_{07}*REFORM*EXPERIENCE_j + \gamma_{08}*REFORM*EXPERTISE + \gamma_{10}*SES7_{ij} + u_{0j} + r_{ij}$$

Is there a difference in student performance on constructed response items associated

with the textbook selected by a school?

Given the research findings that student achievement is positively associated with using a reform curriculum when the outcome measure is open-ended, I performed a similar analysis as above but used the student performance on constructed response items as the outcome measures.

Individual scores on constructed response items on the math WKCE were not typically reported separately because the items do not cover only one strand, so the scores on these items were incorporated into the appropriate strand scores for the purposes of reporting student achievement. I obtained the scores on the seven individual constructed response items, accounting for 11 total possible points. I calculated the total points earned by a student, then calculated a z-score for each student.

As before, I first constructed the null model, then added in SES and the text variables traditional and reform. I then constructed the full level-two model, as shown:

$$ZTOTALSCORE_{ij} = \gamma_{00} + \gamma_{01}*TRADITIONAL_j + \gamma_{02}*REFORM_j + \gamma_{03}*EXPERIENCE_j + \gamma_{04}*EXPERTISE_j + \gamma_{05}*TRADITIONAL_j*EXPERTISE_j + \gamma_{06}*TRADITIONAL_j*EXPERIENCE_j + \gamma_{07}*REFORM_j*EXPERIENCE_j + \gamma_{08}*REFORM_j*EXPERTISE_j + \gamma_{10}*SES7_{ij} + u_{0j} + r_{ij}$$

Are proficiency rankings impacted by the textbook selected?

The above models are limited in that they cannot represent growth due to the limitations of the outcome variables. In an effort to examine an indicator of growth, I compared the change in proficiency rank from grade 6 to grade 8 between students who experienced a reform text for at least one year (in grade 6 and/or grade 7) to all other students.

I regressed longitudinal proficiency on variables as in the prior analysis. In the level-one model, the outcome variable proficiency was examined on three occasions (proficiency ranks for grades 6, 7, and 8 for each student). I assumed linear growth in proficiency scores because of the limited data – 3 data points per student. Only students with all three data points were included in this analysis.

At level 2, I added the variable of SES. The final level-three model included a variable for textbook type (traditional and reform), the two components of human capital, and the interaction variables, given the resulting equation:

$$PROFICIENT_{tij} = \gamma_{000} + \gamma_{001}*TRADITIONAL_j + \gamma_{002}*REFORM_j + \gamma_{003}*EXPERIENCE_j + \gamma_{004}*EXPERTISE_j + \gamma_{005}*TRADITIONAL_j*EXPERTISE_j + \gamma_{006}*TRADITIONAL_j*EXPERIENCE_j + \gamma_{007}*REFORM_j*EXPERIENCE_j + \gamma_{008}*REFORM_j*EXPERTISE_j + \gamma_{010}*SES7_{ij} + \gamma_{100}*OCCASSION_{tij} + r_{0ij} + u_{00j} + e_{tij}$$

Are there differences among the schools that selected different textbooks?

In an effort to better understand the mechanisms that might influence a school in selecting a reform textbook and therefore aid in interpreting the results, I used logistic regression to examine the association between school factors and textbook selection. Using two dichotomous outcome variables, one representing the choice of a reform text and one representing the choice of a traditional text, I examined the relationship between

the outcome and school characteristics of student demographics and teacher human capital.

In the first analysis, using the selection of a reform text as the outcome, I used a variety of school demographic data as independent variables. These variables included the percent of students of each race/ethnicity (Black, Hispanic/Latino, White, Asian, and Native American), the percentage of low SES students, the percentage of ELL students, and the percentage of LD students in a school. Using binary logistic regression, I investigated whether these demographic factors were associated with selecting a reform text, then repeated the analysis using the selection of a traditional text as the outcome variable.

As well as the demographic make up of the students in a school, I also wanted to know if the human capital component scores were associated with selecting each type of textbook. Using the aggregate human capital component scores for each school as the independent variables, I again used logistic regression to investigate this association.

Chapter 4: Results

I divide this chapter into six parts. First I present information on the cleaning of the data. In the second part, I present the results of the principal component analysis (PCA) of teacher characteristics resulting in component scores used to represent teacher human capital. I then present analysis for the research questions about student achievement resulting from the use of a particular type of textbooks mediated by teacher human capital, first using HLM to investigate component and strand scores, then examining data relating to Constructed Response items, and finally an analysis of growth of proficiency ranking. Lastly, I present the logistic regression analysis examining the association between choosing a reform textbook and schools as described by student and teacher data.

Data Cleaning

The data for this study was supplied by the administrative offices of a large urban school district. As such, it included data that was not intended to be included in the study. I was interested in student achievement on the eighth grade mathematics section of the Wisconsin Knowledge and Concepts Exam (WKCE). The WKCE is the annual exam given to all students in grades three through eight and grade ten. It includes a reading/language arts assessment and a math assessment. This exam is given each October.

I received some data such as school assignment and SES for 4,571 students; all eighth graders enrolled in the district in the fall of 2010. Of those, I received testing data for 4,436 of these students. I was able to determine seventh grade school assignments for

4,347 of these students. I was only interested in students in traditional public schools, so I dropped students attending charter schools, resulting in a total of 4,188 cases. After consulting further with the school district, I was able to determine textbook used by 3,826 students, resulting in 362 students being dropped from the study. These 3,826 students were spread over 66 schools.

For the teachers, I only included teachers who taught middle school mathematics during 2008-2010 at one of the 66 schools included in this study. Due to an agreement with the school district, I was not able to match students directly with teachers. This was not a problem, as the intent of the study was to examine human capital as a characteristic of a school. Therefore aggregate teacher component scores were used to represent a characteristic of a school in the analyses.

Principal Component Analysis of Teacher Data

The school district provided teacher data included seven variables: years of experience, years with the school district, certification, highest degree, amount of non-math professional development, math professional development, grades covered by certification. These seven variables were subjected to principal component analysis (PCA) using SPSS version 19. An initial evaluation of the data showed a number of correlation coefficients of 0.3 or greater (see Appendix B). The factorability of the correlation matrix was supported by the statistically significant Bartlett's test of sphericity (KMO = 0.602, $p < 0.01$) (Bartlett, 1954; Kaiser, 1974).

Initially, using Oblimin with Kaiser Normalization rotation and a minimum of 0.30 loading criterion, the first 2 factors explained 56.67% of the variance. However the data were reanalyzed by specifying a Varimax rotation. This allowed for easier use of the

components in additional analyses, and the Oblimin rotation was not necessary because the resulting factors were not highly correlated ($r = 0.05$).

Initial PCA with Varimax rotation resulted in three components with eignevalues greater than 1 (Guttman, 1954). These components explained 71.32% of the variance. Most of the items loaded strongly (above 0.4) on the first 2 components. A visual inspection of the scree plot (Appendix B) did not show any distinct breaks, providing no additional guidance on the appropriate number of components to retain. Using a parallel analysis engine to compute values for 100 randomly generated data matrices of the same size with a 95% confidence interval (Patil et al., 2007), a parallel analysis confirmed the specification of two components as two eigenvalues exceeded the corresponding computed values (see Appendix C).

An inspection of the communalities revealed six communalities above 0.47 and one communality (non-math professional development) of only 0.13. In addition, the variable non-math professional development did not load on the two retained components. Based on these factors, I deleted the non-math professional development as a variable and subjected the remaining six variables to PCA, forcing a two-component solution. The new two-component solution explains 64.81% of the variance. These two components I have labeled Experience and Expertise (see Appendix C).

Individual teacher scores for the two components were calculated and then aggregated using the means in order to find a school score for each component. These scores were then used as two variables, *Experience* and *Expertise*, in the HLM that follows.

Hierarchical Linear Modeling (HLM)

For this analysis, I was interested in examining the extent to which textbook type, teacher human capital, and the interaction of these are associated with middle school student's mathematics achievement as measured by the scale score and strand scores on the WKCE. Because I was interested in the effect at the school level, I focus on the between school variance in test scores. With regards to the conceptual framework of the interaction between textbook and human capital (Stein & Kim, 2009), the model encapsulates the effect of textbook type, components of human capital and the interaction of these on math achievement.

Using HLM 7, I first estimated the unconditional or null model. From that model I was able to calculate intra-class correlations (ICC) for each outcome variable (scale score and 6 strand scores), as shown in Table 1 below. Calculations of ICCs can be found in Appendix D. The ICCs indicated that about 20% of the variances in outcome measures were associated with between school differences.

Table 1

Intra-class correlations

Outcome variable (grade 8 WKCE scores)	Intra-class correlations
Scale score	0.20
SPI – math processes	0.22
SPI – number and operations	0.20
SPI – geometry	0.20
SPI – measurement	0.22
SPI – probability and statistics	0.20
SPI – algebra	0.20

I then regressed the achievement outcomes on three blocks of variables to explore what relationship, if any, these variables had on math achievement. In the level-1 model, I regressed the outcome variables on student SES because SES has been shown to have a close association with standardized test performance. In the second model (a 2-level model), I added dummy variables for textbook type. In the full model, I added the two components of human capital and the four interaction variables.

In this section, I discuss the effects of each of these blocks of variables on student achievement. Table 2 presents all the coefficients of the 2-level model with level-two variables for type of textbook selected. Table 3 presents all the coefficients of the complete between schools models. Appendix D gives descriptive statistics for variables and outcome measures, Appendices E-G display the step by step modeling process for the scale score and component scores (A-F) on the WKCE, starting with the SES (level-1) model through the final model (level-2), including the model with only textbook type considered. Included in Appendix F are coefficients and robust standard errors for all models. Robust errors were used because they are larger and therefore give a more conservative estimate. Also in Appendix G are the variance explained calculations to which I refer throughout the results. The within schools variance explained calculation for SES was determined at level-one by subtracting the variance of the SES model from the null model and then dividing by the variance in the null model (Hox, 2002). In level-two, the between schools variance was determined by subtracting the variance of the level-two component from the variance in the SES model (level-1 model) and then dividing by the variance in the SES model.

Table 2

Final estimation of fixed effect coefficients, 2-level model, textbook type only

	Scale score	Math A Math processes	Math B Numbers-operations	Math C Geometry	Math D Measurement	Math E Probability Statistics	Math F Algebra
Intercept	509.815***	33.716***	41.238***	57.359***	37.754 ***	47.609***	51.544***
Textbook	**variables**						
Traditional	13.040	5.113	4.526	4.490	3.461	3.903	4.824
Reform	23.756**	9.274**	7.414*	7.708**	6.549 *	7.446 **	6.492*
Reliabilities							
Intercept	0.868	0.877	0.868	0.870	0.876	0.863	0.867
Deviance	41167.031	32696.281	32634.291	32416.616	41121.136	31847.772	32499.493

*p < 0.05, **p < 0.01, ***p < 0.001

Table 3

Final estimation of fixed effect coefficients, 2-level model, full model

	Scale score	Math A Math processes	Math B Numbers-Operations	Math C Geometry	Math D Measurement	Math E Probability Statistics	Math F Algebra
Intercept	511.647***	34.386***	41.870***	58.004***	38.192***	48.050***	52.169***
Textbook Variables							
Traditional	10.745	4.261	3.785	3.721	2.858	3.350	4.136
Reform	27.145**	10.095**	8.532*	8.388**	6.951*	8.212*	7.064*
Human Capital Components							
Experience	13.486	4.914	5.024*	4.287	3.104	3.302	3.346
Expertise	5.345	1.922	1.487	2.317	1.447	1.372	2.079
Interaction Terms							
Trad x experience	-11.456	-3.393	-4.379	-3.999	-2.279	-3.038	-5.137
Trad x expertise	-2.670	-0.868	-1.826	-0.306	-0.873	-0.238	-1.449
Reform x experience	-11.727	-2.671	-2.501	-4.683	-3.275	-3.173	-5.473
Reform x expertise	-14,820	-4.008	-4.944	-3.924	-2.666	-3.225	-3.983
Reliabilities							
Intercept	0.868	0.876	0.868	0.868	0.879	0.865	0.867
Deviance	41121.136	32662.610	32601.601	32383.433	30803.49	31817.780	32467.296

*p < 0.05, **p < 0.01, ***p < 0.001

The first block of level 2 variables, textbook type Traditional and textbook type Reform explained about 15% of the variance in both the model which only includes textbook type and in the full model. In all of the models (regardless of outcome variable), students who attended schools using reform textbooks in seventh grade scored about 27 scale score points higher in eighth grade than students using other texts, with component scores 7-10 points higher, and proficiency ranking 0.4 points higher. Outcomes for students who used a traditional text in seventh grade were not significantly different than those students using a hybrid text.

The next block added to the model included the variables for human capital (the components experience and expertise) and the interaction variables. The only additional variable to show statistical significance ($p < 0.05$) in the complete model was Experience and only when the outcome variable was the component score of student performance index (SPI) for strand B (number and operations) ($d = 0.26$). The interaction terms have negative coefficients. None of these reached statistical significance.

These results show there was a difference in achievement for students who were exposed to reform textbooks in 7th grade and then tested in the fall of 8th grade (Appendix H). This difference was statistically significant, with effect sizes ranging from small ($d = 0.12$) to medium ($d = 0.52$). The largest effect size ($d = 0.52$) represents the effect size for mathematical processes. These effect sizes were calculated by dividing the treatment gamma by the square root of the sum of the between school variance and the within school variance from the null model (Raudenbush et al., 2005).

Based on the research literature, which shows a difference in student achievement on open-ended assessments, and in an effort to try to uncover more information about the impact of choosing a reform textbook, I segregated out performance on Constructed Response items for further analysis.

Constructed Response

I began by focusing on the total number of points a student earned for Constructed Response items. On the version of the WKCE used in this analysis, there were 7 scoreable answers, worth a total of 11 points. The mean for points earned, the percent of students who left all question blank, the percentage of students who scored a total of zero points (but did attempt at least one question), and the percent earning a perfect score are summarized below.

Table 4

Constructed response summary

Text type	Mean points earned	No response	Zero points earned	11 points total
Traditional	2.39	1%	27.5%	0.3%
Hybrid	1.92	1.9%	33.1%	0.2%
Reform	3.19	0.3%	16.9%	1.1%

HLM was undertaken in order to examine the extent to which textbook type and human capital components are associated with student achievement represented by a z-score for the constructed response total. In the model incorporating just the textbook type, the variable reform was statistically significant ($p < 0.001$), indicating an increase in total points earned on constructed response items of 0.64 points more than students using the

hybrid text. In the model incorporating textbook type and human capital (but not the interactions), reform text and experience were both statistically significant ($p < 0.001$ and $p < 0.05$ respectively), with the experience component adding an additional 0.138 points to a student's total. An examination of the full model, including the interaction terms, showed the variables reform and experience to again be the only statistically significant variables (Appendix H). Once again, the selection of a reform text by a school was statistically significant, with a medium effect size ($d = 0.62$). Experience was also statistically significant, but with a small effect size ($d = 0.26$).

Growth of Proficiency Rank

In the end, the score that mattered most to children and schools was the proficiency ranking of the students. Using HLM, I investigated a growth model for proficiency, comparing students who experienced a reform text for at least one year (grade 6 and/or grade 7) with all other students. Students received a ranking of 1 for minimal, 2 for basic, 3 for proficient and 4 for advanced, dependent on their scale score. Cut scores are set annually for each administration of the test.

I regressed longitudinal proficiency rankings on three bocks of variables as in the prior analysis. In the first model, I regressed growth in change in proficiency rank on SES. In the second model, I added dummy variables for textbook type (traditional and reform), and in the third model added components of human capital and the interaction variables. (See Appendix I for model summary).

The model assumes linear growth because of the small number of occasions included in the model. Each student had three measurements, 6th grade proficiency rank, 7th grade proficiency rank and 8th grade proficiency rank. Only data for students with all

three measurements were included in this analysis. This resulted in a model that includes

3346 students in 66 schools. The tables below (Table 5 and Table 6) show coefficients

and intercepts for each level of the model. Once again, the only variable that was

statistically significant was whether a school had selected a reform textbook.

Table 5

Final estimation of fixed effect coefficients 2-level model, with textbook type only

	Proficiency Rank
Intercept	2.387***
Textbook Variables	
Traditional	0.203
Reform	0.402**
Reliabilities	
Intercept	0.852
Deviance	10095.905

** $p < 0.01$, *** $p < 0.001$

Table 6

Final estimation of fixed effect coefficients, 2-level model, full model

	Proficiency Rank
Intercept	2.420***
Textbook Variables	
Traditional	0.167
Reform	0.412**
Human Capital Components	
Experience	0.225
Expertise	0.124
Interaction Terms	
Trad x experience	-0.261
Trad x expertise	-0.064
Reform x experience	-0.294
Reform x expertise	-0.158
Reliabilities	
Intercept	0.851
Deviance	100098.779

** $p < 0.01$, *** $p < 0.001$

Despite the limited effect of the chosen variables and models, the consistency of the results suggests that a school selecting a reform textbook may be positively associated with student achievement. However, the complexity of the system of schooling, with many variables and interactions, which have been modeled here, indicates that further research is warranted. In an initial effort to aid interpretation of these results, I investigated whether there is an association between selecting a reform textbook and certain school characteristics that can be found in the student and teacher data.

Logistic Regression

Logistic regression was performed in order to assess the impact of school demographics or teacher human capital components on the likelihood that schools would choose a reform textbook. The first model contained the school demographic data (race/ethnicity, SES, ELL, and special education status) was not statistically significant (see Appendix J).

A second model containing school human capital components (experience and expertise) was statistically significant indicating that the model was able to distinguish between schools that selected a reform textbook and those that did not (see Appendix K). The model explained between 12.9% (Cox & Snell R^2) and 20.5% (Nagelkerke R^2) of the variance in selection of a reform textbook and correctly classified 78.8% of schools. Of the two variables, only one made a statistically significant contribution (expertise). Expertise had an odds ratio of 7.42, indicating that an increase of 1 in the component expertise suggests a school was 7 times more likely to select a reform textbook. But, the 95% confidence interval is quite large, ranging from 1.68 to 32.75, so the probability of correctly predicting that a school will select a reform textbook is only about 33%.

Components of human capital was not consistently associated with increased student achievement (as shown in the HLM), but the component of expertise was associated with the selection of a reform textbook, and selection of a reform textbook was associated with greater achievement.

Chapter 5: Discussion and Conclusions

This study sought to determine whether the type of textbook selected by a school, moderated by the human capital of the teachers teaching mathematics and the interaction of those variables is associated with increased student mathematics achievement on the state-wide standardized eighth grade test. Hierarchical linear modeling was used to investigate the model relating textbook selection, components of teacher human capital and their interaction, based on theory proposed by Stein and Kim (2009). Contrary to the initial hypothesis based on the theory, the interaction of textbook selection and components of human capital were not found to be significant. Selecting a reform textbook was associated with student achievement but accounted for very little of the variance. Logistic regression was used to investigate the association between various school demographics and choosing a reform textbook but did not find any significant variables. However, logistic regression used to investigate the association between components of human capital and choosing a reform textbook were significant. This chapter elaborates on the significance of these findings, the limitations of this study, and practical implications. I also include suggestions for further research.

Background for the Study

The publication of the NCTM standards beginning in 1989 resulted in the development of reform curricula. These curricula, focused on developing mathematical understanding and incorporating new pedagogical techniques, were unlike commonly used curricula, and researchers were pressured to provide evidence of their impact on student achievement (Schoenfeld, 2002). Soon after, the National Research Council

reviewed existing research on the effectiveness of mathematics curricula and found insufficient evidence to support any of the curricula in any of the studies, citing an insufficient number of studies, limitations in the methods used and the uneven quality of the studies (NRC, 2004). At this same time, the passage of No Child Left Behind pushed policy towards an outcome orientation, with an emphasis on testing results over inputs. Effectiveness has become synonymous with impact on standardized test scores, whether talking about teacher effectiveness or curriculum effectiveness.

Since the introduction of reform curricula, effectiveness studies of mathematics curricula have come to varied, sometimes conflicting conclusions. Traditional textbooks may produce higher standardized test scores on statewide annual assessments or other predominantly multiple-choice assessments (Schneider, 2000). Reform textbooks may be associated with increased performance on open-ended or researcher developed assessments. (Ridgeway, 2003; Eddy et al., 2008). Or there may be no difference in performance (Martin, et al., 2012).

For decades, research on teacher effectiveness has suggested that teachers matter and that they impact student achievement, but little of the variance has been explained by examination of individual qualifications. The more recent studies that use more complex analytic tools show fairly consistent results, as far as significance, but still find small effect sizes. Research shows that experience matters (Darling-Hammond, 2001; Clotfelter, Ladd, & Vigdor, 2007); subject specific preparation matters (Pil & Leana, 2009; Aloe, 2013); and certification matters (Paige, 2002; Betts, Zau, & Rice, 2003), particularly when examining associations with middle school mathematics achievement. What is still unclear is what explains the remainder of between teacher variance.

Studies of different types of effectiveness, whether teacher or textbook effectiveness, do not capture the complexity of their interaction. This study attempted to extend the research on textbook effectiveness into a situated investigation of a single large urban school district in which individual schools within the district selected from three middle school mathematics textbooks. The textbook adoption process in this district, with schools choosing from three types of textbooks, is quite unique, so research of this type is uncommon. My interest in the association between textbooks and human capital is rooted in the efforts in the mathematics education community to develop theory around the interaction of teachers and mathematics curriculum materials (Remillard et al., 2009). This study was an effort to look at that interaction in the hopes of furthering this discussion and impacting policy guiding textbook selection. In particular, this study examined the associations among textbook selection, human capital, the interaction of these, and student achievement.

It is important to acknowledge this study's limitations. First, my data came from only one school district. As a result, it is not possible to generalize beyond this district. Secondly, I was only able to utilize data that this school district provided to me. Therefore, these data only included information on which textbook each school selected to purchase. It did not include any information about whether the text was implemented with fidelity or information about the school or classroom culture beyond basic demographic information. Therefore, it is not possible to make any generalizations about how mathematics may or may not differ between schools that selected a reform textbook and others. Thirdly, because schools chose books and were not randomly assigned, this research cannot be taken as an experiment. Therefore, no generalizations can be drawn

about what would happen if schools were assigned a textbook instead of choosing one. Whether the results would persist if schools were assigned books is impossible to know from this study. Finally, the measures of student achievement in this study, the scale score and strand scores from a statewide annual assessment implemented to meet the requirements of NCLB may not be a robust measure of achievement or student learning. As a result, difference in student learning may be difficult to discern from these standardized and predominantly multiple-choice tests, resulting in statistical analyses that may give few significant results.

In addition to the limitations associated with the selection of textbooks and the outcome measures (test scores), the development of components to represent human capital were limited by the available data. For example, data on professional development was limited to professional development taken through the school district; it did not include any additional opportunities pursued by individual teachers. The variable for mathematics endorsement was a dichotomous variable, suggesting all types of math endorsements may be equal. This does not reflect the variability in licensing and endorsement requirements that have changed over time. And, the aggregation of data to create school wide measures of human capital components resulted in the loss of some variability. The aggregation was intentional in the development of the model, but may limit the results from the model.

And, perhaps the biggest limitation is the modeling used to answer the research questions. As this was an initial foray into modeling the association among textbooks, teachers, the interaction between them, and student achievement, the model itself is limited. The model is based in part on theory around textbooks, teachers and interactions,

yet as an initial conception, it was unlikely to adequately describe all the data in this study. It will need refinement in the future.

In addition, as has been noted in other research, inferences about effectiveness from statistical models rely on assumptions about schools, teachers, students, families, and communities in producing student achievement. For example, models rarely account for non-random assignment of students to schools and classrooms (Corcoran, 2010; Linn, 2008). The model developed in this study did not account for non-random assignment of students to schools; there is no variable that considers whether the majority or the students in a school were assigned by the district or chose that school as their first preference, whether it is a neighborhood or city-wide school, whether the school has space for all students that request this school or has to turn some students away.

Another criticism of current statistical models is the inability to completely capture the complexity of schooling. Schools have multiple teachers and variations in parental involvement, after school tutoring programs, available materials, time spent on mathematics instruction, and class size (Baker et al., 2010; Brown, 2005; NRC, 2010). This model is no different. I did not attempt to include all aspects that affect student achievement but, based on the literature, focused only on aspects of schools directly related to textbook selection and human capital that showed promise.

Discussion of Results

I divide the discussion of the results of this study into two main sections: a discussion of the results, with special attention paid to statistically significant variables and a discussion of the hierarchical linear model used.

Significant Results. The most consistent result in this research is the statistically significant association between choosing a reform textbook and student scores whether measured by a scale score or strand scores. In most cases, this was the only significant variable in the model, with effect sizes ranging from $d = 0.12$ to $d = 0.52$, with most around $d = 0.4$, indicating a small effect. Despite being small, these effect sizes are typical of textbook or curriculum effects research (Riordan,Noyce, 2001; Riordan, Noyce & Perda, 2003; Schneider, 2000; Ridgeway et al., 2003; Cai et al., 2011). The size of the effect may be typical of this research, but prior research does not predict this with certainty, as results are inconsistent in the literature (Resendez, 2005; Martin et al., 2013; Callow-Heusser et al., 2005). Also, these results suggest that reform texts are more effective, but only when schools choose to use these texts.

An examination of the full model, which included the human capital components and the interaction terms, showed only one other significant result. When the student achievement measure was the strand score for numbers and operations (strand B), the effect size for the component of experience was 0.27. This strand may be particularly prone to an effect from the length of time a teacher has been teaching. Research shows teachers are more effective after 2-5 years of experience, and may be more adept at supplementing computational skills.

This positive effect from the selection of a reform text is also noted in a comparison of the mean total points earned on the constructed response items. HLM was undertaken to investigate the association between textbook selection, human capital and the interaction terms with the mean total constructed response points earned using z-scores. The selection of a reform textbook ($d = 0.57$) and the human capital component of

experience (*d* =0.26) were found to be significant, but the interaction of these two variables was not found to be significant. The effect of selecting a reform textbook on mean points earned on constructed response items was the largest effect in this study. Similar to findings in other studies (Post et al., 2008; Cai et al., 2011), the use of a reform textbook was associated with higher student achievement on open-ended assessments like constructed response items.

Taken together, the results above are important. Prior research has sometimes found students using reform textbooks are at a disadvantage on skills-based multiple-choice tests. These results show students in schools using the reform text outperform their peers on all measures, whether statewide standardized test scores or a closer examination of open-ended questions which are part of the standardized test. In other words, using a reform text does not disadvantage students on the annual testing. None of the statistically significant coefficients for textbook type were negative, indicating that the only associations between selection of a reform textbook and student achievement are positive (See Appendices F, H and I). In addition, selecting a reform textbook is associated with an increased willingness to attempt constructed response items (See Table 4). This will be important to schools in the future as the assessments being developed in conjunction with the CCSSM (Smarter Balanced (smarterbalanced.org) and Partnership for Assessment of Readiness (parconline.org)) both include extended response or constructed response items and performance tasks. The results of this study suggest students who have used reform texts will be more likely to succeed with this format of testing.

The findings above show an increase in student performance in mathematics in eighth grade for students in schools that chose reform texts but these results do not

examine growth over time. In order to understand the impact on a measure more closely associated with school performance, I used HLM to look for changes in proficiency rankings over time. Students were divided into two groups; those who had experienced the reform curriculum for at least one year between 6th and 7th grade and all others. Again, the use of a reform textbook, but not other types of textbooks, was associated with student proficiency rankings. This adds to the claim that using a reform textbook does not disadvantage students. Not only is the use of a reform text positively associated with student achievement, but it is also positively associated with proficiency rankings of students, the measure used in determining the success of the school. The only statistically significant coefficient in the model was the coefficient for selecting a reform textbook (see Appendix I). As with composite scores, SPI scores, and constructed response scores, the consistent finding was that selecting a reform textbook was positively associated with student outcomes.

The impact of using a reform textbook accounted for about 15% of the variance in the models. The effect of the human capital component of experience was positively associated with student achievement in only a limited number of cases, and the effect was small. The lack of a statistically significant effect from human capital or the interaction terms in most models led me to reconsider the model I used for this study and whether other models might produce more robust results.

The Model. The model for this study was initially conceived in response to writings by Brown (2009) and Stein and Kim (2009). Teachers use textbooks and other materials in the process of designing instruction and teaching students (Brown, 2009). In addition, the available human capital in a school may need to be considered when

selecting a textbook (Stein & Kim, 2009). Based on these theoretical suggestions, I chose to model the human capital within a school and then look at whether the interactions between that human capital and the type of textbook selected were associated with student achievement. This model did not prove fruitful in finding an effect from the interaction of human capital and textbook.

Aware of the possibility that there were characteristics of the schools that impacted which textbook they chose, I then investigated the association between some school characteristics and textbook selection. I found that student demographics for a school were not associated with the selection of a reform textbook, meaning schools with fewer poor, minority, or second language learners were no more or less likely to select a particular textbook. However, my examination of school human capital led to a different result. Schools with higher component scores on the human capital component of expertise were more likely to choose a reform textbook. This higher level of expertise was not associated with increased student achievement but was associated with a school selecting a reform textbook. In light of these findings, changes to the model may be appropriate for future research.

These findings suggest there is an association between selecting a reform textbook and student achievement. But, rather than confirm, as the initial model implied, that human capital interacts with the textbook selection in this association, the results suggest that human capital is associated with selecting a reform textbook in a school, which is then associated with student achievement.

textbook
type

student
achievement

human
capital
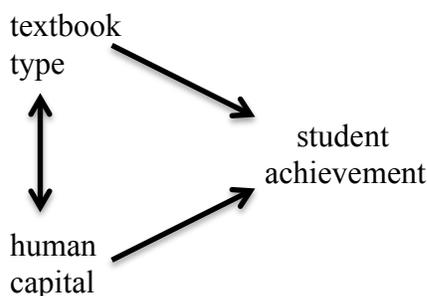
Initial Model

Therefore, one possible modification of the model to reflect the results of this study is

shown below.

human
capital
(expertise)

textbook
type

student
achievement
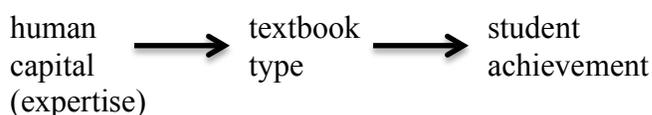
Revised Model

The investigation of the interaction of teachers and textbooks is an under-

researched area within the textbook effectiveness literature. Understanding the

mechanisms that constrain or afford effectiveness and whether or how interactions impact

these affordances and constraints is important in guiding school districts in determining

how best to provide for the education of their students. The initial model proposed in this

study should not be rejected outright. Limitations in the data available, the lack of data on

other contributors to human capital like additional non-district professional development

or coursework could all impact this model. In addition, direct measures of expertise such

as test scores or performance on items such as the Mathematical Knowledge for Teaching

Measures (Hill, Schilling, & Ball, 2004). Making the construct of human capital more robust could result in significant human capital and interaction variables.

The new model proposed in response to the results deserves further consideration. This model suggests teachers and textbooks matter, suggesting policy implications for the selection of middle school mathematics teachers and textbooks. Empowering schools with mathematics teacher expertise to select texts may result in more frequent selection of reform texts, which in turn may produce better student outcomes.

Implications for Policy and Practice

This study has a number of implications for policy and practice, particularly related to textbook selection policies and teacher licensure for middle school. The need for teacher expertise, the impact on the opportunity gap, and guidance in selecting texts that may help meet the Common Core State Standards for Mathematics are addressed.

Teacher Expertise. First, this study suggests that, when a reform textbook is chosen by teachers, its selection is associated with increased student achievement. The likelihood of a school selecting such a text is associated with the expertise of the teachers that teach math. Therefore, an initial step in increasing the likelihood that students will have access to a high quality mathematics textbook in middle school is to increase the presence of teachers with an expertise in mathematics on selection committees. Mathematics textbook selection committees considering selection of middle school mathematics texts need to attend to the preferences of committee members with expertise in mathematics, as this study showed that teachers with greater expertise selected text that were associated with higher student achievement. These findings are limited in scope and further research would be necessary in order to offer specific criteria for selecting

textbooks. However, insuring textbook selection committees include teachers with expertise in mathematics will likely lead to the selection of reform type middle school mathematics textbooks.

Later on, a more ambitious and worthwhile goal for middle schools and school districts in general would be to make sure that more teachers are hired who have expertise in mathematics.. This increased expertise will likely lead to the more frequent selection of high quality textbooks for all students. Of course, licensure policies and teacher education programs should be examined and adjusted to ensure all middle school mathematics teachers have sufficient expertise in mathematics. Requiring proven mathematical expertise for licensure and, in addition, teaching teachers to identify high quality curricula and advocate for their selection could increase the likelihood that all students have access to a high quality textbook.

The Opportunity Gap. Secondly, this study highlights a potential contributor to what many in current policy debates are now referring to as the opportunity gap, particularly for urban students (Welner & Carter, 2013). The difference in teacher quality or characteristics between urban and suburban schools is well documented (Darling-Hammond, 2013). This difference may contribute to the opportunity gap if greater expertise among middle school math teachers leads to the selection of higher quality textbooks in suburban districts over urban districts. In other words, because urban schools are less likely to have teachers with mathematical expertise, they are less likely to select reform texts. This may amplify the opportunity gap because students will not have access to high quality curricula.

Common Core State Standards and Textbooks. Third, this study suggests that using a reform textbook results in increased student achievement in middle school mathematics, and perhaps more importantly, increased achievement on open-ended items. This follows from other studies (Jones & Tarr, 2007; Sood & Jitendra, 2007) that note that reform textbooks contain materials and problems that have a higher level of cognitive demand than other texts.

The Common Core State Standards for Mathematics (CCSM) specify mathematics content and mathematical practices, but as yet there is no funded initiative to develop textbooks to meet these standards. Instead, funds have been allocated to develop assessments that developers claim will align with these standards and have more depth and greater difficulty than current state mathematics tests. Without texts that are developed with backing from professional organizations such as NCTM and NSF, identifying high quality commercial texts and selecting them by committee may be more difficult.

For this reason it is time to move beyond evaluating the alignment between texts and standards as a way to select textbooks. Textbook selection processes are currently under researched and primarily the focus only in dissertation studies (Kalder, 2004). First person accounts of participation by content experts on textbook selection committees highlight the difficulty of the committee process, the lack of guidance, and the under-representation of content specialists in evaluating the texts (e.g. Newman, 2006; Feynman, 1985).

Current policy on textbook selection typically suggests that books be examined for alignment with standards and be reviewed and selected by committees made up of

teachers, administrators and parents. Selecting texts under these policies, however, has not led to the widespread adoption of reform textbooks as parents and teachers frequently prefer texts with content that seems familiar; more traditional texts. Committee members with expertise and data on student learning, particularly data beyond multiple choice scores or proficiency ranks, on the other hand, may be more likely to select reform textbooks and may have adequate data to convince skeptical teachers and parents of the benefits.

Future Research

As more policies focused on outputs are adopted, investigating and understanding the complex system of schooling that produces the desired output (typically student achievement on a standardized test) becomes more important. This study was an initial step towards addressing that concern.

The CCSSM and the accompanying assessments currently in development do not offer direct guidance on how to get there, how to teach students so they will meet the standards. More so than prior reforms, districts, schools, and teachers are left to decide which math textbook to use and how to prepare students for annual assessments. In order to provide some guidance, research that informs the discussion by looking beyond isolated variables to more complex models is needed. One possibility is an ecological approach that attempts to account for more of the mechanisms within the complex system of schooling. A teacher's environment is part of a complex system with many pieces such as the school, teachers, students, materials, and technology, and the relationships between them. They affect one another continuously and relationships are constantly changing (Zhao & Frank, 2003). Examining the relationships among teachers, texts and student

achievement in this way will help schools understand the decisions they are making and offer guidance on how best to make these decisions.

Learning how teacher expertise is connected to textbook selection will take greater understanding of the ecology and require qualitative research that unpacks how teachers select textbooks. And, expanding beyond how teachers with expertise select books, research is needed to understand the ecology of textbook selection committees, how the committees are chosen and how different stakeholders participate. In order to understand how to ensure students have quality texts that prepare them for more challenging tests, we must first understand how this process unfolds before we try to change it.

In addition, the construct of human capital needs further study and development. Due to policy requirements, more data is available on teachers than ever before. This data may be useful in understanding a great deal about developing individual and group human capital. But, research is needed that can guide the use of this data and the development of human capital. Some questions that would increase our understanding include (1) How does mathematical content knowledge impact textbook selection by teachers? (2) How does mathematical knowledge for teaching impact textbook selection by teachers? (3) Is there an association between the construct of expertise as described in this study and mathematical content knowledge and/or mathematical knowledge for teaching? These studies will help clarify what type of knowledge is most valuable for teachers to gain in order to select high quality texts from which to teach.

Conclusion

This study raises more questions than it answers. It was a first effort to develop a quantitative model that explains the effect of mathematics textbooks and teacher human capital on student achievement. Despite the ineffectiveness of the investigated model to explain variance in student test scores, the results of the HLM and logistic regression taken together suggest that there is some relationship among these variables. In addition, theoretical considerations suggest the same (Brown, 2009; Stein & Kim, 2009). These findings, in conjunction with the statistically significant positive effect of the selection of a reform textbook suggest there is something here worthy of modification and further consideration. Greater understanding of the relationships among teachers, textbooks, and achievement will help provide guidance to schools in selecting high quality textbooks, leading to greater access and opportunity for all students.

References

Aaronson, D., Barrow, L., & Sanders, W. (2007). Teachers and student achievement in the Chicago Public Schools. *Journal of Labor Economics*, *25* (1), 95-135.

AERA (2005). Teaching teachers: Professional development to improve student achievement. *AERA Research Points. 3*(1). Retrieved from http://www.aera.net/Portals/38/docs/Publications/Teaching%20Teachers.pdf

Alexander, K. L, Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 171-191.

Aloe, M. A., McLeish, C., Shisler, S. M., & Cantaffa, D. T. (2013). Teacher degree level and student academic achievement. Paper presentation, AERA, San Francisco, CA April 2013

American Association for the Advancement of Science: Project 2061. (2000). *Middle grades mathematics textbooks: A benchmark based evaluation*. Washington, DC.

Angus, D. L. (2001). *Professionalism and the public good: A brief history of teacher certification*. Washington, DC: Fordham Foundation.

Apple, M., & Christian-Smith, L. (1991). *The politics of the textbook.* New York, NY: Routledge.

ARC Center. (2003). *The ARC Tri-State Sudent Achievement Study.* Lexington, MA: ARC.

Argote, L. (1999). *Organizational learning: Creating, retaining and transferring knowledge.* Boston, MA: Kluwer Academic Publishers.

Ashby, C. M. (2010). *No child left behind act: Enhancements in the department of education's review process could improve state academic assessments.* Washington, DC: US Government Accountability Office.

Badgett, B. A. (20110. Toward the Development of a Model to Estimate the Readability of Credentialing-Examination Materials. Unpublished doctoral dissertation.

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, *66*(2), 103-115.

Ballou, D., & Podursky, M. (2000). Reforming teachers' preparation and licensing: What is the evidence? *Teachers College Record*, *102* (1), 2-7.

Barrick, M., Steward, G., Neubert, M., & Mount, M. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology, 55.* 267-306.

Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society, 16* (series B), 296-298.

Becker, G. S. (1964). *Human captial: A theoretical and empirical analysis, with special reference to education.* Chicago, IL: University of Chicago Press.

Ben-Chaim, D., Fey, J. T., Fitzgerald, W., Benedetto, C., & Miller, J. (1998). Proportional reasoning among seventh grade students with different curricula experiences. *Educational Studies in Mathematics*, *36* (3), 247-273.

Betts, J. R., Reuben, K. S., & Danneberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California.* San Francisco, CA: Public Policy Institute of California.

Betts, J., Zau, A., & Rice, L. (2003). *Determination of student achievement: New evidence from San Diego.* Public Policy Institute of California. San Francisco, CA: Public Policy Institute of California.

Borko, H., Davinroy, K. H., Bliem, C. L., & Cumbo, K. B. (2000). Exploring and supporting teacher change: Two third-grade teachers' experiences in a mathematics and literacy staff development project. *Elementary School Journal*, *100*, 273-306.

Boyd, D. G. (2007). The effect of certification and preparation on teacher quality. *The Future of Children*, *17* (1), 45-68.

Braswell, J. S., Lutkus, A. D., Grigg, W. W., Santapau, S. L., Tay-Lim, B., & Johnson, M. (2001). *The nations's report card: Mathematics 2000.* Washington, DC: Department of Education, Office of Educational Research and Improvement.

Britton, B. K., Woodward, A., & Brinkley, M. (1993). *Learning from textbooks: Theory and practice.* Hillsdale, NJ: Erlbaum Associates.

Brown, M. W. (2009). The teacher-tool relationship. In J. T. Remillard, B. A. Herbel-Eisenmann, & G. M. Lloyd (Eds.), *Mathematics Teachers at Work: Connecting Curriculum Materials and Classroom Instuction* (pp. 17-36). New York, NY: Routledge.

Bryant, B. R., Bryant, D. P., Kethley, C., Kim, S. A., Pool, C., & Seo, Y. (2008). Preventing mathematics difficulties in the primary grades: The critical features of instruction in textbooks as part of the equation. *Learning Disability Quarterly*, *31*, 21-35.

Cai, J., Wang, N., Moyer, J. C., Wang, C., & Nie, B. (2011) Longitudinal investigation of the curricular effect: An analysis of student learning outcomes from the LieCal project in the United States. *International Journal of Educational Research. 50.* 1117-136

Callow-Heusser, C., Allred, D., Robertson, D., & Sanborn, W. (2005). *Evidence -based small-sclae study final report: McDougal Littell Middle School Math.* Logan, UT: EndVision Research & Evaluation.

Cavalluzzo, L. (2004). *Is national board certification an effective signal of teacher quality.* Alexandria, VA: The CNA Corporation.

Chronbach, L. (1955). *Text Materials in modern education: A comprehesive theory and platform for research.* Urbana, IL: Univeristy of Illinois Press.

Clark, R. P. (2010). *Developing our human capital: a mixed method study of teacher use of online communities.* (Unpublished Doctoral Dissertation)

Clarkson, L. M. (2001). *The effects of the Connected Mathematics Project on middle school mathematics achievement.* (Unpublished Doctoral Dissertation)

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778-820.

Clotfelter, C., Ladd, H., & Vigdor, J. (2007). *How and why do teacher credentials matter for student learning.* National Bureau of Economic Research. Cambridge, MA: National Bureau of Economic Research.

Clotfelter, C., Ladd, H., & Vigdor, J., (2010). Teacher credentials and student achievement in high school: A cross subject analysis with student fixed effects. *Journal of Human Resources, 45*(3), 655-671.

College Entrance Examination Board (1998). SAT 1998 College Bound Seniors: National Report. New York: Author.

Coff, R. W. (1999). When competitive advantage doesn't lead to performance: the resource based view and stakeholder bargaining power. *Organizational Science*, *10* (2), 119-133.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003) *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd edition). New York, NY: Routledge.

Cohen, D., & Hill, H. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, *102* (2), 294-343.

Cohen, D., & Hill, H. (2001). *Learning policy: When state education reform works.* New Haven, CT: Yale University Press.

Coleman, J. S. (1966). *Equality of Educational Opportunity,* Washington, DC: US Government Printing Office.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, 95-121.

Coley, R. J. (2002). An Uneven Start: Indicators of Inequality in School Readiness. Policy Information Report.

Darling-Hammond, L. (1998). Top 10 list for teacher quality. *NEA Today*, *17* (3), 17-24.

Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives*, *8* (1), 1-9.

Darling-Hammond, L. (2000). Reforming teacher preparation and licensing: Debating the evidence. *Teachers College Record*, *102* (1), 28-56.

Darling-Hammond, L. (2001). Standard setting in teaching: Changes in lincensing, certification and assessment. In V. Richardson (Ed.), *Handbook on research on teaching* (pp. 751-776). Washington, DC: American Educational Research Association.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106* (6), 1047-1085.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13* (42).

Darling-Hammond, L (2013). Inequality and school resources. In P. L. Carter & K. G. Welner (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 77-97). New York, NY: Oxford University Press.

Day, E. A., Arthur, W., Bell, S. T., Edwards, B. D., Bennett, W., Mendoza, J. L., & Tubre, T. C. (2005). Ability-based pairing strategies in the team-based training of a complex skill: Does the intelligence of your training partner matter? *Intelligence*, *33*, 39-45.

DeShon, R., Kozlowski, S., Schmidt, A., Milner, K., & Weichmann, D. (2004). A multiple-goal, multi-level model of feedback effects on the regulation of individual and team performance. *Journal of applied psychology, 89*, 1035-1056.

Eddy, R. M., Berry, T., Aguirre, N., Wahlstrand, G., Ruitman, T., & Mahajan, N. (2008). *The Effects of Connected Mathematics Project 2 on Student Performance: Randomized Control Trial Final Report*. Claremont Graduate University, Claremont, CA. Funding provided by Pearson Education.

Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development.* New York, NY: Albert Shanker Institute.

Faraj, S., & Sproull, L. (2000). Coordinating expertise in software development teams. *Management Science*, *26* (12), 1554-1568.

Farmer, J., Gerretson, H., & Lassak, M. (2003). What teachers take from professional development: Cases and implications. *Journal of Mathematics Teacher Education*, *6* (4), 331-360.

Feynman, R. P. (1985). *" Surely you're joking, Mr. Feynman!": Adventures of a curious character*. New York, NY: WW Norton & Company.

Gamoran, A. & Long, D. A. (2006). *Equality of educational opportunity: A 40-year retrospective* (WCER Working Paper No. 2006-9). Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Educational Research.

Gamoran, A. & An, B. (2005) *Effects of school segregation and school resources in a changing policy context.* Paper presented at meeting of the research committee on social stratification of the international sociological association, Los Angeles, CA

Gimbert, B., Bol, L., & Wallace, D. (2007). The influence of teacher preparation on student achievement and the application of national standards by teachers of mathematics in urban secondary schools. *Education and Urban Society*, *40*(1), 91-117.

Goe, L. (2002). Legislating equity: the distribution of emergency permit teachers in California. *Education Policy Analysis Archives*, *10* (42).

Goldhaber, D. D., & Brewer, D. J. (1996). Evaluation of the effect of the teacher degree level on educational performance. In W. Fowler (Ed.), *Developments in School Finance* (pp. 197-210). Washington, DC: NCES.

Goldhaber, D., & Brewer, D. (1997). Why don't schools and teachers seem to matter more: Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, *3*, 505-523.

Goldhaber, D. D., & Brewer, D. J. (1999). Teacher licensing and student achievement. In M. Kanstoroom, & C. E. Finn (Eds.), *Better Teachers, Better Schools* (pp. 83-103). Washington, DC: The Thomas B. Fordham Foundation.

Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*, 129-145.

Goldstein, P. (1978). *Changing the American schoolbook.* Lexington, MA: DC Heath & Co.

Greenberg, E., Rhodes, D., Ye, X., & Stancavage, F. (2004). Prepared to teach: Teacher preparation and student achievement in eighth grade mathematics. Paper presentation, AERA, San Deigo, CA, 2004.

Greenwald, R., Hedges, L., & Laine, R. (1996). The effect of school resources on student achievement. *Journal of Human Resources*, *30* (2), 280-310.

Grissmer, D., Flanagan, A. & Williamson, S. (1998). Why did the black-white score gap narrow in the 1970's and 1980's. n Jencks, C. & Phillips, M. (Eds.) *The black-white test score gap.* Washington, DC: Brookings Institution Press.

Grouws, D. A., & Smith, M. S. (2000). Findings from NAEP on the preparation and practices of mathematics teachers. In E. A. Silver, & P. A. Kenney (Eds.), *Results from the Seventh Mathematics Assessment of the National Assessment of Educational Progress.* Reston, VA: National Council of Teachers of Mathematics.

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

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economics Literature*, *24* (3), 1141-1177.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (No. w11154). National Bureau of Economic Research.

Hardy, M. A. (1993). *Regression with dummy variables.* Newbury Park, CA: Sage.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, *95*(7), 798-812.

Harwell, M. R., Post, T. R., Maeda, Y., Davis, J. D., Cutler, A. L., Andersen, E., & Kahan, J. A. (2007). Standards-based mathematics curricula and secondary students' performance on standardized achievement tests. *Journal for Research in Mathematics Education*, 71-101.

Hawk, P. P., Coble, C. R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education*, *36* (3), 13-15.

Hawkins, E., Stancavage, F. B., & Dossey, J. A. (1998). *School policies and practices affecting instruction in mathematics.* Washington, DC: National Center for Education Statistics.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *Americal Educational Research Journal*, *42*, 371-406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal,* 105(1), 11.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Ingersoll, R. (2001). *Out of field teaching and the limits of teacher policy.* Center for the Study of Teaching and Policy. Seattle, WA: The Consortium for Policy Research in Education.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evalutation in education. *Journal of Labor Economics*, *26*, 101-136.

Jenks, C., & Phillips, M. (1998) *The black-white test score gap.* Washington, DC: Brookings Institution Press.

Jones, D. L., & Tarr, J. E. (2007). An examination of the levels of cognitive demand required by probability tasks in middle grades mathematics textbooks. *Statistics Eduation Research Journal*, *6* (2), 4-27.

Jones, M. G., Jones, B. D., & Hargroves, T. Y. (2003). *The unintended consequences of high-stakes testing.* Lanham, MD: Rowman & Littlefield.

Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.

Kalder, R. S. (2004). *A Study of the Mathematics Textbook Adoption Process in Selected New York State School Districts*. (Unpublished Doctoral Dissertation)

Kane, T., Rockoff, J. E., & Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness: Evidence from New York City.* Cambridge, MA: National Bureau of Economic Research.

Klecker, B. M. (2008). Is Teacher Quality Related to Eighth-Grade Mathematics Achievement? Evidence from the 2007 NAEP Data. ERIC *Online Submission*.

Klein, D. (2003). A brief history of American K-12 mathematics education in the 20th century. In J. M. Royer (Ed.), *Mathematical Cognition.* Charlotte, NC: Information Age Publishing.

Kornhaber, M. L., & Orfield, G. (2001). High-stakes testing policies: Examining their assumptions and consequences. In M. Kornhaber & G. Orfield (Eds.), *Raising Standards or raising barriers: Inequality and high stakes testing in public education* (pp. 1-18). New York: Century Foundation.

Laczko-Kerr, I., & Berliner, D. C. (2002). The effectiveness of Teach for America and other under-certified teachers on student academic achievement: A case of hamful public policy. *Education Policy Analysis Archives*, *10* (37), 1-52.

Lafferty, J. F. (1994). The links among mathematics text, student achievement, and students' mathematics anxiety: A comparon of the incremental development and traditional texts. (Unpublisher Doctoral Dissertation)

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37-62.

Loeb, S., & Reininger, M. (2004). *Public policy and teacher labor markets: What we know and why it matters.* East Lansing, MI: Michigan State University: The Education Policy Center.

Martin, T., Brasiel, S. J., Turner, H., & Wise, J. C. (2012). Effects of the Connected Mathematics Project 2 (CMP2) on the Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region. Final Report. NCEE 2012-4017. *National Center for Education Evaluation and Regional Assistance*.

Mathematically Correct. (1999). *Mathematically Correct Seventh Grade Mathematics Program Review*. Retrieved August 1, 2011, from Mathematically Correct: http://mathematicallycorrect.com/books7.htm

Mayer, R. E., Sims, V., & Tajika, H. (1995). A comparison of how textbooks teach mathematical problem solving in Japan and the United States. *American Educational Research Journal*, *32* (2), pp. 443-460.

McCaffrey, M. D. (1971). Politics in the schools: A case for partisan board elections. *Educational Administration Quarterly*, *7* (3), 51-63.

Mueller, C. M. (2012). The Impact of Teacher Certification Programs on Teacher Efficacy, Job Satisfaction, and Teacher Performance: A Comparison of Traditional and Alternative Certification.

Murnane, R. J., & Steele, J. L. (2007). What is the problem? The challenge of providing effective teachers for all children. *The Future of Children*, *17*(1), 15-43.

NACOME. (1975). *Overview and analysis of school mathematics grades K-12.* Washington, DC: Conference Board of the Mathematical Sciences.

National Commission on Excellence in Education. (1983). *A Nation at Risk: The imperative for educational reform.* Washington, DC: United States Department of Education

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's Future.* New York: National Commission of Teaching and America's Future.

National Council of Teachers of Mathematics. (1980). *An Agenda for Action.* Reston, VA: NCTM.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: NCTM.

National Mathemaics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* National Math Advisory Panel. Washington, DC: Department of Education.

National Middle School Association. (1996). *Middle level licensure 1996.* Columbus, OH: NMSA.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Newman, W. J. (2004). Serving on a mathematics textbook selection committee: A tale of woe. *School Scienc and Math. 104(8)*. 361-367

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

Orlich, D. (1978). *They still use textbooks, don't they: Selection and criteria for evaluation.* Olympia, WA: School Information and Research Service.

Orfield, G. (2001) *Schools More Separate: Consequences of a Decade of Resegregation.* Cambridge, MA: Civil Rights Project, Harvard University.

Orfield, G., Kucsera, J., & Siegel-Hawley, G. (2012). *E Pluribus…Separation: Deepening double segregation for more students.* Cambridge, MA: Civil Rights Project, Harvard University.

Paige, R. (2002). *Meeting the highly qualified teachers challenge.* Washington, DC: United States Department of Education.

Patil, V. H., Singh, S. N., Mishra, S., & Donavan, D. T. (2007). *Parallel analysis engine to aid determining number of factors to retain* (computer software). Available from http://smishra.faculty.ku.edu/parallelenging.htm

Perie, M., Moran, R., & Lutkus, A. D. (2005). *NAEP 2004 trends in academic progress: Three decades of student performance in reading and mathematics (*NCES 2005-464). Washington, DC: Government Printing Office.

Pil, F. K., & Leana, C. (2009). Applying organizational research to public school reform: The effects of teacher human and social capital on student performance. *The Academy of Management Journal Archive*, *52* (6), 1101-1124.

Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards?. *American Educational Research Journal*, *48*(4), 965-995.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3-14.

Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Remillard, J. T., Herbel-Eisenmann, B. A., & Lloyd, G. M. (Eds.). (2009). Mathematics teachers at work: Connecting curriculum materials and classroom instruction. New York, NY: Routledge

Rentschler, R. V. (1994). The effects of Saxon's incremental review on computational skills and problem solving achievement of sixth-grade students. (Unpublished Doctoral Dissertation)

Resendez, M., & Azin, M. (2005). *A study on the effectiveness of the 2004 Prentice Hall Course 2 Middle School Math Program.* Jackson, WY: PRES Associates.

Resendez, M., Fahmy, A., & Azin, M. (2005). *The relationship between using Saxon Middle School Math and student performance on Texas statewide assessments.* Jackson, WY: PRES Associatees.

Reys, B. J., Reys, R. E., Barnes, D., Beem, J., & Papick, I. (1997). Collaborative curriculum investigation as a vehicle for teacher enhancement and mathematics curriculum reform. *School Science and Mathematics*, *97* (5), 253-259.

Reys, R., Reys, B., & Lapan, R. H. (2003). Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, *34* (1), 74-95.

Resnick, L. B., Rothman, R., Slattery, J. B., & Vranek, J. L. (2004). Benchmarking and alignment of standards and testing. *Educational Assessment 9*(1).

Reys, B., Dingman, S. W., Sutter, A., & Teuscher, D. (2005). Development of state-level mathematics curriculum documents: Report of a survey. *Center for the Study of Mathematics Curriculum publications (MU).*

Reys, B. J., & Reys, R. E. (2006). The development and publication of elemntary mathematics textbooks: Let the buyer beware! *Phi Delta Kappan*, *87*, 337-383.

Ridgway, J., Zawojewski, Z., Hoover, M., & Lambdin, D. (2003) 'Student attainment in the connected mathematics curriculum.', in *Standards-based school mathematics curricula : what are they ? what do students learn ?* New Jersey: Lawrence Erlbaum, pp. 193-224.

Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal of Research in Mathematics Education*, *32*, 368-398.

Riordan, J. E., Noyce, P. E., & Perda, D. (2003). The impact of two standards based mathematics curricula on student achievement in Masachusetts: A follow up study of Connected Mathematics. *Journal of Research in Mathematics Education*, *32* (4), 368-398.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73* (2), 417-458.

Roberts, F. H. (1994). The impact of the Saxon mathematics program on group achievement test scores. (Unpublished Doctoral Dissertation)

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94* (2), 247-252.

Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (2002). *Benchmarking and alignment of standards and testing*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Schneider, C. (2000) *Connected Mathematics and the Texas assessment of academic skills* (Unpublished doctoral dissertation).

Schoenfeld, A. H. (1994). What do we know about mathematics curricula? *Journal of Mathematical Behavior*, *13* (1), 55-80.

Schoenfeld, A. (2004). The Math Wars. *Education Policy*, *18* (1), 253-286.

Senk, S., & Thompson, D. (2003). *Standards-oriented school mathematics curricula: What does the research say about student outcomes?* Mahwah, NJ: Erlbaum.

Sher, L., & O'Reilly, F. (2009). Professional development for K-12 math and science teachers: What do we really know. *Journal of Research on Educational Effectiveness*, *2* (4), 209-249.

Silver, E. A. (1996). Moving beyond learning alone and in silence: Observations from the QUASAR project concerning communication in mathematics classrooms. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments in education* (pp. 127-160). Mahwah, NJ: Erlbaum.

Simon, M., & Schifter, A. D. (1993). Toward a constructivist perspective: The impact of a mathematics teacher inserivce program on students. *Educational Studies in Mathematics*, *25*, 331-340.

Skandera, H., & Sousa, R. (2007). What makes a good teacher. *Hoover Digest*, *2*.

Smith, K. G., Collins, C. J., & Clark, K. D. (2005). Existing knowledge, knowledge creation capability and the rate of new product introduction in high technology firms. *Academy of Management Journal*, *48*, 346-357.

Smith, M. S., & Stein, M. K. (1998). Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School*, *3*, 344-350.

Smithson, J. L., & Blank, R. K. (2006). *Indicators of quality of teacher profesional development and instructional change using data from surveys of enacted curriculum: findings from NSF MSP-RETA Project.* Washington, DC: Council of Chief State School Officers.

Sood, S., & Jitendra, A. K. (2007). A comparative analysis of number sense instruction in reform-based and traditional mathematics textbooks. *Journal of Special Education*, *41* (3), 145-157.

Stein, M. K., & Kim, G. (2009). The role of mathematics curriculum materials in large-scale urban reform: An analysis of demands and opportunities for teacher learning. In J. Remillard, G. Lloyd, & B. Herbel-Eisenmann (Eds.), *Teachers' use of mathematics curriculum materials: Research perspectives on relationships between teachers and curriculum* (pp. 37-55). New York, NY: Routledge.

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection. *Mathematics Teaching in the Middle School*, *3*, 268-275.

Stein, M. K., Smith, M. W., Henningsen, M., & Silver, E. A. (2000). *Implementing standards based mathematics instruction: A casebook for professional development.* New York, NY: Teachers College Press.

Swann, J. M. (1996). An investiation into the effectiveness of Transition Mathematics. (unpublished doctoral dissertation)

Tabachnick, B. G., & Fidell, L. S. (2007). Using Multivariate Statistics. Boston, MA: Pearson Education Inc.

Tarr, J. E., Reys, B. J., Billstein, R., & Baker, D. B. (2006). Selecting high quality mathematics textbooks. *Mathematics Teaching in the Middle School*, *12* (1), 50-54.

Thompson, D. R., Senk, S. L., Witonsky, D., Usiskin, Z., & Kaeley, G. (2005). *An evaluation of the second edition of UCSMP Transition Mathematics.* Chicago, IL: University of Chicago School MathematicsProject.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational research*, *73*(1), 89-122.

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (NISE Research Monograph No. 18) Madison, WI: University of

Wisconsin-Madison, National Institute for Science Education, Washington, DC: Council of Chief State School Officers.

Webb, N. L. (2002) An analysis of the alignment between mathematics standards and assessments for three states. In *Annual meeting of the American Educational Research Association, New Orleans, LA.*

Weiss, I. R. (2007). *Participant involvement in planning professional development in technology integration*. New York, NY.

Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001). *Report of the 2000 national survey of science and mathematics education.* Chapel Hill, NC: Horizon Research, Inc.

Welner, K. G. & Carter, P. L. (2013). Achievement gaps arise from opportunity gaps. In P. L. Carter & K. G. Welner (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 77-97). New York, NY: Oxford University Press.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analyisis Archives*, *10* (12).

Wenglinsky, H. (2002). *How teaching matters: Bringing the classroom back into discussions of teacher quality.* Princeton, NJ: ETS.

Wenglinsky, H. (2006). On ideology, causal inference and the reification of statistical methods: reflections on examining instruction, achievement and equity with NAEP mathematics data. *Educational Policy Analysis Archives*, *14* (17), 1-8.

Woodbury, M. (1979). *Selecting Materials for Instruction: Issues and Policies.* Littleton, CO: Libraries Unilmited, Inc.

Woodward, J. (2004). Mathematics education in the United States: Past to present. *Journal of Learning Disabilities*, *37* (1), 16-31.

Youngs, P., Frank, K., Thum, Y. M., & Low, M. (2012). The motivation of teachers to produce human capital and conform to their social contexts. *Yearbook of the National Society for the Study of Education*, *110*, 248-272.

Zhao, Y., & Frank, K. A. (2003). Factors affecting technology uses in schools: An ecological perspective. *American educational research journal*, *40*(4), 807-840.

Zvoch, K., & Stevens, J. J. (2006). Longitudinal effects of school context and practice on middle school mathematics achievement. *The Journal of Educational Research*, *99* (6), 347-356.

Appendix A

Release Constructed Response Items

Released items, constructed response, grade 8 WKCE, 2005. These are the latest released items available at this time. Retrieve from http://oea.dpi.wi.gov/oea_mathptri

Release question 1:

Kate makes a batch of salsa in a large cylindrical pot. The inside of the pot is 9 inches in diameter, and it is filled with 4 inches of salsa. Kate plans to store the salsa in small cylindrical glass jars that are 3 inches in diameter and 4 inches high.

$$V = \pi r^2 h$$

Step A

How many glass jars will Kate need for all of the salsa? (Use 3.14 to approximate π)

Answer: _____ glass jars

Step B

Explain how you determined the number of glass jars Kate will need. Use words and/or numbers in your explanation.

Release question 2:

Scott has an old fish tank in the shape of a box. It fits exactly onto a rectangular stand that is 12 inches wide and 30 inches long. The tank can be filled with water to a depth of 15 inches.

Step A

What is the total volume of water that Scott's old fish tank can hold?

Answer: _____ cubic inches

Step B

Scott is buying a new fish tank that fits on the same stand as the old tank, but holds up to 7,200 cubic inches of water. Use what you know about volume to explain how to find the depth of the water in the new fish tank. Use words and/or numbers in your explanation.

Appendix B
Principal Component Analysis

Correlation Matrix

	years mps	years total	highest degree	nonmath pd	math pd	grades certification	math endorsement
yearsmps	1.000	.979	.536	-.229	.005	-.061	-.061
yearstotal	.979	1.000	.513	-.191	.004	-.068	-.068
highest degree	.536	.513	1.000	-.159	.035	-.046	-.046
nonmathpd	-.229	-.191	-.159	1.000	.149	-.071	-.071
mathpd	.005	.004	.035	.149	1.000	.260	.260
grades certification	.143	.123	.116	-.042	.218	.237	.237
math endorsement	-.061	-.068	-.046	-.071	.260	1.000	1.000

Scree Plot
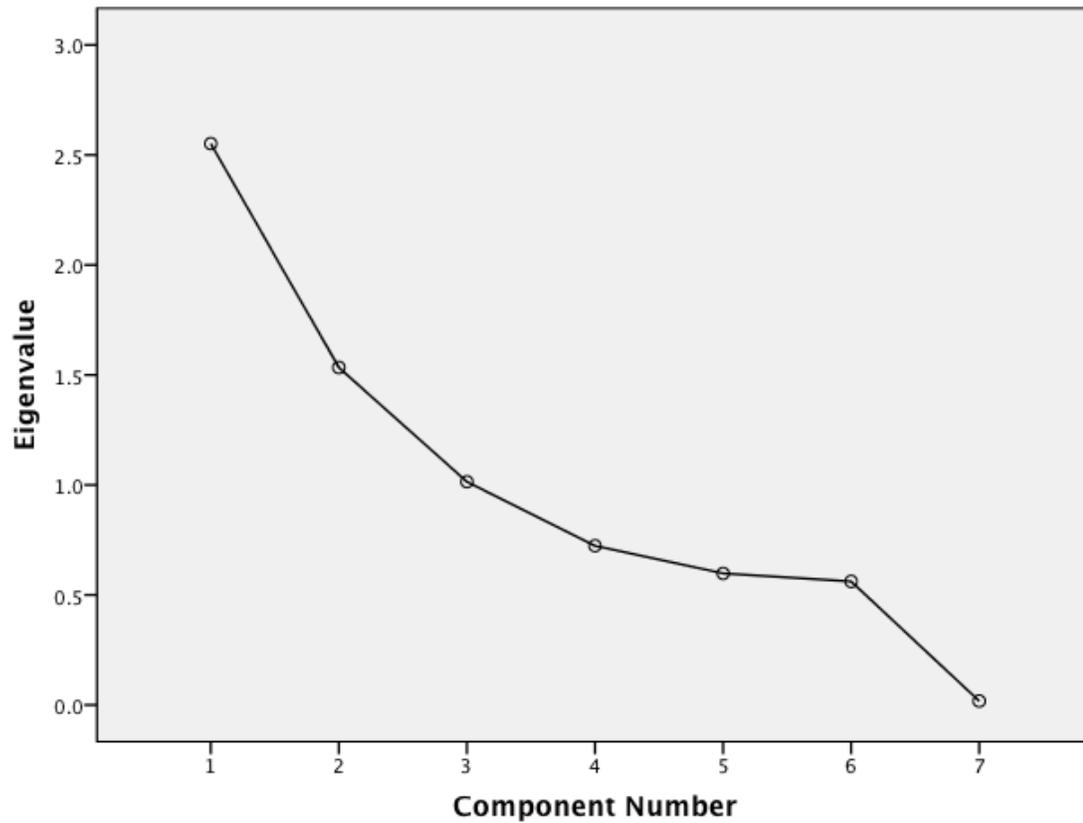
Appendix C
Parallel Analysis and Components

Parallel Analysis Results

Eigenvalue	Observed (PCA)	Mean from PA	Retain/reject
1	2.491	1.203239	Retain
2	1.476	1.114205	Retain
3	1.026	1.051638	Reject
4	0.752	0.996929	Reject
5	0.641	0.944083	Reject
6	0.594	0.881794	Reject
7	0.020	0.808113	Reject

Final Component Matrix
Six variables, forced 2 component solution

Component Matrix[a]

	Component	
	experience	expertise
yearsmps	.959	
yearstotal	.950	
highest degree	.729	
mathpd		.704
grades certification		.652
math endorsement		.735

Extraction Method: Principal Component
Analysis.
a. 2 components extracted.

Appendix D

Intra-class correlation calculations

$$ICC = \frac{\tau_{00}}{\sigma^2 + \tau_{00}}$$

Intra-class correlation computations

Dependent Variable	τ_{00}	σ^2	ICC
Scale score	672.21	2706.81	0.199
Math A	85.83	296.62	0.224
Math B	73.74	291.20	0.202
Math C	70.28	274.77	0.204
Math D	51.34	182.45	0.220
Math E	59.25	237.30	0.200
Math F	68.66	281.32	0.196

Appendix E

Descriptive Statistics

Descriptive Statistics for Variables in HLM

Variables	N	Mean	SD	Min	Max
SES	4136	0.81	0.39	0	1
Traditional	66	0.17	0.38	0	1
Reform	66	0.20	0.40	0	1
Experience	66	-0.10	0.57	-1.70	1.83
Expertise	66	-0.01	0.48	-1.09	1.24

Descriptive Statistics for Outcome Measures in HLM

Outcome measures	N	Mean	SD	Min	Max
Scale Score	3826	503.78	57.38	350	694
SPI Math A mathematical processes	3826	31.35	19.14	0	99
SPI Math B Number & operations	3826	38.69	18.71	0	99
SPI Math C Geometry	3826	55.35	18.33	0	99
SPI Math D Measurement	3826	33.32	14.90	0	99
SPI Math E Probability & Statistics	3826	45.43	16.97	0	99
SPI Math F Algebra	3826	49.10	18.40	0	99
Constructed Response (z-scores)	3826	0	1	0.94	3.80
Proficiency	9044	2.40	0.99	1.00	4.00

Appendix F
HLM Models

Appendix E contains final coefficients and standard errors for the HLM models using outcome variables of scale score each strand score. Also included is the development of the first model (all other models developed in a similar fashion), and variance measures.

Model with the scale score as the dependent varialbe

Null model

Level-1 Model

$SCALESCO_{ij} = \beta_{0j} + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + u_{0j}$

Mixed Model

$SCALESCO_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

$\sigma^2 = 2706.80770$

τ
INTRCPT1,β_0 672.21132

Final estimation of fixed effects, Scale score is dependent variable (with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. d.f.	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	502.214455	3.346716	150.062	65	<0.001

Model including SES only

Level-1 Model

$SCALESCO_{ij} = \beta_{0j} + \beta_{1j}*(SES7_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10}$

Mixed Model

$SCALESCO_{ij} = \gamma_{00}$
$+ \gamma_{10}*SES7_{ij} + u_{0j} + r_{ij}$

$\sigma^2 = 2670.27333$

τ
INTRCPT1,β_0 574.21517

Final estimation of fixed effects, scale score is dependent variable
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	516.665604	4.306004	119.987	65	<0.001
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-17.788796	2.807243	-6.337	3759	<0.001

2-level model including textbook type selected

Level-1 Model

$SCALESCO_{ij} = \beta_{0j} + \beta_{1j}*(SES7_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TRADITIO_j) + \gamma_{02}*(REFORM_j) + u_{0j}$
$\beta_{1j} = \gamma_{10}$

Mixed Model

$SCALESCO_{ij} = \gamma_{00} + \gamma_{01}*TRADITIO_j + \gamma_{02}*REFORM_j$
$+ \gamma_{10}*SES7_{ij}$
$+ u_{0j} + r_{ij}$

$\sigma^2 = 2669.94789$
τ
INTRCPT1,β_0 502.44599

Final estimation of fixed effects, Scale Score is DV
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	509.815482	4.425158	115.208	63	<0.001
TRADITIO, γ_{01}	13.040121	7.590068	1.718	63	0.091
REFORM, γ_{02}	23.756212	8.562192	2.775	63	0.007
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-17.813972	2.791204	-6.382	3759	<0.001

Full 2-level model

Level-1 Model

$$SCALESCO_{ij} = \beta_{0j} + \beta_{1j}*(SES7_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TRADITIO_j) + \gamma_{02}*(REFORM_j) + \gamma_{03}*(EXPERIEN_j) + \gamma_{04}*(EXPERTIS_j)$$
$$+ \gamma_{05}*(TRAD*EXPERIENCE) + \gamma_{06}*(TRAD*EXPERTISE) +$$
$$\gamma_{07}*(REFORM*EXPERIENCE) + \gamma_{08}*(REFORM*EXPERTISE) + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

Mixed Model

$$SCALESCO_{ij} = \gamma_{00} + \gamma_{01}*TRADITIO_j + \gamma_{02}*REFORM_j + \gamma_{03}*EXPERIEN_j$$
$$+ \gamma_{04}*EXPERTIS_j + \gamma_{05}*TRAD*EXPERTISE_j + \gamma_{06}*TRAD*EXPERIENCE_j +$$
$$\gamma_{07}*REFORM*EXPERIENCE_j + \gamma_{08}*REFORM*EXPERTISE + \gamma_{10}*SES7_{ij} + u_{0j} + r_{ij}$$

$\sigma^2 = 2670.31680$

τ

INTRCPT1,β_0 477.19127

Final estimation of fixed effects, Scale Score is DV
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	511.647024	4.436859	115.317	57	<0.001
TRADITIO, γ_{01}	10.745164	7.159804	1.501	57	0.139
REFORM, γ_{02}	27.145027	9.274514	2.927	57	0.005
EXPERIEN, γ_{03}	13.486351	8.190464	1.647	57	0.105
EXPERTIS, γ_{04}	5.345179	7.401772	0.722	57	0.473
TRAD*EXPERTISE, γ_{05}	-11.456496	17.725193	-0.646	57	0.521
TRAD*EXPERIENCE, γ_{06}	-2.670134	12.602976	-0.212	57	0.833
REFORM*EXPERIENCE, γ_{07}	-11.727136	16.695090	-0.702	57	0.485
REFORM*EXPERTISE, γ_{08}	-14.820325	22.560113	-0.657	57	0.514
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-17.800077	2.798666	-6.360	3759	<0.001

The remaining tables show the final estimates of fixed effects for similar models.

DV: Math A (mathematical processes)
$\sigma^2 = 290.74444$
τ
INTRCPT1,β_0 59.07593

Final estimation of fixed effects, Math A (matheamtical processes) is DV
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. d.f.	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	34.386363	1.515202	22.694	57	<0.001
TRADITIO, γ_{01}	4.260772	2.439338	1.747	57	0.086
REFORM, γ_{02}	10.094554	3.576638	2.822	57	0.007
EXPERIEN, γ_{03}	4.914494	2.509093	1.959	57	0.055
EXPERTIS, γ_{04}	1.921843	2.283503	0.842	57	0.404
TRAD*EXPERTISE, γ_{05}	-3.393070	5.852166	-0.580	57	0.564
TRAD*EXPERIENCE, γ_{06}	-0.868051	4.120464	-0.211	57	0.834
REFORM*EXPERIENCE, γ_{07}	-2.670854	5.748796	-0.465	57	0.644
REFORM*EXPERTISE, γ_{08}	-4.007883	7.273549	-0.551	57	0.584
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-7.040932	1.053365	-6.684	3759	<0.001

DV: Math B (Number and Operations)
$\sigma^2 = 286.41390$
τ
INTRCPT1,β_0 53.97445

Final estimation of fixed effects, Math B (number and operations) is DV
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. d.f.	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	41.870447	1.482695	28.239	57	<0.001
TRADITIO, γ_{01}	3.784680	2.271681	1.666	57	0.101
REFORM, γ_{02}	8.532098	3.321853	2.568	57	0.013
EXPERIEN, γ_{03}	5.023939	2.378407	2.112	57	0.039
EXPERTIS, γ_{04}	1.486685	1.991972	0.746	57	0.459
TRAD*EXPERTISE, γ_{05}	-4.378666	5.320890	-0.823	57	0.414
TRAD*EXPERIENCE, γ_{06}	-1.826436	4.101664	-0.445	57	0.658
REFORM*EXPERIENCE, γ_{07}	-2.500609	5.116643	-0.489	57	0.627
REFORM*EXPERTISE, γ_{08}	-4.944070	6.735062	-0.734	57	0.466
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-6.436687	1.171526	-5.494	3759	<0.001

DV: Math C (Geometry)

$\sigma^2 = 270.50055$

τ

INTRCPT1,β_0 50.93335

Final estimation of fixed effects, Math C (Geometry) is DV
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. d.f.	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	58.003648	1.409646	41.148	57	<0.001
TRADITIO, γ_{01}	3.721416	2.304090	1.615	57	0.112
REFORM, γ_{02}	8.387701	2.867267	2.925	57	0.005
EXPERIEN, γ_{03}	4.287045	2.663722	1.609	57	0.113
EXPERTIS, γ_{04}	2.316958	2.468189	0.939	57	0.352
TRAD*EXPERTISE, γ_{05}	-3.998577	5.458349	-0.733	57	0.467
TRAD*EXPERIENCE, γ_{06}	-0.305725	3.784862	-0.081	57	0.936
REFORM*EXPERIENCE, γ_{07}	-4.683331	5.229351	-0.896	57	0.374
REFORM*EXPERTISE, γ_{08}	-3.924462	6.468392	-0.607	57	0.546
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-6.053884	0.899544	-6.730	3759	<0.001

DV: Math D (Measurement)

$\sigma^2 = 178.78349$

τ

INTRCPT1,β_0 37.41947

Final estimation of fixed effects, Math D (Measurement) as DV
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. d.f.	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	38.192324	1.212847	31.490	57	<0.001
TRADITIO, γ_{01}	2.857685	1.777157	1.608	57	0.113
REFORM, γ_{02}	6.950688	2.860517	2.430	57	0.018
EXPERIEN, γ_{03}	3.104327	2.005671	1.548	57	0.127
EXPERTIS, γ_{04}	1.447421	1.624597	0.891	57	0.377
TRAD*EXPERTISE, γ_{05}	-2.278652	4.401789	-0.518	57	0.607
TRAD*EXPERIENCE, γ_{06}	-0.873034	3.539478	-0.247	57	0.806
REFORM*EXPERIENCE, γ_{07}	-3.275339	4.088522	-0.801	57	0.426
REFORM*EXPERTISE, γ_{08}	-2.666295	5.917814	-0.451	57	0.654
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-5.582226	0.902788	-6.183	3759	<0.001

DV: Math E (Probability and Statistics)

$\sigma^2 = 233.30854$

τ

INTRCPT1,β_0 42.86920

Final estimation of fixed effects, Math E (Probability and Statistics) is DV
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. d.f.	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	48.050310	1.362907	35.256	57	<0.001
TRADITIO, γ_{01}	3.350397	2.014976	1.663	57	0.102
REFORM, γ_{02}	8.211556	2.837245	2.894	57	0.005
EXPERIEN, γ_{03}	3.301547	2.762882	1.195	57	0.237
EXPERTIS, γ_{04}	1.371517	2.225202	0.616	57	0.540
TRAD*EXPERTISE, γ_{05}	-3.038344	5.153237	-0.590	57	0.558
TRAD*EXPERIENCE, γ_{06}	-0.237714	3.919973	-0.061	57	0.952
REFORM*EXPERIENCE, γ_{07}	-3.173978	4.836776	-0.656	57	0.514
REFORM*EXPERTISE, γ_{08}	-3.224571	6.281242	-0.513	57	0.610
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-5.865126	0.965445	-6.075	3759	<0.001

DV: Math F (Algebra)

$\sigma^2 = 276.55417$

τ

INTRCPT1,β_0 51.44741

Final estimation of fixed effects, Math F (Algebra) is DV
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. d.f.	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	52.168515	1.471539	35.452	57	<0.001
TRADITIO, γ_{01}	4.136356	2.336827	1.770	57	0.082
REFORM, γ_{02}	7.064293	2.955466	2.390	57	0.020
EXPERIEN, γ_{03}	4.345804	2.619210	1.659	57	0.103
EXPERTIS, γ_{04}	2.078615	2.233910	0.930	57	0.356
TRAD*EXPERTISE, γ_{05}	-5.136659	5.574403	-0.921	57	0.361
TRAD*EXPERIENCE, γ_{06}	-1.448996	4.015186	-0.361	57	0.720
REFORM*EXPERIENCE, γ_{07}	-5.473169	5.106594	-1.072	57	0.288
REFORM*EXPERTISE, γ_{08}	-3.982827	6.883369	-0.579	57	0.565
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-6.392935	1.036546	-6.168	3759	<0.001

Appendix G

Variance Explained Calculations

Variance explained by full model, given indicated outcome scores

	τ_{00} null model	τ_{00} level-1 model, SES only	τ_{00} 2-level model with textbook type variables	Proportion of variance accounted for by SES (above null model)	Proportion of variance accounted for by 2-level model with variables for textbook type (above SES model)
Scale Score Math	672.211	574.215	502.446	0.147	0.125
Strand A Math	85.824	70.078	59.425	0.183	0.152
Strand B Math	73.741	60.903	53.754	0.174	0.117
Strand C Math	70.279	58.169	50.193	0.173	0.137
Strand D Math	51.340	41.837	36.295	0.185	0.132
Strand E Math	59.249	49.053	41.872	0.172	0.146
Strand F z-score (constructed response)	68.660	57.057	51.349	0.169	0.100
d response)	0.237	0.206	0.151	0.131	0.267

Appendix H
Constructed Response Models

2-level model with SES and textbook selection only

Level-1 Model

$ZTOTALSC_{ij} = \beta_{0j} + \beta_{1j}*(SES7_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TRADITIONAL_j) + \gamma_{02}*(REFORM_j) + u_{0j}$
$\beta_{1j} = \gamma_{10}$

REFORM has been centered around the grand mean.

Mixed Model

$ZTOTALSC_{ij} = \gamma_{00} + \gamma_{01}*TRADITIONAL_j + \gamma_{02}*REFORM_j$
$+ \gamma_{10}*SES7_{ij} + u_{0j} + r_{ij}$

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	0.226767	0.072635	3.122	63	0.003
TRADITIO, γ_{01}	0.217027	0.130085	1.668	63	0.100
REFORM, γ_{02}	0.636636	0.170094	3.743	63	<0.001
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-0.319159	0.050952	-6.264	4069	<0.001

2- level full model, Constructed response z-scores as DV

Level-1 Model

$ZTOTALSC_{ij} = \beta_{0j} + \beta_{1j}*(SES7_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(TRADITIONAL_j) + \gamma_{02}*(REFORM_j) + \gamma_{03}*(EXPERIENCE_j) + \gamma_{04}*(EXPERTISE_j) + \gamma_{05}*(TRAD*EXPERIENCE) + \gamma_{06}*(TRAD*EXPERTISE) + \gamma_{07}*(REFORM*EXPERIENCE) + \gamma_{08}*(REFORM*EXPERTISE) + u_{0j}$
$\beta_{1j} = \gamma_{10}$

Mixed Model

$ZTOTALSC_{ij} = \gamma_{00} + \gamma_{01}*TRADITIONAL_j + \gamma_{02}*REFORM_j + \gamma_{03}*EXPERIENCE_j + \gamma_{04}*EXPERTISE_j + \gamma_{05}*TRAD*EXPERTISE_j + \gamma_{06}*TRAD*EXPERIENCE_j + \gamma_{07}*REFORM*EXPERIENCE_j + \gamma_{08}*REFORM*EXPERTISE + \gamma_{10}*SES7_{ij} + u_{0j} + r_{ij}$

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	0.245986	0.072471	3.394	57	0.001
TRADITIO, γ_{01}	0.181221	0.121140	1.496	57	0.140
REFORM, γ_{02}	0.587417	0.175980	3.338	57	0.001
EXPERIEN, γ_{03}	0.268867	0.109991	2.444	57	0.018
EXPERTIS, γ_{04}	0.046325	0.099679	0.465	57	0.644
TRAD*EXPERTISE, γ_{05}	-0.275090	0.302538	-0.909	57	0.367
TRAD*EXPERIENCE, γ_{06}	-0.126344	0.202203	-0.625	57	0.535
REFORM*EXPERIENCE, γ_{07}	-0.197325	0.281012	-0.702	57	0.485
REFORM*EXPERTISE, γ_{08}	0.103991	0.350599	0.297	57	0.768
For SES7 slope, β_1					
INTRCPT2, γ_{10}	-0.317082	0.050747	-6.248	4069	<0.001

<center>**Appendix I**</center>

<center>**Growth Models in HLM**</center>

Growth Models, Proficiency Rank as DV

Level-1 Model

$$PROFICIE_{tij} = \pi_{0ij} + \pi_{1ij}*(OCCASSIO_{tij}) + e_{tij}$$

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j}$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + u_{00j}$$
$$\beta_{10j} = \gamma_{100}$$

Mixed Model

$$PROFICIE_{tij} = \gamma_{000} + \gamma_{100}*OCCASSIO_{tij} + r_{0ij} + u_{00j} + e_{tij}$$

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	2.268586	0.054859	41.353	65	<0.001
For OCCASSIO slope, π_1					
For INTRCPT2, β_{10}					
INTRCPT3, γ_{100}	0.038511	0.015957	2.413	5631	0.016

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	0.77131	0.59492	3280	23982.92635	<0.001
level-1, e	0.50478	0.25481			

Growth model with SES, Proficiency Rank as DV

Level-1 Model

$$PROFICIE_{tij} = \pi_{0ij} + \pi_{1ij}*(OCCASSIO_{tij}) + e_{tij}$$

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}*(SES7_{ij}) + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j}$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + u_{00j}$$
$$\beta_{01j} = \gamma_{010}$$
$$\beta_{10j} = \gamma_{100}$$

Mixed Model

$$PROFICIE_{tij} = \gamma_{000} + \gamma_{010}*SES7_{ij} + \gamma_{100}*OCCASSIO_{tij} + r_{0ij} + u_{00j} + e_{tij}$$

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	2.536958	0.063034	40.247	65	<0.001
For SES7, β_{01}					
INTRCPT3, γ_{010}	-0.331463	0.046922	-7.064	3279	<0.001
For OCCASSIO slope, π_1					
For INTRCPT2, β_{10}					
INTRCPT3, γ_{100}	0.038338	0.015940	2.405	5631	0.016

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1,r_0	0.76234	0.58117	3279	23485.82548	<0.001
level-1, e	0.50481	0.25484			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2,u_{00}	0.34305	0.11768	65	613.87216	<0.001

Statistics for the current model

Deviance = 19955.011247
Number of estimated parameters = 6

Growth model including textbook selection, Proficiency Rank as DV

Level-1 Model

$$PROFICIE_{tij} = \pi_{0ij} + \pi_{1ij}*(OCCASSIO_{tij}) + e_{tij}$$

$\sigma^2 = 0.25481$
Standard error of $\sigma^2 = 0.00477$

τ_π
INTRCPT1,π_0 0.59492
Standard error of τ_π
INTRCPT1,π_0 0.01723

Random level-1 coefficient	Reliability estimate
INTRCPT1,π_0	0.856

τ_β
INTRCPT1

INTRCPT2,β_{00} 0.14386

Standard error of τ_β
INTRCPT1

INTRCPT2,β_{00} 0.02899

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}*(SES7_{ij}) + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j}$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(TRADITIO_j) + \gamma_{002}(REFORM_j) + u_{00j}$$
$$\beta_{01j} = \gamma_{010}$$
$$\beta_{10j} = \gamma_{100}$$

Mixed Model

$$PROFICIE_{tij} = \gamma_{000} + \gamma_{001}*TRADITIO_j + \gamma_{002}*REFORM_j + \gamma_{010}*SES7_{ij}$$
$$+ \gamma_{100}*OCCASSIO_{tij} + r_{0ij} + u_{00j} + e_{tij}$$

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	2.445468	0.067470	36.245	63	<0.001
TRADITIO, γ_{001}	0.173549	0.115393	1.504	63	0.138
REFORM, γ_{002}	0.315640	0.122077	2.586	63	0.012
For SES7, β_{01}					
INTRCPT3, γ_{010}	-0.331974	0.046622	-7.121	3279	<0.001
For OCCASSIO slope, π_1					
For INTRCPT2, β_{10}					
INTRCPT3, γ_{100}	0.038347	0.015938	2.406	5631	0.016

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	0.76227	0.58105	3279	23485.62440	<0.001
level-1, e	0.50482	0.25484			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2,u_{00}	0.31971	0.10222	63	597.26179	<0.001

Statistics for the current model

Deviance = 19946.824102
Number of estimated parameters = 8

Full Model, Proficiency Rank as DV

Level-1 Model

$$PROFICIE_{tij} = \pi_{0ij} + \pi_{1ij}*(OCCASSIO_{tij}) + e_{tij}$$

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}*(SES7_{ij}) + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j}$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(TRADITIO_j) + \gamma_{002}(REFORM_j) + \gamma_{003}(EXPERIEN_j) + \gamma_{004}(EXPERTIS_j)$$
$$+ \gamma_{005}(TRAD*EXPERIENCE) + \gamma_{006}(TRAD*EXPERTISE) +$$
$$\gamma_{007}(REFORM*EXPERIENCE) + \gamma_{008}(REFORM*EXPERTISE) + u_{00j}$$
$$\beta_{01j} = \gamma_{010}$$
$$\beta_{10j} = \gamma_{100}$$

Mixed Model

$$PROFICIE_{tij} = \gamma_{000} + \gamma_{001}*TRADITIO_j + \gamma_{002}*REFORM_j + \gamma_{003}*EXPERIEN_j$$
$$+ \gamma_{004}*EXPERTIS_j + \gamma_{005}*TRAD*EXPERTISE_j + \gamma_{006}*TRAD*EXPERIENCE_j +$$
$$\gamma_{007}*REFORM*EXPERIENCE_j$$
$$+ \gamma_{008}*REFORM*EXPERTISE + \gamma_{010}*SES7_{ij} + \gamma_{100}*OCCASSIO_{tij}$$
$$+ r_{0ij} + u_{00j} + e_{tij}$$

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
For INTRCPT2, β_{00}					
INTRCPT3, γ_{000}	2.471394	0.067754	36.476	57	<0.001
TRADITIO, γ_{001}	0.154100	0.104900	1.469	57	0.147
REFORM, γ_{002}	0.392763	0.140171	2.802	57	0.007
EXPERIEN, γ_{003}	0.239651	0.148576	1.613	57	0.112
EXPERTIS, γ_{004}	0.021421	0.118624	0.181	57	0.857
TRAD*EXPERTISE, γ_{005}	-0.349007	0.292982	-1.191	57	0.239
TRAD*EXPERIENCE, γ_{006}	-0.093135	0.190693	-0.488	57	0.627
REFORM*EXPERIENCE, γ_{007}	-0.191870	0.259430	-0.740	57	0.463
REFORM*EXPERTISE, γ_{008}	-0.250286	0.296135	-0.845	57	0.402
For SES7, β_{01}					
INTRCPT3, γ_{010}	-0.334027	0.047098	-7.092	3279	<0.001
For OCCASSIO slope, π_1					
For INTRCPT2, β_{10}					
INTRCPT3, γ_{100}	0.038316	0.015937	2.404	5631	0.016

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, r_0	0.76239	0.58124	3279	23485.53403	<0.001
level-1, e	0.50482	0.25485			

Final estimation of level-3 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2, u_{00}	0.29982	0.08989	57	505.05798	<0.001

Deviance = 19941.035128
Number of estimated parameters = 14

Appendix J
Binary Logistic Regression with Demographics

Textbook selection as DV, Reform = 1; Independent variables – demographics of school

	B	S.E.	Wald	df	Sig.
black_mean	-1874.378	3926070.945	.000	1	1.000
hispaniclatino_mean	-1874.189	3926070.945	.000	1	1.000
white_mean	-1868.779	3926070.945	.000	1	1.000
asian_mean	-1884.958	3926070.945	.000	1	1.000
natam_mean	-1922.429	3926070.945	.000	1	1.000
gender_mean	-13.217	8.352	2.504	1	.114
disadvantage_mean	1.215	4.547	.071	1	.789
sped_mean	-4.391	6.626	.439	1	.508
ELL_mean_1	-.248	6.737	.001	1	.971
Constant	1879.472	3926070.945	.000	1	1.000

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.023	4	.554
	Block	3.023	4	.554
	Model	12.055	9	.210

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	51.622[a]	.177	.275

a. Estimation terminated at iteration number 20 because maximum
iterations has been reached. Final solution cannot be found.

Classification Table[a]

Observed			Predicted		
			dummyreform		Percentage Correct
			0	1	
Step 1	dummyreform	0	46	3	93.9
		1	9	4	30.8
	Overall Percentage				80.6

a. The cut value is .500

Appendix K
Binary Logistic Regression with Human Capital

Textbook selection as DV, Reform = 1; Independent variables – human capital
component scores

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
	Step	9.120	2	.010
Step 1	Block	9.120	2	.010
	Model	9.120	2	.010

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	56.375[a]	.129	.205

a. Estimation terminated at iteration number 5 because parameter
estimates changed by less than .001.

Classification Table[a]

Observed			Predicted		
			dummyreform		Percentage Correct
			0	1	
Step 1	dummyreform	0	51	2	96.2
		1	12	1	7.7
	Overall Percentage				78.8

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Experience	-.424	.559	.575	1	.448	.654	.219	1.958
	Expertise	2.005	.757	7.010	1	.008	7.425	1.683	32.750
	Constant	-1.659	.372	19.852	1	.000	.190		

a. Variable(s) entered on step 1: experience01, expertise02.